Published in final edited form as:

Nat Chem. 2015 May ; 7(5): 411–417. doi:10.1038/nchem.2221.

The colibactin warhead crosslinks DNA

Maria I. Vizcaino[1,2] and **Jason M. Crawford**[1,2,3,*]

[1]Department of Chemistry, Yale University, New Haven, Connecticut 06510, United States

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

[3]Department of Microbial Pathogenesis, Yale School of Medicine, New Haven, Connecticut 06536, Unites States

Abstract

Members of the human microbiota are increasingly being correlated to human health and disease states, but the majority of the underlying microbial metabolites that regulate host-microbe interactions remain largely unexplored. Select strains of *E. coli* present in the human colon have been linked to initiating inflammation-induced colorectal cancer through an unknown small molecule-mediated process. The responsible nonribosomal peptide-polyketide hybrid pathway encodes "colibactin," a largely uncharacterized family of small molecules. Genotoxic small molecules from this pathway capable of initiating cancer formation have remained elusive due to their high instability. Guided by metabolomic analyses, here we employ a combination of NMR spectroscopy and bioinformatics-guided isotopic labeling studies to characterize the colibactin warhead, an unprecedented substituted spirobicyclic structure. The warhead crosslinks duplex DNA *in vitro*, providing direct experimental evidence for colibactin's DNA-damaging activity. The data support unexpected models for both colibactin biosynthesis and its mode of action.

The human microbiome encodes at least two orders of magnitude more genes than the human genome[1,2], representing a large biocatalytic reservoir for small molecule biosynthesis and processing. Microbial genomics and bioinformatic predictions have unveiled a remarkable array of unknown human-associated microbial secondary metabolites in various structural classes, including nonribosomal peptides, polyketides, saccharides, terpenoids, and ribosomally-encoded peptides[3]. These bacterial-derived small molecules can affect both microbial community structure and host physiology by regulating a variety of metabolic processes, such as antibiotic-associated microbiota imbalance and benzodiazepine-mediated hemorrhagic colitis[4]. While select bacteria are being correlated at the nucleotide-sequence level to host health and diseases, including various cancers, cystic fibrosis, cardiovascular

*Correspondence and requests for materials should be addressed to J.M.C. jason.crawford@yale.edu.

Author contributions

M.I.V. and J.M.C. conceived and designed the experiments; M.I.V. performed the experiments; M.I.V. and J.M.C. analyzed the data and wrote the paper.

Supplementary information is available in the online version of the paper.

Competing financial interests

The authors declare no competing financial interests

diseases, human behaviors, and others[5], little to nothing is known about their small molecule contributions to human health and their regulatory interkingdom chemical signaling processes.

The “colibactin” pathway illustrates how a bacterial nonribosomal peptide synthetase (NRPS)-polyketide synthase (PKS) hybrid gene cluster (*clb* genomic island, Supplementary Fig. S1) has been phenotypically linked to colorectal cancer pathogenesis while its encoded genotoxic small molecules have eluded structural characterization. Presence of the orphan 54-kilobase *clb* gene cluster in select strains of *E. coli* ultimately leads to mammalian cell DNA double-strand breaks *in vitro*[6] and *in vivo*[7], contributing to colorectal cancer formation under host inflammatory conditions[8,9]. The pathway modulates host immunity[10], exacerbates lymphopenia leading to sepsis[11], endows its bacterial producer with long-term persistence in the gut[12], and is found at a much higher prevalence in colorectal cancer patients[8]. Recently, the *clb* pathway was interestingly found to induce cellular senescence in transiently infected mammalian cells, leading to DNA double-strand breaks in bystander mammalian cells through downstream paracrine signaling even in the absence of bacterial cells[13-15].

Detailed mode-of-action studies for the *clb* pathway have remained experimentally intractable without its corresponding small molecule structures. However, colibactin’s instability has thwarted structural characterization efforts despite attempts from various labs over the last nine years[6]. Employing molecular networking tools[16], we recently developed a pathway-targeted structural network analysis approach to map the *clb* pathway within *E. coli*’s complex metabolome. To validate the approach, we isolated, structurally characterized, and synthesized the most abundant molecular features in the network map, including the first authentic precolibactin assembly line derailment product **15**[17] (Fig. 1; Structures **1-32** are numbered in order of biosynthetic complexity as detailed below). **15** is proteolytically cleaved by peptidase ClbP to liberate small alkyl-amine **16**, which accumulated in organic extracts as cyclic imine **17**, and N-myristoyl-D-Asn **1**, which exhibited *in vitro* bacterial growth inhibitory activity and served as an antagonist of the 5-hydroxytryptamine-7 receptor implicated in colitis[17]. Here we elucidate the structure of the relatively stable precolibactin derailment product **27**, illuminating the colibactin warhead, an unprecedented substituted spirobicyclic structural feature. Correspondingly, we employ [U-^{13}C]-isotopic labeling studies in various auxotrophic strain backgrounds in conjunction with pathway-targeted molecular network analyses among wildtype, peptidase *clbP* mutant, and control pathways to provide a system-wide analysis of colibactin biosynthesis. The data collectively support the structures of highly unstable advanced precolibactins and an unexpected biosynthetic model that accounts for 32 predicted *clb*-dependent molecular features. In contrast to the reported DNA double-strand break phenotype, we demonstrate that the colibactin warhead crosslinks DNA *in vitro*, supporting a new model for colibactin’s mode of action.

Results and discussion

Structural characterization of the colibactin warhead

Colibactin belongs to a subset of hybrid polyketide-nonribosomal peptides that undergo a prodrug activation mechanism[18-21]. During colibactin maturation, the inner membrane peptidase ClbP cleaves “precolibactins” in the periplasm[22,23], liberating N-terminal N-acyl-D-asparagines and the unknown C-terminal “colibactins”[17,20,24]. To focus our colibactin structural characterization efforts, we initially assessed secondary metabolic flux of *clb* pathway-dependent molecular features in wildtype (*clb*+) and *clbP* mutant (Δ*clbP*) strains (Supplementary Table S1) by comparative metabolomics (Supplementary Table S2) and network analyses (Supplementary Fig. S2). By comparing MS ionization intensities of closely related features in a given experiment, a qualitative assessment of metabolite abundances can be determined relative to one another in a molecular network cluster. We analyzed the organic-extractable metabolomes from *clb*+ and Δ*clbP* heterologous systems, containing the *clb* pathways from the meningitis isolate *E. coli* IHE3034 expressed in *E. coli* DH10B, and homologous systems of the probiotic *E. coli* Nissle 1917[17]. ClbP was previously determined to cleave **15**[17], and our metabolomic analyses indicate that ClbP is promiscuous. Consequently, we focused on the structural characterization of the most abundant advanced precolibactins from the *clbP* mutants (Fig. 2, Supplementary Table S2) to correlate the colibactin biosynthetic pathway to precolibactin structure(s).

We initially focused on the isolation of three Δ*clbP* metabolites with ion masses of *m/z* 816.3780, *m/z* 609.3862, and *m/z* 547.3859, which were biosynthesized by both Δ*clbP* IHE3034 and Δ*clbP* Nissle 1917 strains (Fig. 2, Supplementary Table S2). The molecules could be detected by Q-TOF HRMS (Supplementary Fig. S3) in freshly prepared organic extracts, but *m/z* 816.3780 and *m/z* 609.3862 reproducibly deteriorated in subsequent isolation attempts despite substantial efforts to identify stabilizing conditions in the complex extracts. The Δ*clbP* metabolite with *m/z* 547.3859 was relatively stable and survived extensive normal- and reverse-phase chromatographic processing, leading to ~1.0 mg of pure material derived from an 18 L Δ*clbP* culture. This molecule was subjected to detailed one- (^{1}H, ^{13}C) and two-dimensional (gCOSY, gHSQCAD, and gHMBCAD) NMR studies in DMSO-d_6 (Supplementary Table S3, Supplementary Figs. S4-S7) and methanol-d_4 (Supplementary Table S4, Supplementary Figs. S8-S13). The NMR studies strongly support the presence of a substituted spirobicyclic structure (Fig. 1), illuminating shunt precolibactin **27** and its attenuated colibactin warhead. The heterocyclic 4-azaspiro[2,4]hept-6-en-5-one substructure in **27** is not found in any natural molecule described to date, revealing unprecedented biosynthetic potential. Crucially, this structure could not have been predicted from the biosynthetic pathway and current NRPS-PKS logic. The spirobicyclic core substructure with different substitutions appears in some synthetic molecules, and their ^{1}H and ^{13}C chemical shifts are in agreement with our proposed structure (Supplementary Figs. S14-S15). We also observed hydrogen-deuterium exchange at the warhead methyl substituent in NMR (Supplementary Figs. S4 and S16) and post-NMR HRMS spectra (Supplementary Fig. S16), indicating exchange of acidic methyl protons in the attenuated warhead. The structure was further supported by HRMS (*m/z* 547.3859) consistent with the $[M+H]^+$ ion formula $[C_{30}H_{51}N_4O_5]^+$ (theoretical *m/z* 547.3859, Δppm 0.0), MS/MS

fragmentation (Supplementary Fig. S17), and subsequent isotopic labeling studies (detailed below). The *R, S*-configurations of the Asn- and Ala-derived stereocenters were previously determined by synthesis of **15**[17]. Analogous to these earlier characterizations of **14** and **15, 27** could result from decarboxylation of an unstable assembly line derailment product **26** (Fig. 1). LC-HRMS analyses of wildtype and *clbP* mutant strains support **27** as a substrate for promiscuous ClbP-mediated cleavage, leading to **1** and **30** (Fig. 1, Supplementary Table S2).

Isotope-labeled precursors illuminate colibactin biosynthesis

Merging our new understanding of the spirobicyclic structure with retrospective bioinformatic analyses of the biosynthetic pathway and metabolomic network analyses, we predicted the structures for the two other targeted advanced precolibactins (*m/z* 609.3862 and *m/z* 816.3780). Our proposed structures were supported by HRMS and MS/MS fragmentation (Supplementary Figs. S18-S20). To gain additional structural support for all detectable *clb* pathway-dependent metabolites, including these highly unstable advanced precolibactins, we conducted an extensive series of isotopic labeling studies using universally [U-^{13}C]-labeled L-amino acids in an *E. coli* BW25113 parent strain and in a variety of amino acid auxotrophic strain variants (Supplementary Table S1). Studies were conducted using a control bacterial artificial chromosome (*clb*-) and the IHE3034-derived bacterial artificial chromosomes bearing *clb*+ and Δ*clbP* colibactin pathways. Labeling studies in defined minimal media were guided by bioinformatic predictions of adenylation (A)-domain specificities (Supplementary Fig. S1), established NRPS-PKS logic, and our structural characterizations of the two shunt precolibactins **15**[17] and **27**. By comparing HRMS analyses of ^{13}C-labeled cultures to our detectable comprehensive list of *clb*-dependent metabolites, we determined system-wide specific L-amino acid (Asn, Ala, Met, Gly, Cys, Ser) isotopic incorporations (Fig. 2, Supplementary Table S2). From this data, 32 predicted metabolites (Supplementary Table S5) were mapped onto the biosynthetic pathway model (Fig. 3). The labeling studies below are described in the context of this model.

As previously reported, **15** supports the sequential action of NRPS ClbN, NRPS-PKS hybrid ClbB, and *trans*-acyl-transferase (*trans*-AT) PKS ClbC (Fig. 3a)[17]. Specifically, the ClbN A domain activates L-Asn and transfers it onto its cognate thiolation (T) domain, the epimerization (E) domain epimerizes the tethered amino acid, and the condensation (C) domain condenses acyl-CoA substrates to generate an N-acyl-D-Asn-tethered T domain, which is a substrate for downstream ClbB[20]. The NRPS segment of ClbB, in which the A domain is predicted to activate Val, rather preferentially condenses an L-Ala *in vitro*[20] and *in vivo*[17]. The PKS portion of ClbB loads a malonyl-CoA substrate by its *cis*-AT domain and catalyzes one polyketide extension with complete reductive processing to the saturated hydrocarbon chain. The *trans*-AT PKS ClbC, which contains a deteriorated nonfunctional AT domain[25], catalyzes one additional round of polyketide extension with no reductive processing. Here, malonyl-CoA loading could occur in trans by interacting with fatty acid biosynthesis, other embedded ATs in the pathway, or the discretely expressed AT ClbG consistent with other trans-AT PKS systems[26]. Incorporation of L-[U-^{13}C]-Asn was observed in 47 BW25113 *clb*+ and Δ*clbP*-derived metabolites (Fig. 2, Supplementary Table

S2), supporting the processed acyl-D-Asn structures **1-5** (Fig. 3c, Supplementary Table S5). As the preferred amino acid for the ClbB A domain, Ala incorporated into 37 metabolites (Fig. 2, Supplementary Table S2, Supplementary Figs. S21-S24), 11 of which could readily be assigned as shunt products from ClbN-ClbB-ClbC (**10-21**; Fig. 3a, Supplementary Table S5).

The structure of **27** provided further biosynthetic insights, allowing expansion of the biosynthetic model. The spirocyclopropane moiety suggested assembly line utilization of a 1-aminocyclopropane-1-carboxylic acid (ACC) extender unit, a substrate found in other bacterial small molecules[27]. In the biosynthesis of the cytotoxin cytotrienin by a *Streptomyces* sp., the ACC unit is derived from Met as determined by isotopic labeling studies[28]. We hypothesized that Met would similarly be used as a precursor for the ACC unit in colibactin warhead biosynthesis. To test our prediction, L-[U-^{13}C]-Met was fed to a methionine auxotroph containing *clb*+, Δ*clbP*, or *clb*- pathways. As expected, we only observed specific incorporation of four carbons throughout the Met-labeled network (Fig. 2b) in contrast to the substrate's five-labeled carbons, supporting that the ACC unit is derived from the aminobutyryl moiety of L-Met. Incorporation of free deuterium [2,2,3,3-D]–labeled ACC was not observed in any of these Met-labeled metabolites, suggesting cyclization of Met or a Met-derived precursor (*e.g.*, S-adenosyl-Met) directly on the assembly line. The labeling experiments further support the structure and biosynthesis of **27** and its C-terminal ClbP cleavage product **30** (Fig. 1), in addition to six other proposed Δ*clbP* metabolites (**24-25** and **28-32**; Fig. 3b; Supplementary Tables S2 and S5).

While it was proposed that a cryptic chlorination and subsequent cyclization event might be involved in ACC-unit formation in cytotrienin biosynthesis[29,30], the *clb* gene cluster lacks such enzymes. The *clb* gene cluster similarly lacks the type of pyridoxal-5′-phosphate-dependent enzymes found in plants necessary to make ACC from S-adenosyl-Met[27], suggesting a new and convergent evolutionary route for this ring-strained NRPS building block. Nevertheless, NRPS ClbH-dependent condensation of an ACC unit, PKS ClbI polyketide extension, and subsequent warhead cyclization (intramolecular Knoevenagel-type condensation) could account for **27** biosynthesis (Fig. 3b, Supplementary Fig. S25). The first A domain in ClbH (ClbH-A_1) is predicted to activate Ser while the second A domain has only very poor predictions for Val activation (Supplementary Fig. S1). Based on protein homologies to previously characterized enzyme systems and prior to determining the structure of **27**, we initially predicted that ClbH-A_1 might participate with free-standing carrier protein (T, ClbE), dehydrogenases ClbDF, and discretely expressed AT ClbG to generate and load Ser-derived α-amino malonate polyketide extender units, which are found in related systems, such as in zwittermicin biosynthesis[31]. However, we did not detect intact L-[U-^{13}C]-Ser labeling ($^{13}C_3$) of any pathway-dependent features in our network, nor did we detect Ser-derived α-amino malonate polyketide extender unit incorporation ($^{13}C_2$), supporting either undetectable production under our experimental conditions or alternative enzymatic roles in colibactin biosynthesis. Our study now necessitates further mechanistic enzymatic studies for ClbH-dependent amino acid activation and ACC cyclization. Polyketide extension (ClbI) would be required before the intramolecular Knoevenagel-type condensation could take place, leading to the spirobicyclic colibactin warhead

(Supplementary Fig. S25). The network data and accumulation of predicted **25** (linear structure supported by MS/MS fragmentation, Supplementary Figs. S18-S19) and cyclized **27** suggest co-assembly line warhead cyclization (Fig. 3b).

While instability of advanced precolibactins precluded their NMR-based structure elucidation, a retrospective bioinformatic analysis in combination with the network data led to a proposed biosynthesis for their construction. Ensuing warhead cyclization, a largely "co-linear" biosynthetic interpretation of ClbJK leads to advanced precolibactin A (**32**, Fig. 3b,c) with the ketosynthase (KS) domain in ClbK serving a transthioesterification role to interface ClbJK catalytic activities. To gain support for this prediction, we conducted [U-^{13}C]-Gly and L-[U-^{13}C]-Cys labeling studies, as ClbJK were predicted to canonically activate and process these amino acids. Initially, [U-^{13}C]-labeled and non-labeled amino acids were fed in a 1:1 ratio, and subsequently, only [U-^{13}C]-labeled amino acids were supplemented to confirm the observed results. These relative dose-response studies were necessary, as the advanced precolibactins were observed at low abundances. ClbJK activities appeared to be very processive, as all Gly labeled molecules also contained Cys labeling. As predicted, precolibactin A (**32**) incorporated one Gly and two Cys substrates as supported by HRMS of labeled and nonlabeled precolibactin A (Supplementary Fig. S24). Additionally, **32** incorporated four deuteriums in a L-[2,3,3-D]-Cys feeding experiment, supporting the presence of one thiazoline and one thiazole in its predicted structure (Supplementary Fig. S26).

The colibactin warhead crosslinks DNA *in vitro*

The unprecedented structural features illuminated in our characterized shunt precolibactin **27** and predicted advanced precolibactin **32** served as a chemical guide for theoretical insights into colibactin's modes of action. The colibactin warhead shares the structural hallmarks of "cyclopropane trigger compounds"[32] with the requisite labile ring-strained spirocyclopropyl substituent for nucleophile-induced ring opening and irreversible covalent binding to its targets. Additionally, the proposed ClbP-catalyzed cleavage of precolibactin **32** would liberate the warhead with its N-terminal primary amine and predicted C-terminal thiazolinyl-thiazole tail (Fig. 3c, Fig. 4a). These flanking features are consistent with DNA/RNA being colibactin's primary targets. Terminal amines are common in small molecules that bind DNA and RNA, such as the classically studied bleomycins, and participate in electrostatic interactions with the macromolecules' phosphate moieties[33]. Additionally, the thiazolinyl-thiazole tail of related phleomycins contributes to their partial intercalative mode of DNA binding[34]. In contrast to the bleomycins' and phleomycins' ability to induce DNA double-strand breaks, the colibactin warhead is reminiscent of the duocarmycin family of DNA alkylators[35]. In an analogous reaction, the colibactin warhead could alkylate its targets via a homo-Michael addition reaction (Fig. 4a). NMR solution studies demonstrate that 2-hydroxypyrroles with acyl-substituents in the 3-position, such as the proposed alkylated-intermediate, strongly favor their keto-tautomeric forms[36], as shown boxed in Figure 4a. Intriguingly, this suggests that upon alkylation, colibactin could present a second Michael acceptor, expanding its covalent functional utilities relative to the duocarmycins. In the case of DNA alkylation, a pseudo-intramolecular Michael addition reaction could generate DNA interstrand crosslinks (Fig. 4b).

To initiate functional studies and provide support for our hypothesis of the colibactin warhead's mode of action, we conducted a series of small-scale *in vitro* reactions with duplex DNA and the minute amount (~1.0 mg) of isolated precolibactin **27**. Since it was previously speculated that colibactin induces DNA double-strand breaks, we first tested **27**'s ability to cleave double-strand DNA. We did not detect *in vitro* double-strand break activity for **27** using plasmid pBR322 DNA in the presence or absence of reducing agents, dithiothreitol and β-mercaptoethanol (Supplementary Fig. S27). Instead, we consistently observed weak DNA interstrand crosslinks as hypothesized above by alkaline gel assays using **27** and *Eco*RI-linearized pBR322 DNA (Fig. 4c). The weak activity could be detected with and without reducing agents. Notably, control precolibactin **15** lacking the spirobicyclic structure was inactive under the conditions of the assays (Fig. 4c), suggesting that the warhead directly alkylates and crosslinks DNA. While the observed activity was weak compared to crosslinker mitomycin C (positive control), we anticipate that ClbP-mediated liberation of the N-terminal amine (or its cyclized imine form; Supplementary Fig. S28), presence of the apparently destabilizing C-terminal exocyclic amide, and inclusion of the C-terminal tail would increase DNA binding and enhance reactivity leading to more efficient crosslinking. DNA interstrand crosslinks can be quite toxic, impairing multiple cellular processes, such as replication and transcription, and inducing multiple DNA repair machineries and downstream DNA double-strand breaks[37,38]. Our data support a model in which the colibactin warhead's *in vitro* alkylation and DNA interstrand crosslink activities may account for initiating *clb*-dependent genotoxic activities.

Conclusion

We combined NMR-based structural elucidation, bioinformatics, system-wide isotopic labeling studies, pathway-dependent structural network analyses, and *in vitro* activity studies to map the colibactin pathway, propose a biosynthetic model for its largely intractable family of molecules linked to colorectal cancer formation, and assess its unexpected *in vitro* activity. With our current understanding of nonribosomal peptide and polyketide biosynthesis, the proposed structure for precolibactin A accounts for the majority of biosynthetic enzymes in the colibactin pathway (Fig. 3). Our data provide a working foundation for determining the remaining (pre)colibactin structures (*e.g.*, the detectable metabolites listed in Supplementary Table S2) and enzymatic roles in colibactin biosynthesis, processing, and cellular resistance. Indeed, speculative proposals can now be made for the unaccounted *clb* proteins that currently do not have experimentally supported roles (Supplementary Fig. S1). For example, the protein we call "ClbS" is not required for *clb* pathway activity[6], but it was identified in an activity-based protein profiling experiment in *clb*-positive pathogenic *E. coli*, in which ClbS covalently captured Michael acceptors[39]. In light of our proposed structures and warhead activity, predicted hydrolase ClbS could serve a role in (pre)colibactin binding, trafficking, and/or resistance in addition to the canonical *clb* drug transporter ClbM. The warhead structure itself, which crosslinks DNA in our *in vitro* assays, does not support speculations that colibactin could directly induce DNA double-strand breaks. The double-strand break response in human cells is latently observed well after transient *clb*-positive bacterial infection[6]. Combined with the notion that chronic DNA double-strand breaks can be induced in bystander mammalian cells after transient

bacterial infection[13,14], our structural and *in vitro* functional data support that the mammalian cell DNA double-strand break response could be downstream of colibactin alkylation and crosslinking. Collectively, our studies here will enable structure-guided approaches to mechanistically elucidate colibactin's biosynthesis, synthesis, product diversification, processing, cellular trafficking, and modes of action.

Methods

Production, purification, and structural elucidation of 27

An 18 L culture of *E. coli* DH10B pBAC Δ*clbP* was grown in a defined M9 minimal medium for 72 h at 25 °C (further described in Supplementary Information). Cell pellets were harvested and extracted with 1.8 L methanol. The organic extract was dried *in vacuo* and fractionated over extensive normal- and reverse-phase chromatography. **27** was monitored by UV at 280 nm and by LC-MS in positive ion mode. The structure of isolated **27** was elucidated by extensive interpretation of 1D (^{1}H, ^{13}C) and 2D (gCOSY, gHSQCAD, gHMBCAD) NMR spectral data in DMSO-d_6 and methanol-d_4, and MS experiments (HRMS, MS/MS) (see Extended Experimental Methods in Supplementary Information).

Isotopic labeling experiments

Guided by bioinformatic predictions of *clb* NRPS adenylation domain specificities, amino acid labeling experiments were performed with [U-^{13}C] L-Asn, L-Ala, L-Met, Gly, L-Cys, and L-Ser, in addition to [2,2,3,3-D]–ACC and L-[2,3,3-D]-Cys. *E. coli* BW25113 parent strain and a set of amino acid auxotrophs were transformed with pBAC *clb*-, pBAC *clb*+, or pBAC Δ*clbP* bacterial artificial chromosomes and used in the labeling experiments (see Extended Experimental Methods in Supplementary Information).

Crosslinking assay

The alkaline agarose gel electrophoresis assay was conducted to assess DNA interstrand crosslink activity and was adapted from previous protocols[40-42]. Small-scale reactions (10 μL) were conducted with EcoRI-linearized pBR322 plasmid DNA inoculated with DMSO (negative control), mitomycin C (positive control), compound **15** or **27** in the presence or absence of a reducing agent (see Extended Experimental Methods in Supplementary Information).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank T. Tørring, H-B. Park, and E. Trautman for feedback and critically reviewing a preliminary version of the manuscript. This work was supported by the National Institutes of Health (National Cancer Institute grant, 1DP2CA186575) and the Damon Runyon Cancer Research Foundation (DFS:05-12).

References

1. Gill SR, et al. Metagenomic analysis of the human distal gut microbiome. Science. 2006; 312:1355–1359. [PubMed: 16741115]
2. Turnbaugh PJ, et al. The human microbiome project. Nature. 2007; 449:804–810. [PubMed: 17943116]
3. Donia MS, et al. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. Cell. 2014; 158:1402–1414. [PubMed: 25215495]
4. Schneditz G, et al. Enterotoxicity of a nonribosomal peptide causes antibiotic-associated colitis. Proc Natl Acad Sci U S A. 2014; 111:13181–13186. [PubMed: 25157164]
5. Sharon G, et al. Specialized metabolites from the microbiome in health and disease. Cell Metab. 2014; 20:719–730. [PubMed: 25440054]
6. Nougayrède J-P, et al. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. Science. 2006; 313:848–851. [PubMed: 16902142]
7. Cuevas-Ramos G, et al. *Escherichia coli* induces DNA damage *in vivo* and triggers genomic instability in mammalian cells. Proc Natl Acad Sci U S A. 2010; 107:11537–11542. [PubMed: 20534522]
8. Arthur JC, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science. 2012; 338:120–123. [PubMed: 22903521]
9. Arthur JC, et al. Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. Nat Commun. 2014; 5:4724. [PubMed: 25182170]
10. Olier M, et al. Genotoxicity of *Escherichia coli* Nissle 1917 strain cannot be dissociated from its probiotic activity. Gut Microbes. 2012; 3:501–509. [PubMed: 22895085]
11. Marcq I, et al. The genotoxin colibactin exacerbates lymphopenia and decreases survival rate in mice infected with septicemic *Escherichia coli*. J Infect Dis. 2014; 210:285–294. [PubMed: 24489107]
12. Nowrouzian FL, Oswald E. *Escherichia coli* strains with the capacity for long-term persistence in the bowel microbiota carry the potentially genotoxic pks island. Microb Pathog. 2012; 53:180–182. [PubMed: 22709536]
13. Secher T, Samba-Louaka A, Oswald E, Nougayrede JP. *Escherichia coli* producing colibactin triggers premature and transmissible senescence in mammalian cells. PLoS One. 2013; 8:e77157. [PubMed: 24116215]
14. Cougnoux A, et al. Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. Gut. 2014; 63:1932–1942. [PubMed: 24658599]
15. Dalmasso G, Cougnoux A, Delmas J, Darfeuille-Michaud A, Bonnet R. The bacterial genotoxin colibactin promotes colon tumor growth by modifying the tumor microenvironment. Gut Microbes. 2014; 5:675–680. [PubMed: 25483338]
16. Watrous J, et al. Mass spectral molecular networking of living microbial colonies. Proc Natl Acad Sci U S A. 2012; 109:E1743–E1752. [PubMed: 22586093]
17. 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]
18. Kevany BM, Rasko DA, Thomas MG. Characterization of the complete zwittermicin A biosynthesis gene cluster from *Bacillus cereus*. Appl Environ Microb. 2009; 75:1144–1155.
19. Reimer D, Pos KM, Thines M, Grün P, Bode HB. A natural prodrug activation mechanism in nonribosomal peptide synthesis. Nat Chem Biol. 2011; 7:888–890. [PubMed: 21926994]
20. 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]
21. Reimer D, Bode HB. A natural prodrug activation mechanism in the biosynthesis of nonribosomal peptides. Nat Prod Rep. 2014; 31:154–159. [PubMed: 24356302]
22. 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]

23. Cougnoux A, et al. Analysis of structure-function relationships in the colibactin-maturating enzyme ClbP. J Mol Biol. 2012; 424:203–214. [PubMed: 23041299]
24. Bian X, et al. In vivo evidence for a prodrug activation mechanism during colibactin maturation. ChemBioChem. 2013; 14:1194–1197. [PubMed: 23744512]
25. Engel P, Vizcaino MI, Crawford JM. Gut symbionts from distinct hosts exhibit genotoxic activity via divergent colibactin biosynthetic pathways. Appl Environ Microb. 201410.1128/aem.03283-14
26. Piel J. Biosynthesis of polyketides by trans-AT polyketide synthases. Nat Prod Rep. 2010; 27:996–1047. [PubMed: 20464003]
27. Thibodeaux CJ, Chang WC, Liu HW. Enzymatic chemistry of cyclopropane, epoxide, and aziridine biosynthesis. Chem Rev. 2012; 112:1681–1709. [PubMed: 22017381]
28. Zhang H-P, Kakeya H, Osada H. Biosynthesis of 1-aminocyclopropane-1-carboxylic acid moiety on cytotrienin A in *Streptomyces* sp. Tetrahedron Lett. 1998; 39:6947–6948.
29. Ueki M, et al. Enzymatic generation of the antimetabolite γ,γ-dichloroaminobutyrate by NRPS and mononuclear iron halogenase action in a Streptomycete. Chem Biol. 2006; 13:1183–1191. [PubMed: 17114000]
30. Kelly WL, et al. Characterization of the aminocarboxycyclopropane-forming enzyme CmaC. Biochemistry. 2007; 46:359–368. [PubMed: 17209546]
31. Chan YA, et al. Hydroxymalonyl-acyl carrier protein (ACP) and aminomalonyl-ACP are two additional type I polyketide synthase extender units. Proc Natl Acad Sci U S A. 2006; 103:14349–14354. [PubMed: 16983083]
32. Wessjohann LA, Brandt W, Thiemann T. Biosynthesis and metabolism of cyclopropane rings in natural compounds. Chem Rev. 2003; 103:1625–1648. [PubMed: 12683792]
33. Hecht SM. Bleomycin: new perspectives on the mechanism of action. J Nat Prod. 2000; 63:158–168. [PubMed: 10650103]
34. Wu W, et al. Solution structure of the hydroperoxide of Co(III) phleomycin complexed with d(CCAGGCCTGG)2: evidence for binding by partial intercalation. Nucleic Acids Res. 2002; 30:4881–4891. [PubMed: 12433991]
35. MacMillan KS, Boger DL. Fundamental relationships between structure, reactivity, and biological activity for the duocarmycins and CC-1065. J Med Chem. 2009; 52:5771–5780. [PubMed: 19639994]
36. Egorova AY, Timofeeva ZY. Reactivity of pyrrol-2-ones. Chem Heterocycl Comp. 2004; 40:1243–1261.
37. Muniandy PA, Liu J, Majumdar A, Liu ST, Seidman MM. DNA interstrand crosslink repair in mammalian cells: step by step. Crit Rev Biochem Mol Biol. 2010; 45:23–49. [PubMed: 20039786]
38. Deans AJ, West SC. DNA interstrand crosslink repair and cancer. Nat Rev Cancer. 2011; 11:467–480. [PubMed: 21701511]
39. Kunzmann MH, Sieber SA. Target analysis of α-alkylidene-γ-butyrolactones in uropathogenic *E. coli*. Mol Biosyst. 2012; 8:3061–3067. [PubMed: 22990910]
40. Cech TR. Alkaline gel electrophoresis of deoxyribonucleic acid photoreacted with trimethylpsoralen: rapid and sensitive detection of interstrand cross-links. Biochemistry. 1981; 20:1431–1437. [PubMed: 6261794]
41. Tepe JJ, Williams RM. DNA cross-linking by a phototriggered dehydromonocrotaline progenitor. J Am Chem Soc. 1999; 121:2951–2955.
42. Kim JJ, Kim HR, Lee SH. Studies on activation mechanism of a mitomycin dimer, 7-N,7'-N'-(1", 2"-dithiepanyl-3",7"-dimethylenyl)bismitomycin C. Arch Pharm Res. 2012; 35:1629–1637. [PubMed: 23054720]

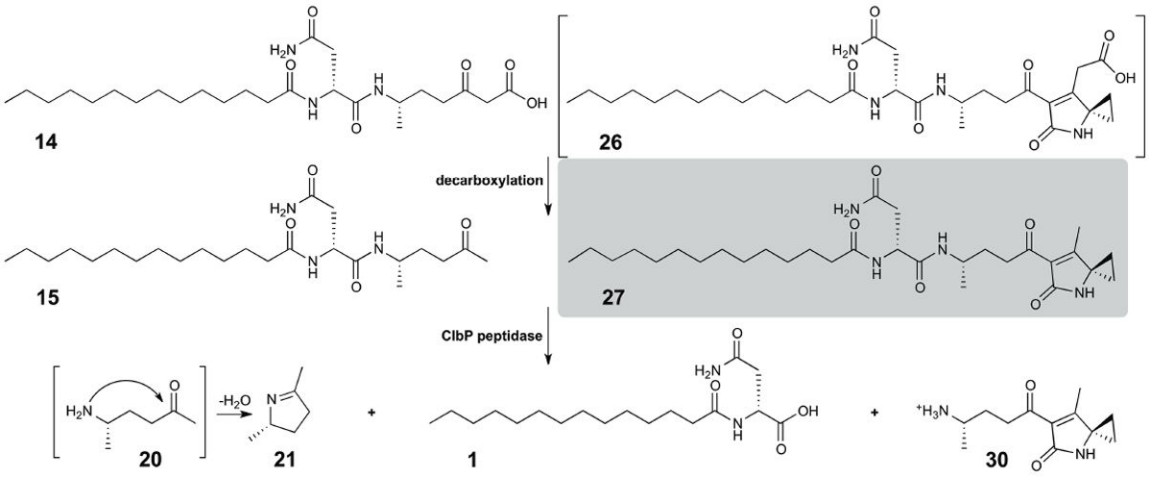

Figure 1. Key colibactin pathway (*clb*)-dependent shunt metabolites

Analogous to the decarboxylation of *clb*-assembly line derailment product **14** to **15**, characterized shunt precolibactin **27** most likely arises from decarboxylation of transient derailment product **26**. ClbP-mediated cleavage appears to be promiscuous in our analysis, leading to N-terminal N-acyl-D-asparagines, such as **1**, and detectable C-terminal products, such as **21** and **30**. Structures are numbered in accordance with increasing biosynthetic complexity as illustrated in Figure 3.

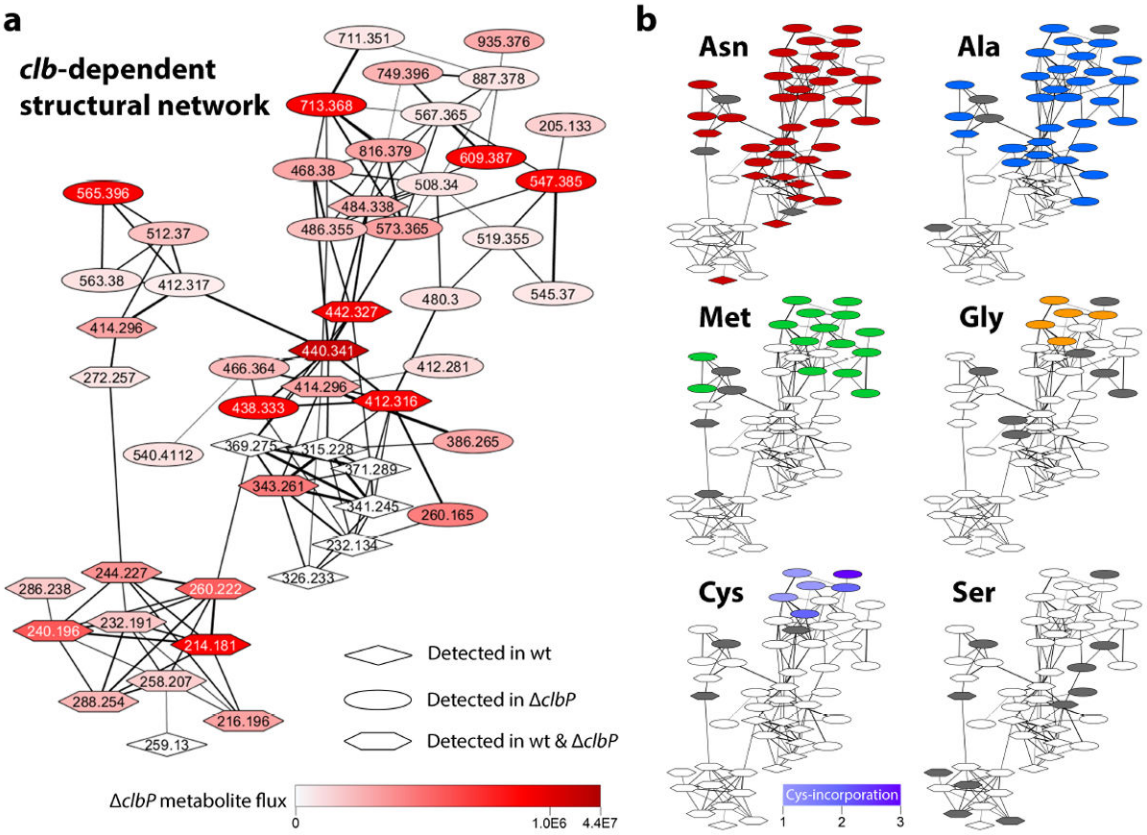

Figure 2. Colibactin pathway (*clb*)-dependent molecular network

(a) *clb*-dependent metabolites detected in IHE3034-derived wildtype (*clb*+ or wt) and *clbP* (Δ*clbP*) mutant cultures. A heat map of ionization intensities for the Δ*clbP* metabolites is shown. (b) System-wide ^{13}C-isotopic incorporations determined from HRMS analysis of L-[U-^{13}C]-amino acid substrates, Asn, Ala, Met, Gly, Cys, and Ser. If a specific amino acid incorporation was detected, the metabolite node was color coded as follows: Asn (red), Ala (blue), Met (green), Gly (orange), and Cys (purple). For L-[$^{13}C_5$]-Met, we only observed $^{13}C_4$ products, indicating amino-butyryl incorporation (green), which were not labeled by [2,2,3,3-D]–ACC. We also observed 1, 2, and 3 Cys incorporations as denoted by the colored map. L-[U-^{13}C]-Ser ($^{13}C_3$ and $^{13}C_2$) labeled metabolites were not detected. Grey nodes were not detected in the labeling experiments. Connectivity strength is represented by the thickness of the lines linking individual nodes.

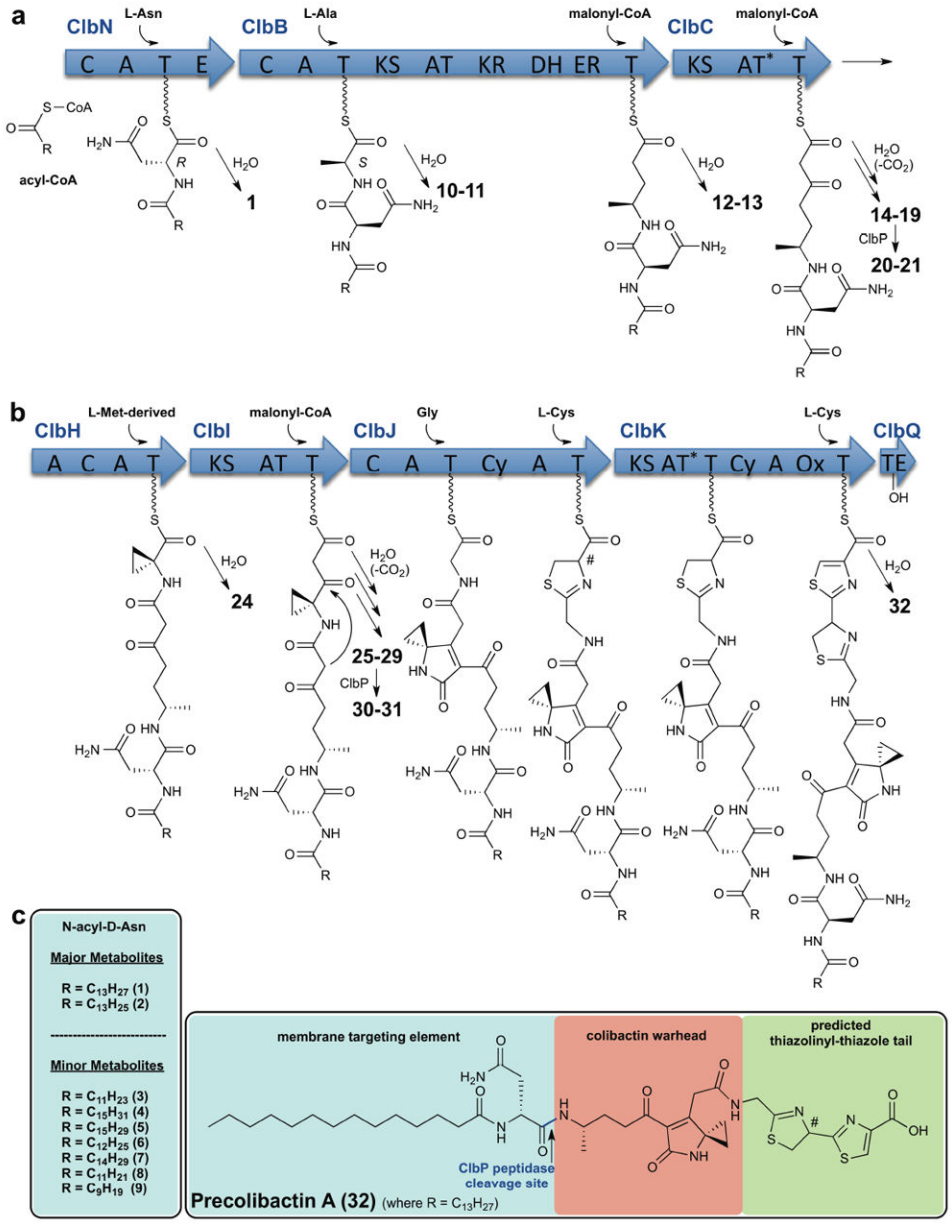

Figure 3. Proposed assembly line biosynthetic model for precolibactin A

(a) Proposed biosynthesis for the *clb* assembly line derailment product **15** and its structurally related shunt metabolites (**1, 10-21**; Supplementary Table S5). Experimentally supported metabolites are indicated with bold numbers, which can result from thioester hydrolysis. Arrows represent NRPS and PKS enzymes with each acronym representing a distinct catalytic domain. Malonyl-CoA and amino acid substrate incorporations, supported by universally ^{13}C-labeled amino acid feeding experiments, are indicated at their proposed cognate carrier proteins (T domains). (b) Predicted and experimentally supported assembly line biosynthesis of advanced derailment products (**24-31**) and precolibactin A (**32**). Our studies indicate that the proposed ClbH-dependent ACC formation is ultimately derived from the aminobutyryl moiety of L-Met. (c) Proposed structure of precolibactin A (**32**) and

detected N-acyl-D-Asn ClbP cleavage products (**1**-**9**; Supplementary Table S5). NRPS and PKS domains: C, condensation; A, adenylation; T, thiolation sequence of acyl- and peptidyl-carrier proteins; E, epimerization; KS, ketosynthase; AT, acyl-transferase; KR, ketoreductase; DH, dehydratase; ER, enoyl-reductase; Cy, condensation/cyclization; Ox, oxidase; TE, thioesterase. *, Denotes evolutionarily deteriorated cis-AT domain in a trans-AT PKS. The bioinformatics predicted thiazoline and thiazole-containing tail was supported by HRMS, MS/MS, and isotopic labeling studies, and its predicted heterocycle order and stereochemistry (#) need further validation.

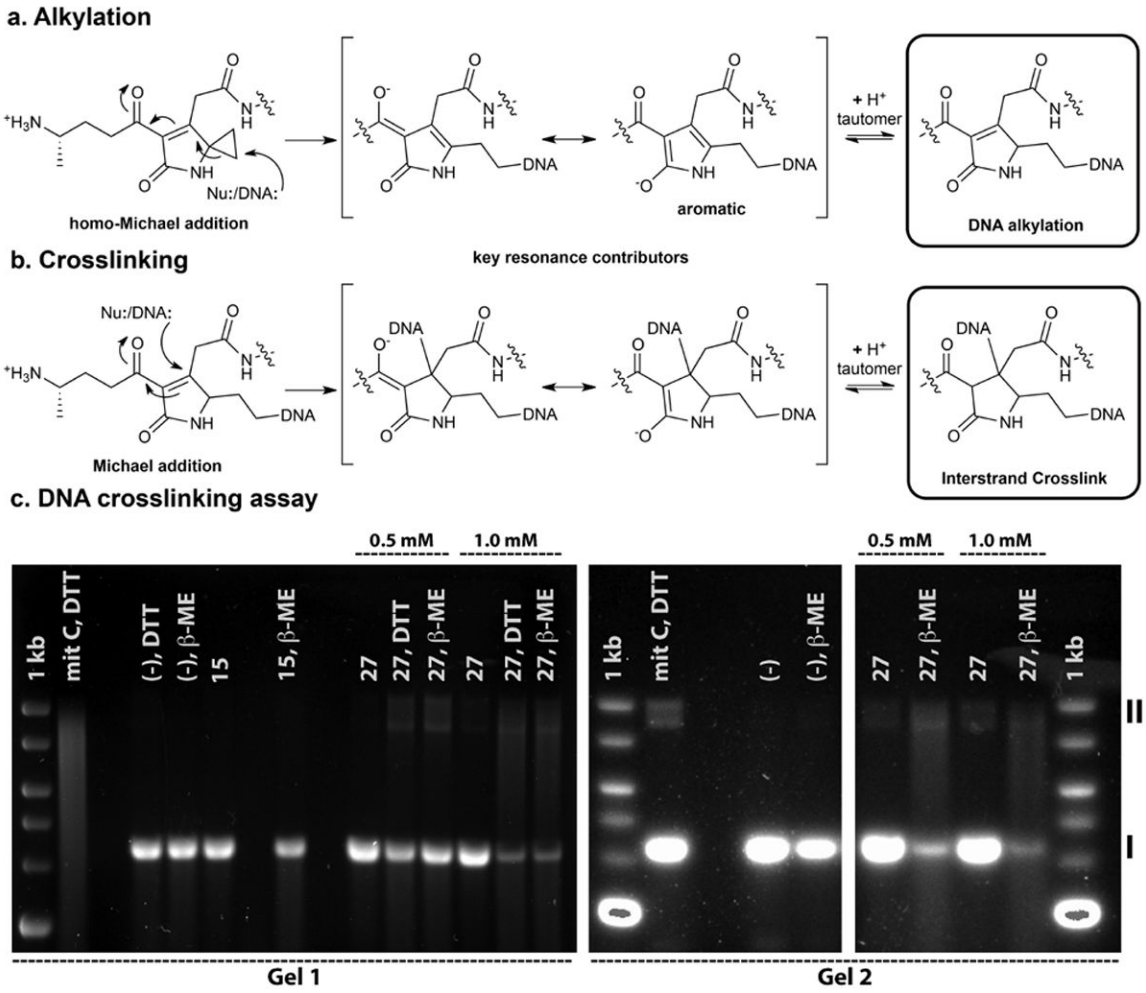

Figure 4. The colibactin warhead crosslinks DNA

(a) DNA alkylation by the warhead is hypothesized to occur through a homo-Michael addition reaction, followed by a (b) pseudo-intramolecular Michael addition reaction, generating a DNA interstrand crosslink. (c) DNA crosslinking was observed using an *Eco*RI-linearized pBR322 plasmid in the presence of **27** (0.5-1.0 mM) with or without reducing agents. DMSO (-) and **15** controls did not lead to detectable activity. Reactions were performed at 37 °C for 20 h (Gel 1). Under these conditions, positive control mitomycin C + DTT caused substantial DNA degradation. Consequently, experiment was repeated with a shorter incubation time and reduced temperature for mitomycin C + DTT (2 h, 20 °C), while the DMSO (-) control and **27** were incubated at 37 °C for 20 h with and without β-ME (Gel 2). mit C, mitomycin C; DTT, dithiothreitol; β-ME, β-mercaptoethanol; I, single-stranded DNA; II, cross-linked DNA.