The clb gene cluster encodes the biosynthesis of metabolites known as precolibactins and colibactins. The clb pathway is found in gut commensal E. coli, and clb metabolites are thought to initiate colorectal cancer via DNA cross-linking. Precolibactin 886 (1) is one of the most complex isolated clb metabolites; it contains a 15-atom macrocycle and an unusual 5-hydroxy-3-oxazoline ring. Here we report confirmation of the structural assignment via a biomimetic synthesis of precolibactin 886 (1) proceeding through the amino alcohol 9. Double oxidation of 9 afforded the unstable α-ketoimine 2 which underwent macrocyclization to precolibactin 886 (1) upon HPLC purification (3% from 9). Studies of the putative precolibactin 886 (1) biosynthetic precursor 2, the model α-ketoimine 25, and the α-dicarbonyl 26 revealed that these compounds are susceptible to nucleophilic rupture of the C36–C37 bond. Moreover, cleavage of 2 produces other known clb metabolites or biosynthetic intermediates. This unexpected reactivity explains the difficulties in isolating full clb metabolites and accounts for the structure of a recently identified colibactin–adenine adduct. The colibactin peptidase ClbP deacylates synthetic precolibactin 886 (1) to form a non-genotoxic pyridone, suggesting precolibactin 886 (1) lies off-path of the major biosynthetic route.
Title: Synthesis and reactivity of precolibactin 886.

Authors: Alan R. Healy, a,b Kevin M. Wernke, a Chung Sub Kim, a,b Nicholas R. Lees, a Jason M. Crawford, a,b,c and Seth B. Herzon a,d

Affiliations:

aDepartment of Chemistry, Yale University, New Haven, Connecticut 06520, United States.
bChemical Biology Institute, Yale University, West Haven, Connecticut 06516, United States.
cDepartment of Microbial Pathogenesis, Yale School of Medicine, New Haven, Connecticut 06536, United States.
dDepartment of Pharmacology, Yale School of Medicine, New Haven, Connecticut 06520, United States.

Email: seth.herzon@yale.edu; jason.crawford@yale.edu

Abstract. The clb gene cluster encodes the biosynthesis of metabolites known as precolibactins and colibactins. The clb pathway is found in gut commensal E. coli, and clb metabolites are thought to initiate colorectal cancer via DNA cross-linking. Precolibactin 886 (1) is one of the most complex isolated clb metabolites; it contains a 15-atom macrocycle and an unusual 5-hydroxy-3-oxazoline ring. Here we report confirmation of the structural assignment via a biomimetic synthesis of precolibactin 886 (1) proceeding through the amino alcohol 9. Double oxidation of 9 afforded the unstable α-ketoimine 2 which underwent macrocyclization to precolibactin 886 (1) upon HPLC purification (3% from 9). Studies of the putative precolibactin 886 (1) biosynthetic precursor 2, the model α-ketoimine 25, and the α-dicarbonyl 26 revealed that these compounds are susceptible to nucleophilic rupture of the C36–C37 bond. Moreover, cleavage of 2 produces other known clb metabolites or biosynthetic intermediates. This unexpected reactivity explains the difficulties in isolating full clb metabolites and accounts for the structure of a recently identified colibactin–adenine adduct. The colibactin peptidase ClbP deacylates synthetic precolibactin 886 (1) to form a non-genotoxic pyridone, suggesting precolibactin 886 (1) lies off-path of the major biosynthetic route.
Main text.

In a seminal report, Oswald and co-workers discovered that strains of commensal *E. coli* containing a biosynthetic gene cluster—referred to as *clb* or *pks*—induce DNA damage in eukaryotic cells.\(^1\) Subsequent studies established that the presence of *clb*\(^+\) *E. coli* is epidemiologically correlated with colorectal cancer (CRC) in humans,\(^2,3\) and a growing body of literature suggests these bacteria are tumorigenic.\(^4-9\) Therefore, understanding the structures and mechanism of action of *clb* metabolites may provide insights into the pathophysiology of CRC.

Precolibactin 886 (1, Fig. 1A)\(^10\) is one of the most complex isolated metabolites derived from the *clb* gene cluster; it accounts for all but three (*clbL, clbO, and clbQ*) biosynthetic enzymes in the *clb* pathway.\(^10,11\) Precolibactin 886 (1) is also the first known *clb* metabolite that contains atoms deriving from an unusual α-aminomalonate building block. The α-aminomalonate originates biosynthetically from serine\(^10,12,13\) and provides the C37 and C38 atoms of precolibactin 886 (1). Bacteria deficient in the genes responsible for the biosynthesis and transfer of this polyketide extender unit (*clbDFEG* and an adenylation (A) domain of *clbH-AI*) are not genotoxic, suggesting that the α-aminomalonate building block is required for cytotoxicity.\(^1,12,13\)

**A.**

![precolibactin-886 (1)](image1)

**B.**

![wild type (ClbP proficient)](image2)

![mutant (ClbP deficient)](image3)

Fig. 1. **A.** Structure of precolibactin 886 (1) and the putative biosynthetic precursor 2. The α-ketoimine 2 served as a synthetic precursor to precolibactin 886 (1). **B.** Genetic studies established that linear intermediates resembling 3 are off-loaded from the *clb* biosynthetic assembly line. In wild-type strains, they undergo ClbP deacylation and transformation to the genotoxic imines 4. In mutant strains lacking functional ClbP, they transform to non-genotoxic pyridones, such as precolibactins A–C (6–8).\(^13-17\) The adenine adduct 5 was derived from DNA that had been exposed to *clb*\(^+\) *E. coli* cultures.
Precolibactin 886 (1) was isolated by Qian, Moore, and co-workers from a large-scale fermentation of a clbP/clbQ double mutant.\textsuperscript{10} Recently, a derivative of precolibactin 886 (1) bearing a terminal oxazole was also disclosed.\textsuperscript{18} Mutation of clbP, a pathway-dedicated serine protease,\textsuperscript{19,20} leads to the persistence of a terminal N-myristoyl-D-Asn residue (blue in 1)\textsuperscript{21-23} and accumulation of more stable biosynthetic precursors referred to as precolibactins. This genetic modification has been widely used to facilitate isolation of clb metabolites. Deactivation of the type-II thioesterase ClbQ,\textsuperscript{24} which is required for genotoxicity,\textsuperscript{1} increased the production of precolibactin 886 (1) 22-fold versus the wild-type strain by reducing the release of advanced intermediates from the biosynthetic assembly line.\textsuperscript{10} Despite the use of this strategically-optimized double mutant, only 2.8 mg of purified precolibactin 886 (1) were obtained from a 1000-L fermentation.\textsuperscript{10} Qian and Moore proposed that the unique 15-atom macrocycle and unusual 5-hydroxy-3-oxazoline ring (green in 1) found in precolibactin 886 (1) arises from the cyclization and enzymatic oxidation of a linear precursor, a proposal supported by intermolecular annulations between α-ketoimines and ketones reported by Hoffman (eq 1).\textsuperscript{25}

\[
\begin{align*}
\text{Ph} & \quad \text{N}_3 \\
& \quad \text{DBU acetone} \\
\rightarrow & \quad \text{Ph} \quad \text{O} \quad \text{N} \quad \text{O} \\
& \quad \text{Ph} \quad \text{Ph} \\
\rightarrow & \quad \text{74%} \\
& \quad \text{Ph} \quad \text{Ph} \\
& \quad \text{NH} \\
& \quad \text{CH}_3, \text{CH}_3 \\
\end{align*}
\]

In wild-type strains possessing functional ClbP, N-deacylation initiates an alternative series of cyclization reactions that ultimately form unsaturated imines resembling 4 (Fig. 1B).\textsuperscript{11,17} Synthetic derivatives of 4 alkylated linearized plasmid DNA by nucleotide addition to the cyclopropane and activated the γH2AX DNA damage response in human cells.\textsuperscript{17,26} A dimer of 4 formed stable DNA cross-links, and a derivative possessing a gem-dimethyl group in place of the cyclopropane did not damage DNA, providing robust support for DNA alkylation by cyclopropane opening.\textsuperscript{16} Further supporting this model, the adenine adduct 5 was subsequently identified in clb\textsuperscript{-} E. coli cultures treated with exogenous DNA.\textsuperscript{27} This product was later established as an N3-substituted adenine adduct, which could also be detected in human cells and in colonic epithelial cells of a mouse model when treated with clb\textsuperscript{+} E. coli.\textsuperscript{28} These studies suggest a basis for clb\textsuperscript{+} E. coli genotoxicity and explain earlier determinations that functional ClbP is required for cytopathic effects.\textsuperscript{1,19-23}

In order to probe its reactivity and biochemistry, we set out to synthesize precolibactin 886 (1) by cyclization of a fully linear precursor. We discovered that cyclization to precolibactin 886 (1) occurs upon HPLC purification of the linear α-ketoimine 2. Additionally, we report the surprising discovery that the C36–C37 bond in 2 is unstable toward nucleophilic cleavage. This degradation generates other known clb metabolites or biosynthetic precursors, and provides an explanation for the difficulties in isolating advanced clb metabolites [and the low yield of precolibactin 886 (1)]. Finally, we show that ClbP deacylation of precolibactin 886 (1) generates a non-genotoxic pyridone, suggesting it lies off of the major colibactin biosynthetic pathway.

**Results.**

**Total synthesis of the linear precursor 2.**

Our retrosynthetic analysis of the linear precursor 2 is shown in Fig. 2. The α-ketoimine function was expected to be unstable and was installed in the final step of the sequence. We envisioned accessing this intermediate by double oxidation of the amino alcohol 9. The linear precursor 9
was readily deconstructed to the β-ketothioesters 10 and 11 and the azido alcohol 12, via sequential N-acylation reactions.

![Fig. 2. Retrosynthetic analysis of the linear precursor 2.](image)

The synthesis of the azido alcohol 12 is shown in Fig. 3. The thiazole 13 is accessible in one step and 74% yield from commercial reagents. Semi-reduction of the ester (di-iso-butylaluminum hydride, DIBAL) provided the aldehyde 14 (88%). Alternatively, removal of the carbamate (hydrochloric acid) followed by condensation with benzophenone imine generated the imine 15 (85%, two steps). This synthetic approach exploits the homology between the thiazoles in the target and expedited material throughput. Silver-catalyzed coupling between 14 and 15, followed by hydrolysis of the resulting addition product (not shown), then provided the amino alcohol 16 as an inconsequential 2:1 mixture of diastereomers (stereochemistry not assigned). Wong diazo transfer to 16 using imidazole sulfonyl azide yielded the azido alcohol 17 (56% over three steps). Treatment of the diazo transfer product 17 with hydrochloric acid in dichloromethane–dioxane then formed the primary ammonium ion 12. Coupling with the β-ketothioester 11 (silver trifluoroacetate, triethylamine) generated the β-ketoamide 18 (85%, two steps). Saponification of the ester (lithium hydroxide, 87%) followed by removal of the tert-butyl carbamate (hydrochloric acid) produced the carboxylic acid 19. A second silver-mediated fragment coupling between 19 and the β-ketothioester 10 provided the linear intermediate 20 (78%, two steps). Staudinger reduction of 20 (trimethylphosphine) proceeded smoothly to generate the target amino alcohol 9 (>99%).
Fig. 3. Synthesis of the amino alcohol 9.

Oxidation and cyclization studies.

With access to the amino alcohol 9, we were poised to probe the oxidation to the α-ketoimine 2 (Fig. 4). In the event, all attempts to effect two-fold oxidation of 9 and isolation of the product provided complex mixtures. We surmised that the α-ketoimine 2, if formed, was undergoing hydration or hydrolysis on attempted isolation.\(^{34,35}\) To circumvent this, we oxidized 9 in methyl sulfoxide-\(d_6\) and monitored the reaction by \(^1\)H NMR spectroscopy. The addition of 2-iodoxybenzoic acid (IBX, 4.00 equiv) to a solution of 9 (10.0 mM) in methyl sulfoxide-\(d_6\) at 23 °C resulted in consumption of the amino alcohol 9 within 30 min. A sharp singlet that we attributed to the \(N\)-H resonance of the α-ketoimine emerged at \(\delta\) 12.02 ppm. However, this resonance under-integrated relative to the C20 methyl group, suggesting partial transformation of the starting material to other unidentified products.

Nonetheless, we attempted to promote macrocyclization of the α-ketoimine 2 formed in situ under thermal, basic, or acidic conditions. However, in each instance we were unable to detect (by \(^1\)H NMR spectroscopy) any signals that could be attributed to precobilactin 886 (1). Under most conditions, we observed hydrolysis of the imine to the α-diketone 22, as evidenced by a gradual disappearance of the \(N\)-H resonance and the appearance of a new signal at 185.8 ppm in the \(^13\)C NMR spectrum (assigned as the C37 carbonyl). The hydrolysis was faster under acidic conditions, as expected. Surprisingly, under basic conditions precobilactin B (7) was formed, by an
unexpected hydrolytic cleavage of the C36–C37 bond, followed by cyclodehydration.\textsuperscript{16} Attempts to promote the cyclization thermally led to decomposition.

In light of these difficulties, we reasoned that perhaps the macrocyclization to precolibactin 886 (1) occurred during HPLC analysis and purification in the isolation studies.\textsuperscript{10} To our knowledge, all fermentations of clb\textsuperscript{+} E. coli have been analyzed by HPLC/MS; therefore, cyclization of the α-ketoimine 2 during analytical or preparative HPLC could obfuscate the presence of the linear precursor 2 in the fermentation broth. To test this, we directly injected solutions of the α-ketoimine 2 onto a semipreparative reverse phase HPLC system. We found that under these conditions cyclization did occur, leading to the isolation of precolibactin 886 (1). From 10 mg of 2 we obtained \~0.3 mg of purified precolibactin 886 (1, \~3\%). \textsuperscript{1}H, \textsuperscript{13}C, HRMS, and tandem MS data of synthetic precolibactin 886 (1) were in full agreement with the material isolated from bacterial cultures\textsuperscript{10} (Table S1, Fig. S2). The C36 stereocenter was formed as a \~1:1:9 mixture of diastereomers in the ring closure (stereochemistry not assigned). The dr of natural precolibactin 886 (1), which also formed non-stereoselectively, was not reported.\textsuperscript{10} Importantly, the diastereotopic protons H32 in the linear precursor 2 are nearly magnetically-equivalent and appear as an apparent doublet at \(\delta 4.52\) (\(J = 6.0\) Hz). Formation of the macrocycle breaks this degeneracy leading to a distinct doublet of doublets resonating at \(\delta 4.27\) (\(J = 17.2, 5.0\) Hz) and \(\delta 4.79\) (\(J = 17.2, 7.1\) Hz; DMSO-\(d_6\), 600 MHz). These experimental results suggest precolibactin 886 (1) may derive from cyclization of an unstable linear precursor during HPLC analysis and purification. We did not observe cyclization of the amino alcohol 9 or the monooxidation product 21, suggesting the imine is formed prior to macrocyclization in the bacterial extracts. Finally, precolibactin 886 (1) could also be accessed in comparable yield from the azido alcohol 20. Oxidation of 20 with IBX formed the α-ketoimine 2 directly, presumably via loss of dinitrogen from the intermediate α-azidoketone.\textsuperscript{30} The generation of precolibactin 886 (1) from the azido alcohol 20 provides further support for the intermediacy of the α-ketoimine 2.
On the chemical instability of the α-ketoimine 2.

The low yield of precolibactin 886 (1) and the unexpected generation of precolibactin B (7) from 9 motivated us to study the reactivity of the truncated α-ketoimine model 25 and its hydrolysis product, the α-diketone 26, in detail (Fig. 5A and Table 1). The amino alcohol 23 (see Fig. S1 and the Supporting Information for synthesis) was efficiently oxidized to the α-ketoimine 25 by treatment with excess Swern reagent (2.17 equiv; attempts to effect Swern oxidation of 9 were unsuccessful). 1H NMR analysis indicated that the sample was >95% pure (nominally 93% yield). The imine N–H resonance appeared at δ 12.07 ppm in methyl sulfoxide-d$_6$, in good agreement with the fully elongated α-ketoimine 2 (δ 12.02 ppm). As expected, the imine readily underwent hydrolysis. For example, the α-diketone 26 was formed if the α-ketoimine 25 was exposed to aqueous conditions for an extended time during work-up. Alternatively, the addition of silica gel to solutions of the α-ketoimine 25 formed the α-diketone 26 (38% from 23). The α-diketone 26 was also independently synthesized by Wong diazo transfer$^{31,32}$ to 23 using imidazole sulfonyl azide$^{33}$ (56%), followed by oxidation with the Dess–Martin periodinane$^{36}$ (2.50 equiv, 54%). The C37 carbon atom of 26 was observed at δ 184.6 in methyl sulfoxide-d$_6$, in good agreement with the fully elongated α-diketone 22 (δ 185.8).
Studies of the oxidation of the \( N\)-(tert-butoxycarbonyl)-1,2-amino alcohol \( 27 \) provided unexpected results (Fig. 5B). Oxidation with an excess of the Dess–Martin periodinane, under Swern conditions, or using \( \text{SO}_3\cdot\text{pyridine} \) formed the product of two-fold oxidation, the hemiaminal \( 29 \). Attempted oxidation of \( 27 \) using 0.9 equiv of the Dess–Martin periodinane, followed by immediate purification by flash-column chromatography, generated the expected \( \alpha \)-amino ketone \( 28 \), but the product spontaneously oxidized to \( 29 \) on standing. These data suggest intermediates such as \( 28 \) undergo facile oxidation and support a pathway for the spontaneous generation of the \( \alpha \)-ketoimine in the precolibactin 886 precursor 2 from its expected \( \alpha \)-aminoketone substrate.

Fig. 5. A. Synthesis of the \( \alpha \)-ketoimine \( 25 \) and the \( \alpha \)-dicarbonyl \( 26 \). B. Oxidation of the \( N\)-(tert-butoxycarbonyl)-1,2-amino alcohol \( 27 \).

With access to analytically-pure samples of the \( \alpha \)-ketomine \( 25 \) and the \( \alpha \)-diketone \( 26 \) we were in position to study their chemistry in detail (Table 1). We found, surprisingly, that these compounds were unstable toward nucleophiles and readily underwent cleavage of the C36–C37 bond [precolibactin 886 (1) numbering]. For example, dissolution of \( 25 \) in anhydrous methanol (64 mM) followed by the addition of sodium bicarbonate (9.0 equiv) resulted in rapid C36–C37 bond cleavage and formation of the ester \( 30a \) (42%), as well as the carboxamide \( 30d \) derived from the right-hand thiazole. Bond cleavage even occurred in the absence of base, albeit more slowly: the half-life of the \( \alpha \)-ketomine \( 25 \) in neutral, distilled methanol-\( d_1 \) was ~12 h at 24 °C (\(^1\)H NMR analysis). Alternatively, the \( \alpha \)-ketomine \( 25 \) could be hydrolyzed to the carboxylic acid \( 30b \) by treatment with saturated aqueous sodium bicarbonate solution in tetrahydrofuran at 23 °C (51%) and the amide \( 30c \) was formed when the \( \alpha \)-ketomine \( 25 \) was treated with pyrrolidine (14 equiv) in dichloromethane (36 mM in \( 25 \) at 23 °C (36%). In the case of entries 2 and 3 the products derived from the right-hand thiazole were not readily identified. The \( \alpha \)-diketone \( 26 \) underwent...
parallel transformations forming the methyl esters 30a and 30e, the carboxylic acids 30b and 30f, and the amide 30c.

Table 1. Nucleophilic cleavage of the α-ketoimine 25 and the α-diketone 26.

| entry | substrate | nucleophile | product(s) (yield) |
|-------|-----------|-------------|-------------------|
| 1     | 25        | methanol\(^b\) | 30a (42%) + 30d (~42%)\(^c\) |
| 2     | 25        | water\(^c\)  | 30b (51%)         |
| 3     | 25        | pyrrolidine\(^d\) | 30c (36%)         |
| 4     | 26        | methanol\(^b\) | 30a (45%) + 30e (41%) |
| 5     | 26        | water\(^c\)  | 30b (49%) + 30f (25%) |
| 6     | 26        | pyrrolidine\(^d\) | 30c (41%)         |

\(^a\)Isolated yield after purification by flash-column chromatography. \(^b\)25 or 26 (1 equiv), NaHCO\(_3\) (9.0 equiv), methanol, 23 °C, 48 h (for 25 and 26). \(^c\)25 or 26 (1 equiv), tetrahydrofuran–saturated aqueous sodium bicarbonate solution (1:1, v/v), 23 °C, 48 h (for 25), 72 h (for 26). \(^d\)25 or 26 (1 equiv), pyrrolidine (14 equiv), dichloromethane, 23 °C, 48 h (for 25), 72 h (for 26). \(^e\)Estimated by HPLC analysis of the product mixture.

In light of this discovery, we re-examined the products resulting from oxidation of the amino alcohol 9 by LC/MS and tandem MS. As expected, we detected hydrolysis of the α-ketoimine 2 to its corresponding α-diketone 22 (Fig. 6A, Fig. S4). We also detected hydrolytic cleavage of the C36–C37 bond to generate the linear precursor to precolibactin B (32). As previously reported in a synthetic study,\(^16\) the linear precursor 32 underwent double ring closure to provide precolibactin B (7). The addition of L-cysteine (7.5 equiv based on starting amino alcohol 9) resulted in generation of 33, the linear precursor to precolibactin A (6, Fig. 1B). The structures of 31–33 and precolibactin B (7) were confirmed by tandem MS and LC/MS co-injection with the bacterial extracts derived from a clb\(^+\) ΔclbP E. coli culture (Figs. S3, S5–S7). We also detected competitive hydrolysis of the C23–C24 bond to generate the acylasparagine derivative 31, which has previously been detected in clb cultures.\(^23\)
To establish precolibactin 886 (1) as a potential substrate of the promiscuous colibactin peptidase ClbP, we supplemented 1 to *E. coli* expressing functional ClbP (pPEB018) versus a control strain lacking ClbP (pBAD18) (Fig. 6B). The expected deacylation product 34 was formed in the ClbP expressing strain but it was undetectable in the ClbP-deficient strain. The concentration of 34 increased in a time-dependent fashion. However, we determined that the macrocycle 34 was unstable under the cultivation conditions. As it was produced, 34 underwent conversion to the pyridone 35. The structure of 35 formed in this experiment was confirmed by LC/MS co-injection and tandem MS with a synthetic standard (Fig. S8). The pyridone 35 derives from C36–C37 bond cleavage and double cyclodehydration.
Discussion.

Precobilactin 886 (I) is one of the most complex isolated metabolites derived from the clb gene cluster. It is the first reported clb metabolite that contains two carbon atoms (C37, C38, see Fig. 1A) derived from an α-aminomalonate building block. This unit is biosynthesized from serine\textsuperscript{10,12,13} by the enzymes ClbDEF and ClbH-A\textsubscript{1}. Deletion of any of these enzymes abolishes the cytopathic effects of clb\textsuperscript{+} E. coli suggesting that this unit is necessary for the observed genotoxicity.\textsuperscript{1,12}

The prevailing model for colibactin biosynthesis and genotoxicity involves off-loading of fully linear products (see Fig. 1B) from the biosynthetic assembly line. In wild-type strains these intermediates are deacylated in the periplasm by ClbP and transform, via spontaneous cyclization reactions, to genotoxic unsaturated imines (see Fig. 1B). In mutant strains lacking functional ClbP, artificial cyclization reactions prevail resulting in the generation of non-genotoxic pyridone-containing compounds (e.g., 6–8, Fig. 1B).\textsuperscript{7,17} The studies reported herein allow us to expand this model to accommodate the structure of precobilactin 886 (1, Fig. 7A). Our data suggest a biosynthetic pathway for 1 involving off-loading of the linear α-amino ketone 36 from the assembly line. It has previously been suggested that 36 is oxidized by the enzyme ClbK;\textsuperscript{10} however, our studies of the N-(tert-butoxycarbonyl)-1,2-amino alcohol 27 (Fig. 5B) raise the possibility that oxidation occurs spontaneously. Irrespective of how it is formed, the resulting α-ketoimine 2 then partitions among several reaction pathways. We speculate that in wild-type or mutant strains hydrolysis of the α-ketoimine to the α-dicarbonyl 21 is rapid and predominates. In mutant strains lacking functional ClbP, the hydrolysis product 21 degrades by cleavage of the C23–C24 bond and/or cleavage of the C36–C37 bond. The latter pathway generates metabolites such as precobilactin A (6)\textsuperscript{14,16} and precobilactin B (7)\textsuperscript{15} that have previously been detected in ΔclbP cultures. In wild-type strains, deacylation followed by cyclocondensation generates unsaturated imines 4 structurally analogous to those formed from simpler clb metabolites. With the difficulties we encountered in inducing cyclization of the linear α-ketoimine 2 to precobilactin 886 (1), and the very low yield of the latter from prior isolation studies,\textsuperscript{10} we suggest that macrocyclization is a minor pathway. This is also consistent with our observation that ClbP cleavage of 1 leads to the non-genotoxic pyridone breakdown product 35 as the major detectable product in cell culture.
The facile C36–C37 bond cleavage we have observed has several important implications. First the adenine adduct 5 (Fig. 1B) was characterized in the digestion mixtures of DNA that had suffered interstrand cross-links after exposure to clb+E. coli. While it is conceivable that 5 derives from more advanced monoalkylation products, based on our reactivity data we believe that 5 is generated by degradation of a cross-linked product. Additionally, cleavage of this bond provides
an explanation for the difficulties encountered in identifying fully-elaborated clb metabolites. The remaining biosynthetically unaccounted clb enzymes ClbL and ClbO, which are required for genotoxicity, append substituents to the terminal carboxylic acid end of advanced precolibactins. C36–C37 bond cleavage would effectively remove these substituents from advanced (pre)colibactins. A possible mechanism for the hydrolytic cleavage of the C36–C37 bond in the α-ketoimine 2 is shown in Fig. 7B. Formation of the hemiaminal 37, followed by rearrangement would generate the ester 39. Hydrolysis of the ester would form 32 and liberate the enol 40. Oxidation of the enol 40 would generate the observed carboxamide 30d. This mechanism is reminiscent of thiazolium ion reactivity, but employs a thiazole rather than a thiazolium ion as the nascent leaving group. Related pathways for the cleavage of 1,2-dicarbonyl compounds by thiazolium ions have been proposed.

**Conclusion.** In sum, we have reported the first total synthesis of precolibactin 886 (1), one of the most advanced clb products. Our studies confirm the structural assignment of this unusual metabolite, and suggest it may be an artifact deriving from analysis and purification of the bacterial extracts. Moreover, our data have revealed unexpected electrophilic reactivity of the α-keto imine of precolibactin 886 (1); the mild cleavage of the C36–C37 bond we have observed provides an explanation for the difficulties in isolating advanced clb metabolites and accounts for the structures of recently isolated colibactin–adenine adducts. Finally, our synthetic studies suggest the unusual α-ketoimine may generate by spontaneous oxidation of an unstable α-amino ketone. This work defines the unexpected reactivities of advanced clb metabolites, provide a path for characterizing the structures of DNA crosslinked products, and explain the underlying cellular and chemical degradation pathways of (pre)colibactins.
Acknowledgement. We thank Professor David Spiegel for helpful discussions regarding the mechanism of hydrolytic cleavage of 2. Financial support from the National Institutes of Health (R01GM110506 to S.B.H., IDP2-CA186575 to J.M.C., R01CA215553 to S.B.H. and J.M.C.), the Chemistry Biology Interface Training Program (T32GM067543 to K.M.W.), the Charles H. Revson foundation (postdoctoral fellowship to A.R.H.) and Yale University is gratefully acknowledged.

References.

1 Nougayrède, J.-P. et al. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* **313**, 848-851 (2006).
2 Arthur, J. C. et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338**, 120-123 (2012).
3 Buc, E. et al. High prevalence of mucosa-associated *e. Coli* producing cyclomodulin and genotoxin in colon cancer. *PLoS One* **8**, e56964 (2013).
4 Trautman, E. P. & Crawford, J. M. Linking biosynthetic gene clusters to their metabolites via pathway-targeted molecular networking. *Curr. Top. Med. Chem.* **16**, 1-12 (2015).
5 Balskus, E. P. Colibactin: Understanding an elusive gut bacterial genotoxin. *Nat. Prod. Rep.* **32**, 1534-1540 (2015).
6 Taieb, F., Pett, C., Nougayrede, J. P. & Oswald, E. The enterobacterial genotoxins: Cytolethal distending toxin and colibactin. *EcoSal Plus* **7**, doi:10.1128/ecosalplus.ESP-0008-2016 (2016).
7 Healy, A. R. & Herzon, S. B. Molecular basis of gut microbiome-associated colorectal cancer: A synthetic perspective. *J. Am. Chem. Soc.* **139**, 14817-14824 (2017).
8 Faïs, T., Delmas, J., Barnich, N., Bonnet, R. & Dalmasso, G. Colibactin: More than a new bacterial toxin. *Toxins* **10**, 151 (2018).
9 Dejea, C. M. et al. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* **359**, 592 (2018).
10 Li, Z. R. et al. Divergent biosynthesis yields a cytotoxic aminomalonate-containing precolibactin. *Nat. Chem. Biol.* **12**, 773-775 (2016).
11 Trautman, E. P., Healy, A. R., Shine, E. E., Herzon, S. B. & Crawford, J. M. Domain-targeted metabolomics delineates the heterocycle assembly steps of colibactin biosynthesis. *J. Am. Chem. Soc.* **139**, 4195-4201 (2017).
12 Brachmann, A. O. et al. Colibactin biosynthesis and biological activity depend on the rare aminomalonyl polyketide precursor. *Chem. Commun.* **51**, 13138-13141 (2015).
13 Zha, L., Wilson, M. R., Brotherton, C. A. & Balskus, E. P. Characterization of polyketide synthase machinery from the pks island facilitates isolation of a candidate precolibactin. *ACS Chem. Biol.* **11**, 1287-1295 (2016).
14 Vizcaino, M. I. & Crawford, J. M. The colibactin warhead crosslinks DNA. *Nat. Chem.* **7**, 411-417 (2015).
15 Li, Z. R. et al. Critical intermediates reveal new biosynthetic events in the enigmatic colibactin pathway. *Chembiochem* **16**, 1715-1719 (2015).
16 Healy, A. R., Vizcaino, M. I., Crawford, J. M. & Herzon, S. B. Convergent and modular synthesis of candidate precolibactins. Structural revision of precolibactin a. *J. Am. Chem. Soc.* **138**, 5426-5432 (2016).
Healy, A. R., Nikolayevskiy, H., Patel, J. R., Crawford, J. M. & Herzon, S. B. A mechanistic model for colibactin-induced genotoxicity. *J. Am. Chem. Soc.* **138**, 15563-15570 (2016).

Li, Z.-R. *et al.* Macrocyclic colibactin induces DNA double-strand breaks via copper-mediated oxidative cleavage. *bioRxiv*, 530204 (2019).

Dubois, D. *et al.* Clbp is a prototype of a peptidase subgroup involved in biosynthesis of nonribosomal peptides. *J. Biol. Chem.* **286**, 35562-35570 (2011).

Cougnoux, A. *et al.* Analysis of structure-function relationships in the colibactin-maturating enzyme clbp. *J. Mol. Biol.* **424**, 203-214 (2012).

Bian, X. *et al.* In vivo evidence for a prodrug activation mechanism during colibactin maturation. *Chembiochem* **14**, 1194-1197 (2013).

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

Guntaka, N. S., Healy, A. R., Crawford, J. M., Herzon, S. B. & Bruner, S. D. Structure and functional analysis of clbq, an unusual intermediate-releasing thioesterase from the colibactin biosynthetic pathway. *ACS Chem Biol* **12**, 2598-2608 (2017).

Patonay, T. & Hoffman, R. V. Base-promoted reactions of α-azido ketones with aldehydes and ketones: A novel entry to α-azido-β-hydroxy ketones and 2,5-dihydro-5-hydroxyoxazoles. *J. Org. Chem.* **60**, 2368-2377 (1995).

Shine, E. E. *et al.* Model colibactins exhibit human cell genotoxicity in the absence of host bacteria. *ACS Chem. Biol.* **13**, 3286-3293 (2018).

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

Wilson, M. R. *et al.* The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **363**, eaar7785 (2019).

Seashore-Ludlow, B., Torssell, S. & Somfai, P. Addition of azomethine ylides to aldehydes: Mechanistic dichotomy of differentially substituted α-imino esters. *Eur. J. Org. Chem.* **2010**, 3927-3933 (2010).

Lou, S., Ramirez, A. & Conlon, D. A. Catalytic syn-selective direct aldol reactions of benzophenone glycine imine with aromatic, heteroaromatic and aliphatic aldehydes. *Adv. Synth. Catal.* **357**, 28-34 (2015).

Alper, P. B., Hung, S.-C. & Wong, C.-H. Metal catalyzed diazo transfer for the synthesis of azides from amines. *Tetrahedron Lett.* **37**, 6029-6032 (1996).

Nyffeler, P. T., Liang, C.-H., Koeller, K. M. & Wong, C.-H. The chemistry of amine–azide interconversion: Catalytic diazotransfer and regioselective azide reduction. *J. Am. Chem. Soc.* **124**, 10773-10778 (2002).

Goddard-Borger, E. D. & Stick, R. V. An efficient, inexpensive, and shelf-stable diazotransfer reagent: Imidazole-1-sulfonyl azide hydrochloride. *Org. Lett.* **9**, 3797-3800 (2007).

Edwards, O. E. & Purushothaman, K. K. Some reactions of alicyclic α-azidoketones. *Can. J. Chem.* **42**, 712-716 (1964).
Faiz, S. et al. Synthesis and consecutive reactions of alpha-azido ketones: A review. 
_Molecules_ **20**, 14699-14745 (2015).

Dess, D. B. & Martin, J. C. A useful 12-i-5 triacetoxyperiodinane (the dess-martin periodinane) for the selective oxidation of primary or secondary alcohols and a variety of related 12-i-5 species. _J. Am. Chem. Soc._ **113**, 7277-7287 (1991).

White, M. J. & Leeper, F. J. Kinetics of the thiazolium ion-catalyzed benzoin condensation. _J. Org. Chem._ **66**, 5124-5131 (2001).

Kim, T. & Spiegel, D. A. The unique reactivity of n-phenacyl-derived thiazolium salts toward α-dicarbonyl compounds. _Rejuvenation Res._ **16**, 43-50 (2012).
