ACCOUNTS of chemical research

pubs.acs.org/accounts Article

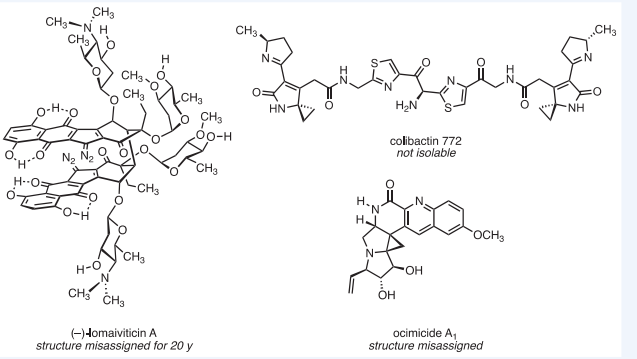
Structure Elucidation of Secondary Metabolites: Current Frontiers and Lingering Pitfalls

Mikaela DiBello, Alan R. Healy, Herman Nikolayevskiy, Zhi Xu, and Seth B. Herzon*

Cite This: *Acc. Chem. Res.* 2023, 56, 1656−1668 Read Online

ACCESS | Metrics & More | Article Recommendations

CONSPECTUS: Analytical methods allow for the structure determination of submilligram quantities of complex secondary metabolites. This has been driven in large part by advances in NMR spectroscopic capabilities, including access to high-field magnets equipped with cryogenic probes. Experimental NMR spectroscopy may now be complemented by remarkably accurate carbon-13 NMR calculations using state-of-the-art DFT software packages. Additionally, microED analysis stands to have a profound effect on structure elucidation by providing X-ray-like images of microcrystalline samples of analytes. Nonetheless, lingering pitfalls in structure elucidation remain, particularly for isolates that are unstable or highly oxidized. In this Account, we discuss three projects from our laboratory that highlight nonoverlapping challenges to the field, with implications for chemical, synthetic, and mechanism of action studies. We first discuss the lomaiviticins, complex unsaturated polyketide natural products disclosed in 2001. The original structures were derived from NMR, HRMS, UV−vis, and IR analysis. Owing to the synthetic challenges presented by their structures and the absence of X-ray crystallographic data, the structure assignments remained untested for nearly two decades. In 2021, the Nelson group at Caltech carried out microED analysis of (−)-lomaiviticin C, leading to the startling discovery that the original structure assignment of the lomaiviticins was incorrect. Acquisition of higher-field (800 MHz ^{1}H, cold probe) NMR data as well as DFT calculations provided insights into the basis for the original misassignment and lent further support to the new structure identified by microED. Reanalysis of the 2001 data set reveals that the two structure assignments are nearly indistinguishable, underscoring the limitations of NMR-based characterization. We then discuss the structure elucidation of colibactin, a complex, nonisolable microbiome metabolite implicated in colorectal cancer. The colibactin biosynthetic gene cluster was detected in 2006, but owing to colibactin's instability and low levels of production, it could not be isolated or characterized. We used a combination of chemical synthesis, mechanism of action studies, and biosynthetic analysis to identify the substructures in colibactin. These studies, coupled with isotope labeling and tandem MS analysis of colibactin-derived DNA interstrand cross-links, ultimately led to a structure assignment for the metabolite. We then discuss the ocimicides, plant secondary metabolites that were studied as agents against drug-resistant *P. falciparum*. We synthesized the core structure of the ocimicides and found significant discrepancies between our experimental NMR spectroscopic data and that reported for the natural products. We determined the theoretical carbon-13 NMR shifts for 32 diastereomers of the ocimicides. These studies indicated that a revision of the connectivity of the metabolites is likely needed. We end with some thoughts on the frontiers of secondary metabolite structure determination. As modern NMR computational methods are straightforward to execute, we advocate for their systematic use in validating the assignments of novel secondary metabolites.



KEY REFERENCES

- Nikolayevskiy, H.; Moe Tun, M. K.; Rablen, P. R.; Ben Mamoun, C.; Herzon, S. B. A complex stereochemical relay approach to the antimalarial alkaloid ocimicide A1. Evidence for a structural revision. *Chem. Sci.* **2017**, *8*, 4867.[1] This manuscript describes the synthesis of the core structure of the ocimicides and computational investigations of their structures that suggest they are misassigned.
- Healy, A. R.; Wernke, K. M.; Kim, C. S.; Lees, N. R.; Crawford, J. M.; Herzon, S. B. Synthesis and reactivity of precolibactin 886. *Nat. Chem.* **2019**, *11*, 890.[2] This

Received: March 24, 2023
Published: May 23, 2023

https://doi.org/10.1021/acs.accounts.3c00183
Acc. Chem. Res. 2023, 56, 1656−1668

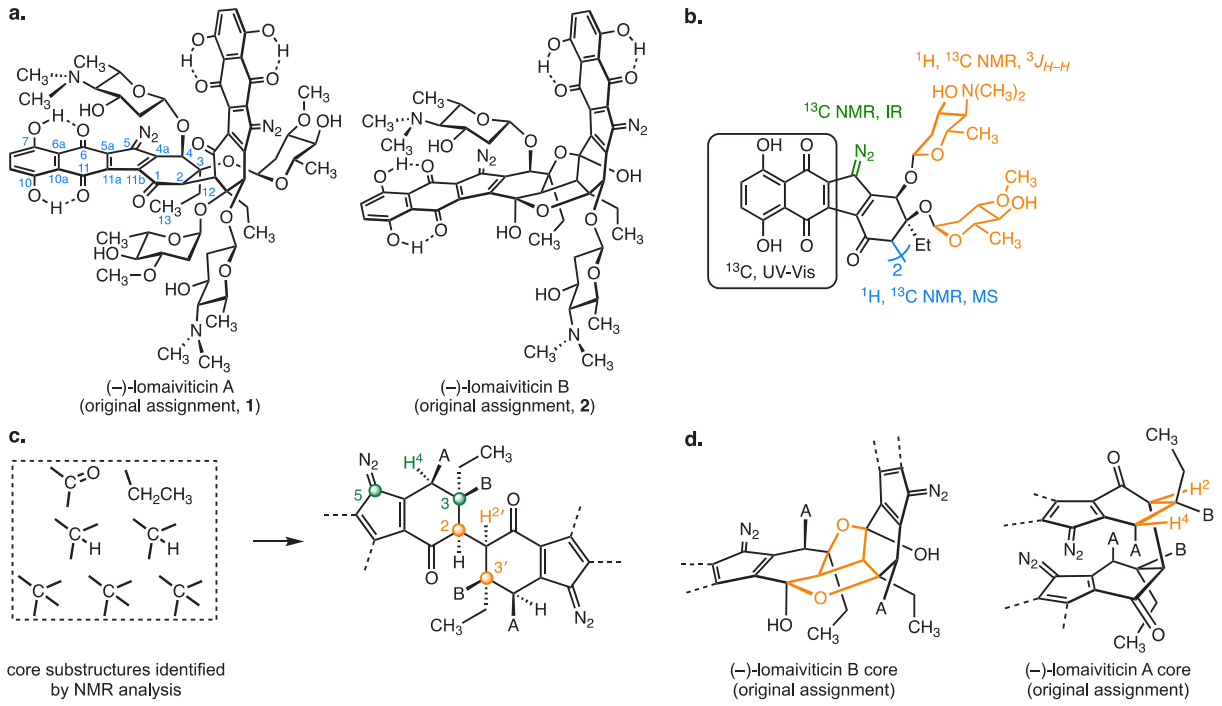


Figure 1. a. Structures **1** and **2** were advanced for (−)-lomaiviticins A and B in 2001. b. Spectroscopic methods used to identify the substructures in (−)-lomaiviticins A and B. c. Core fragments identified by NMR analysis (^{13}C, HMBC, and HSQC) and their assembly into structure **1**. Key HMBC correlations are shown in orange and green (Note: due to the C_2 symmetry of (−)-lomaiviticin A, these correlations apply to both cyclohexenone rings but are depicted on separate rings for clarity.) d. Connectivity and relative configuration of the cyclohexenone derived from the bis(furanol) structure of (−)-lomaiviticin B and putative long-range (W-plane) coupling between H2 and H4 in (−)-lomaiviticin A. This correlation was assigned as a $^{4}J_{H2-H4}$ (W) coupling. A and B represent the oxygen-linked *β*-*N*,*N*-dimethyl-L-pyrrolosamine and *α*-L-oleandrose residues, respectively (see the text).

manuscript describes the synthesis of precolibactin 886 and establishes that advanced *clb* metabolites undergo unexpected carbon−carbon bond cleavage at the two-carbon linker between the thiazole rings. This reactivity explains the difficulties in isolating colibactin directly.

- Xue, M.; Kim, C. S.; Healy, A. R.; Wernke, K. M.; Wang, Z.; Frischling, M. C.; Shine, E. E.; Wang, W.; Herzon, S. B.; Crawford, J. M. Structure elucidation of colibactin and its DNA cross-links. *Science* **2019**, *365*, eaax2685.[3] This manuscript describes the indirect characterization of colibactin 770 by the characterization of DNA interstrand cross-links formed when cultures of *clb*$^{+}$ *E. coli* are treated with exogenous DNA.
- Kim, L. J.; Xue, M.; Li, X.; Xu, Z.; Paulson, E.; Mercado, B.; Nelson, H. M.; Herzon, S. B. Structure revision of the lomaiviticins. *J. Am. Chem. Soc.* **2021**, *143*, 6578.[4] This manuscript describes structural studies of the lomaiviticins by microED, high-field NMR, and computational methods, leading to the revision discussed here.

INTRODUCTION

Methods for the structure elucidation of complex secondary metabolites have advanced considerably. The availability of high-field NMR magnets, cryogenic probes, and advanced two-dimensional pulse sequences has positioned NMR spectroscopy as the primary mode of structure determination.[5] Nonetheless, NMR-guided structure elucidation remains imperfect, especially for metabolites that are highly oxidized (e.g., low H:C ratio). Additionally, NMR-based structure elucidation presupposes that the metabolites of interest are stable to fractionation and are produced in quantities sufficient for isolation. Advances in genome-based approaches to metabolite discovery have made clear that many biosynthetic gene clusters (BGCs) are poorly expressed under laboratory conditions, if at all. Additionally, many secondary metabolites are unstable and not amenable to the multiple rounds of purification required by the classical "grind-and-find" approach.[6]

Here we discuss three projects involving the attempted synthesis and study of secondary metabolites for which the structure assignment was incorrect (lomaiviticins and ocimicides) or for which a structure assignment was not possible based on available data (colibactin). For two of these projects (colibactin and lomaiviticins), we show how we were able to overcome these obstacles to arrive at the correct structure assignment. The correct structure of the ocimicides remains unknown.

LOMAIVITICINS

In 2001, researchers at Wyeth Pharmaceuticals and the University of Utah described two cytotoxic dimeric metabolites from a strain of *M. lomaivitiensis*[7] (later classified as *S. pacifica*).[8] The more abundant isolate, (−)-lomaiviticin A, was advanced as structure **1** (Figure 1a). A second isolate, (−)-lomaiviticin B, was depicted as structure **2**. The lomaiviticins possess diazotetrahydro[*b*]benzofluorene (diazofluorene) functional groups and two or four deoxyglycoside residues. Formally, structure **2** derives from the loss of the oleandrose residues of **1** and the addition of the resulting C3/C3′ hydroxyl substituents to the C1/C1′ carbonyl groups. (−)-Lomaiviticin A displayed antiproliferative activities in the low nM−pM range and antimicrobial minimum inhibitory

https://doi.org/10.1021/acs.accounts.3c00183
Acc. Chem. Res. 2023, 56, 1656−1668

concentrations (MICs) in the ng/spot range. The antiproliferative activity of (−)-lomaiviticin B was not evaluated (owing to its lower abundance), but it also displayed potent antimicrobial effects. The cytotoxic effects of (−)-lomaiviticin A originate from the induction of double-strand breaks in DNA.[9]

The structure elucidation of the lomaiviticins was difficult owing to their C_2 symmetry and unsaturation (Ω = 30−32).[7] High-resolution mass spectrometry established the molecular formula of (−)-lomaiviticin A as $C_{68}H_{80}N_6O_{24}$. The observation of half the expected carbon and proton NMR signals pointed toward a C_2-symmetric structure. The connectivity and relative configuration relied on UV−vis and infrared spectroscopy and 1- and 2-dimensional NMR experiments (Figure 1b). The 5,8-dihydroxy-1,4-naphthoquinone was inferred based upon ^{1}H and ^{13}C NMR analysis, HMBC correlations, and a UV absorption at 525 nm, which is characteristic of these structures. The diazo cyclopentadiene was supported by observation of a carbon-13 resonance at δ 78.8 ppm (diazo carbon atom, C5/5′) and a strong infrared stretch at 2148 cm^{-1}, diagnostic features of this functional group found in the monomeric diazofluorenes known as the kinamycins.[10] The deoxyglycosides β-*N*,*N*-dimethylpyrrolosamine and α-oleandrose were identified by ^{1}H and ^{13}C NMR, the analysis of $^3J_{H-H}$ coupling constants, and a comparison to isolates containing these carbohydrate residues. The absolute configuration of either carbohydrate was not defined but was subsequently established as L by the degradation of (−)-lomaiviticin C (vide infra). The absolute configuration of the aglycon was assigned by analogy to the kinamycins.

Carbon-13, HSQC, and HMQC spectral data suggested that each cyclohexenone contained one carbonyl group, one ethyl substituent, two tertiary carbons, and three quaternary carbons (Figure 1c). A correlation between H2/H2′ and C2/C2′ in both the HMBC and HMQC spectra established the position of the cojoining carbon−carbon bond at C2. The location of the oleandrose residue was assigned by $^2J_{H2-C3}$ and $^2J_{H4-C3}$ couplings. A weak HMBC correlation between H4/H4′ and C5/C5′ (diazo carbon) was interpreted as a $^3J_{C-H}$ coupling, anchoring the orientation of the cyclohexenone ring relative to the diazofluorene residue.

The structure elucidation of (−)-lomaiviticin B established the relative configuration of the cyclohexenone in both isolates. The fused furanol structure **2** was logically advanced based on the absence of the oleandrose and ketone residues (HRMS and NMR) in (−)-lomaiviticin A (Figure 1a and d). It is plausible that (−)-lomaiviticin B derives from the 2-fold deglycosylation of (−)-lomaiviticin A, followed by cyclization. This structure necessitates a *cis*-disposition of the dimeric bond and C3 oxygen substituents. This stereochemical assignment was extended to (−)-lomaiviticin A based on the assumption that the configuration of the cyclohexenone core is identical in the two isolates. Consistent with this, both H2 (H2′) and H4 (H4′) appear as singlets in the 500 MHz ^{1}H NMR spectrum of (−)-lomaiviticin A. However, a correlation between these protons was observed in the COSY spectrum and was attributed to a four-bond W-coupling, where $^4J_{H-H}$ < 2 Hz is typical.[11] This putative W-coupling requires a *cis*-diequatorial disposition of H2 (H2′) and H4 (H4′), supporting the same relative stereochemical arrangement for both isolates (Figure 1d).

From 2012 to 2013, (−)-lomaiviticin C was isolated by our laboratory[12] and by Moore and co-workers (Figure 2a).[13] NMR and HRMS analysis revealed that (−)-lomaiviticin C is

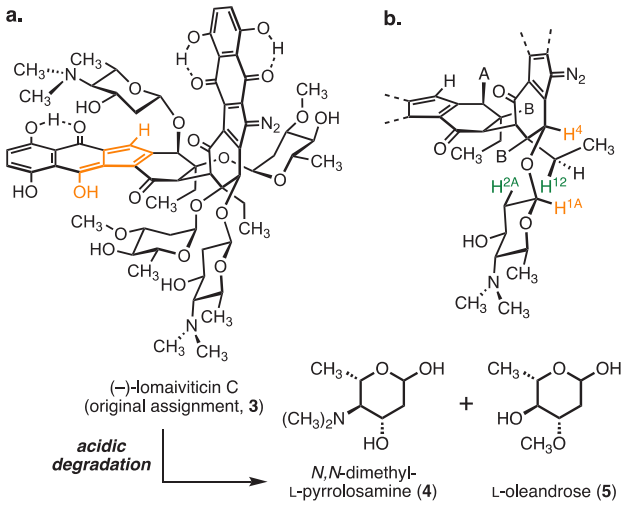


Figure 2. a. Structure **3** advanced for (−)-lomaiviticin C in 2012. Acid-catalyzed hydrolysis of the carbohydrate residues followed by optical analysis established that both possess the L configuration. b. ROESY correlations used to elucidate the absolute configuration of the aglycon in the lomaiviticins are shown in orange and green. **A** and **B** represent the oxygen-linked β-*N*,*N*-dimethyl-L-pyrrolosamine and α-L-oleandrose residues, respectively.

nearly identical to (−)-lomaiviticin A, leading to the assignment of structure **3** (Figure 2a). The only difference between structures **1** and **3** is the presence of a hydroxyfulvene in (−)-lomaiviticin C (orange in Figure 2a), as opposed to two diazofluorenes in (−)-lomaiviticin A. The hydroxyfulvene was supported by features at δ 6.72 ppm (H5′, s, 1H) and δ 120.9 ppm (C5′) in the H-1 and C-13 NMR spectra, respectively, of (−)-lomaiviticin C, which are similar to synthetic hydroxyfulvenes.[14] Doubling of most signals (relative to (−)-lomaiviticin A) was observed, supporting a C_1-symmetric structure.

Acidic digestion of (−)-lomaiviticin C allowed for isolation of the carbohydrate residues; optical analysis established that both are of the L-form. The configuration of the aminosugar was used to infer the absolute configuration of the aglycon (Figure 2b). A ROESY correlation between H1A and H4 requires a *cis*-diaxial disposition of these protons relative to the C−O−C plane. An additional ROESY correlation between $H2A_{eq}$ and H12 is then accommodated only by the diastereomer shown, since it requires both protons to be on the same face of the cyclohexenone ring. Diazotransfer to (−)-lomaiviticin C provided semisynthetic (−)-lomaiviticin A,[12] and hydrolysis of the oleandrose residues of semisynthetic (−)-lomaiviticin A provided semisynthetic (−)-lomaiviticin B.[15] Semisynthetic samples of (−)-lomaiviticins A and B obtained in this way were indistinguishable from natural material, establishing the structural homology among the three isolates.

In 2019, we were contacted by Prof. Hosea Nelson and co-workers, who were interested in obtaining complex isolates for microED analysis. Initially developed for elucidating the structures of biological macromolecules as frozen hydrated crystals, microED analysis is rapidly emerging as a powerful method for the structure determination of small organic molecules using nanocrystalline material (~10^{-15} g).[16] We provided a sample of (−)-lomaiviticin C fully expecting that the microED analysis would be confirmatory. A preliminary structure obtained in early 2021[17] led to the startling discovery that the original assignment of the lomaiviticins was incorrect.[4]

https://doi.org/10.1021/acs.accounts.3c00183
Acc. Chem. Res. 2023, 56, 1656−1668

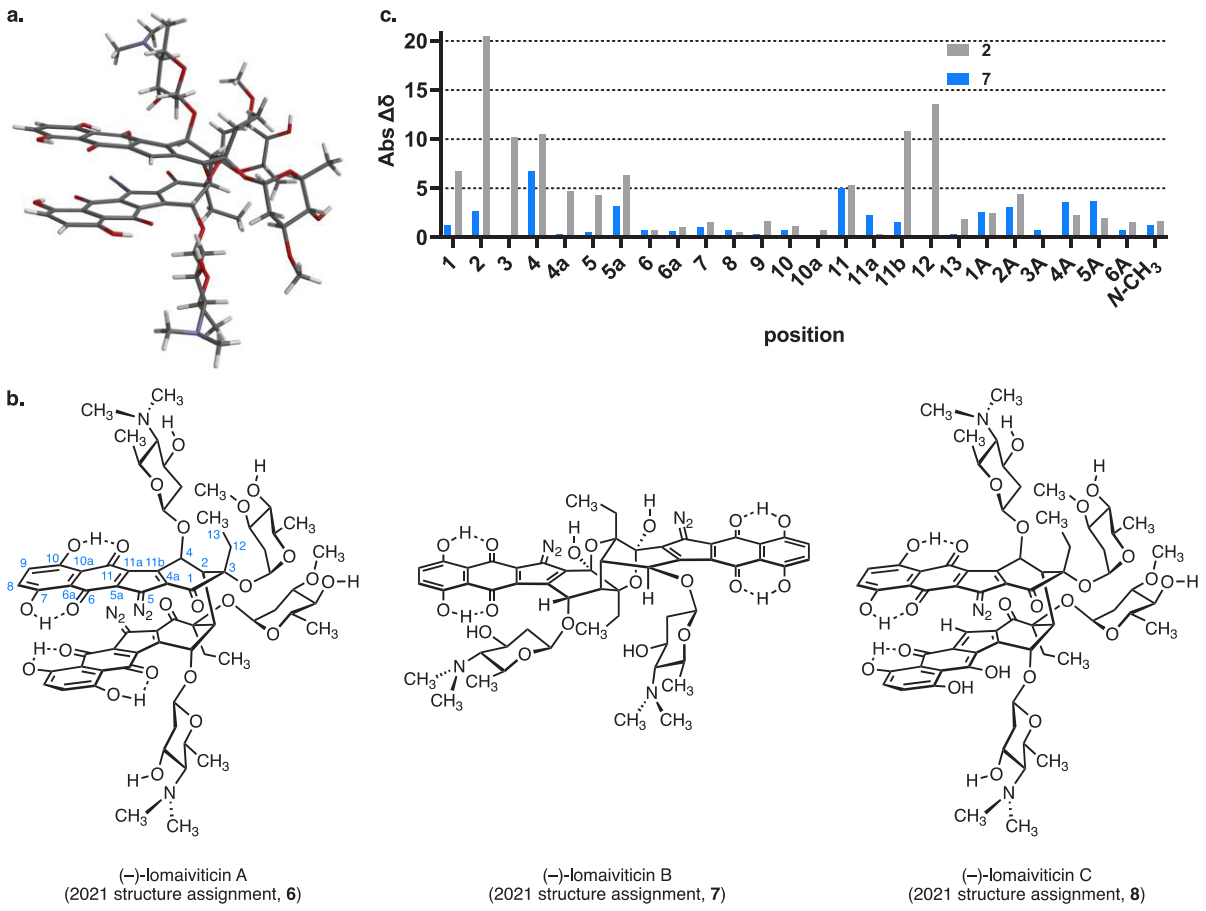


Figure 3. a. MicroED structure of (−)-lomaiviticin C. b. Revised structures of the lomaiviticins. c. Absolute difference between theoretical carbon-13 chemical shifts of structures **2** (see Figure 1a) and **7** and natural (−)-lomaiviticin B. Carbon-13 shifts were determined from the Boltzmann average of the theoretical carbon-13 shifts of the three low-energy conformers of **2** and **7** (ωB97X-V/6-311+G(2DF,2P)[6-311G*]). Gray series: **2** (root-mean-square error (RMSE) = 6.63); blue series: **7** (RMSE = 2.83).

The microED study (Figure 3a) revealed that the C1 and C4 positions were exchanged and the configuration of C4 was inverted relative to the original assignment. Since the homology among (−)-lomaiviticins A, B, and C was established (vide supra), the structures of each isolate were revised as shown in Figure 3b.

DFT calculations[18] were conducted to compare the theoretical carbon-13 shifts of the original (**2**) and new (**7**) (−)-lomaiviticin B structures to experimental values of the natural isolate. (−)-Lomaiviticin B was selected due to the presence of fewer carbohydrate residues and a more rigid structure. The root-mean square errors (RMSEs) between the theoretical and experimental carbon-13 shifts of (−)-lomaiviticin B were 6.63 and 2.83 for structures **2** and **7**, respectively (Figure 3c). The theoretical C−H and H−H coupling constants for the new structure of (−)-lomaiviticin C (**8**) were also in better agreement with the experimental values.

We carried out high-field NMR studies (800 MHz ^{1}H, 200 MHz ^{13}C, cold probe) of natural (−)-lomaiviticin C to probe the revised structure and gain insight into where the original structure determination went awry. We observed many signals consistent with the revised structure but inconsistent with the original assignment. A selection of diagnostic HMBC and ROSEY correlations is shown in Figure 4. HMBC correlations observed between H4 and C2′ (and H4′ and C2) and H12 and C1 (and H12′ and C1′) would require $^{4}J_{C-H}$ couplings through an aliphatic system in the original structure **3**, which are rare (Figure 4a).[19] In the revised structure **8**, these correlations are assigned as common $^{3}J_{C-H}$ couplings. Molecular dynamics simulations were used to determine the lowest-energy conformers of **3** and **8** (Figures 2a and 3b). The lowest-energy conformation of **8** was in qualitative agreement with the microED structure (Figure 3a). These structures were then used to guide ROESY analysis (Figure 4b). A ROESY correlation observed between H2 and H4 (and H2′ and H4′) seems implausible in structure **3**, as it would require a ring inversion to place H2/H4 or H2′/H4′ in pseudoaxial positions. In structure **8**, H2 and H4 (and H2′ and H4′) occupy a vicinal, *trans*-diequatorial relationship, and a ROESY correlation would be expected. Not depicted in Figure 4b (for clarity) are weak ROESY correlations between H2 and H4′ (and H2′ and H4). These would be unlikely in structure **3**, as the protons are located on opposite faces of each cyclohexenone ring. However, these correlations can be interpreted as a transannular interaction arising from the *cis*-disposition of each pair of protons (with respect to each cyclohexenone ring) in structure **8**.

Though we do not have access to the NMR data files from the 2001 study (and the Wyeth team has, to the detriment of the field, been disbanded), an inspection of tabulated and graphical NMR spectra suggests that the two structural series discussed above are essentially indistinguishable using the 2001 data. Figure 4c depicts the HMBC correlations to H4 in (−)-lomaiviticin A available in this data set (except for C1A).

https://doi.org/10.1021/acs.accounts.3c00183
Acc. Chem. Res. 2023, 56, 1656−1668

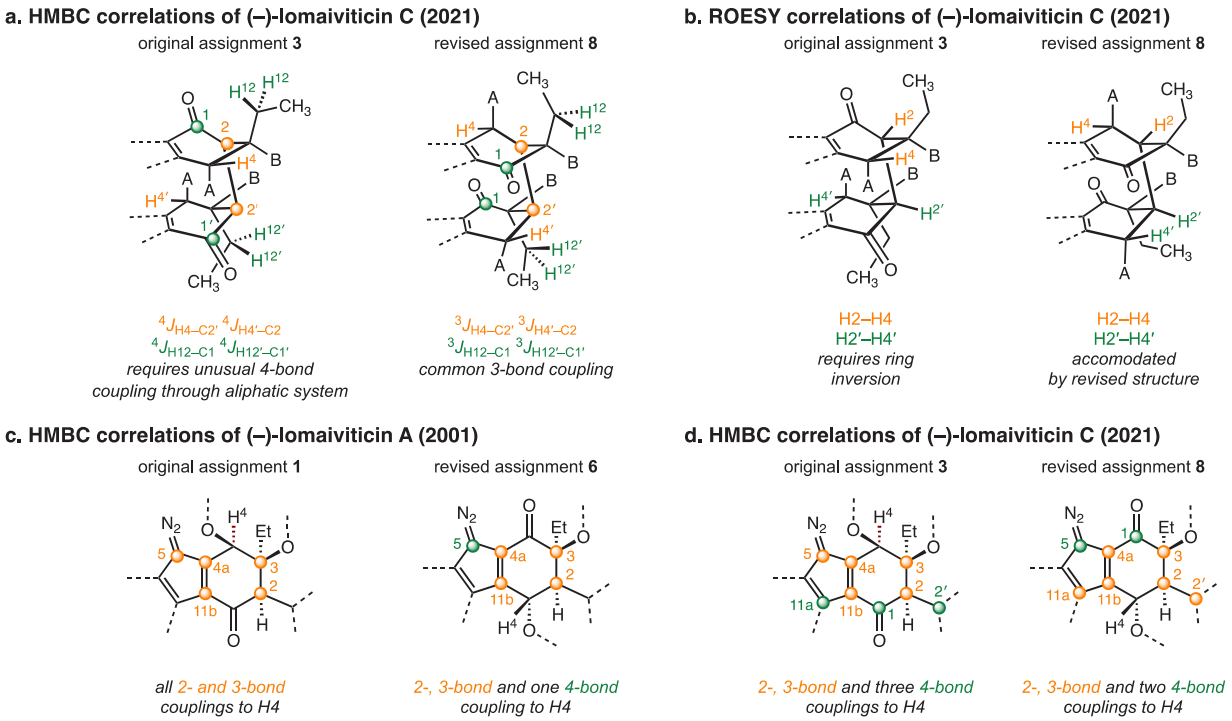


Figure 4. a. Selected HMBC correlations observed for (−)-lomaiviticin C in 2021 (800 MHz ^{1}H, cold probe) and their application to structures **3** and **8**. b. Selected ROESY correlations observed for (−)-lomaiviticin C in 2021 (800 MHz ^{1}H, cold probe) and their application to structures **3** and **8**. c. Reported HMBC correlations observed for (−) lomaiviticin A in 2001 (500 MHz ^{1}H) and their application to structures **1** and **6**. d. Additional HMBC correlations observed for (−)-lomaiviticin C in 2021 (800 MHz ^{1}H, cold probe) and their application to structures **3** and **8**. A and B in Figure 4a,b represent the oxygen-linked *β*-*N*,*N*-dimethyl-L-pyrrolosamine and *α*-L-oleandrose residues, respectively.

All of these correlations correspond to common $^2J_{C–H}$ and $^3J_{C–H}$ couplings in structure **1**. We detected a weak HMBC correlation between H4 and the diazo carbon (C5) in our 2021 studies of (−)-lomaiviticin C. In the revised structure of (−)-lomaiviticins A and C (**6** and **8**, respectively), this correlation is assigned as a well-precedented allylic $^4J_{C–H}$ coupling.[20] Additionally, the HMBC spectrum of (−)-lomaiviticin C obtained in 2021 revealed new correlations between H4 and C2′, H4 and C1, and H4 and C11a (Figure 4d). The correlations from H4 to C11a and H4 to C1 are theoretically accommodated by both **3** and **8** (4-bond allylic couplings in **3**; 3-bond and 4-bond allylic couplings in **8**), but the correlation from H4 to C2′ occurs through an aliphatic system and would be expected only in **8**. This correlation is present only in C_1-symmetric (−)-lomaiviticin C, as the C_2-symmetric structure of (−)-lomaiviticin A makes this indistinguishable from a $^3J_{H4–C2}$ coupling. This signal was overlooked in the original isolation of (−)-lomaiviticin C due to the inability to clearly resolve C2 and C2′ on a 500 MHz instrument.

■ COLIBACTIN

The intestine contains diverse microorganisms that impact physiology and disease.[21] Numerous correlations between microbiome composition and host physiology are documented, but the isolation and characterization of metabolites that may underpin these correlations are challenging. Such characterization is required if researchers are to probe for metabolite-dependent causal relationships between the microbiome and human disease.

In 2006, it was discovered that certain *E. coli* and other proteobacteria harbor a 54-kb BGC termed *clb* (aka *pks*) which encodes the biosynthesis of the genotoxin colibactin.[22] Mammalian cells exposed to *clb*$^+$ *E. coli* accumulated DNA double-strand breaks (DSBs). Mice infected with *clb*$^+$ *E. coli* developed tumors under inflammatory conditions, and clinical data revealed an increased prevalence of *clb*$^+$ bacteria in colorectal cancer (CRC) patients.[23] Despite the strong evidence supporting a role for colibactin in CRC, our inability to isolate the metabolite and elucidate its structure hampered the study of its role in carcinogenesis. In our minds, this points to an important frontier in the field: Advances in genomics provide methods to rapidly identify novel BGCs, but how does one pursue the characterization of these metabolites when they are unstable or produced in quantities too small for isolation? Certainly, this challenge has been recognized by the community.[24] Here we describe how we used a multidisciplinary approach to elucidate the structure and mechanism of action of colibactin. For a more detailed discussion, the reader is directed to ref 25.

The *clb* gene cluster encodes an inactive biosynthetic intermediate termed precolibactin, which contains an *N*-myrisotyl-D-Asn residue.[26] Precolibactin is transported to the periplasm[27] where it is converted to colibactin by removal of the *N*-myrisotyl-D-Asn residue by the serine protease ClbP.[26] As researchers were unable to isolate colibactin from wild-type bacteria possessing an intact *clb* pathway, a strategy involving the large-scale fermentation of genetically modified *clb*$^+$ *E. coli* strains was pursued.[26c] These studies primarily employed *clbP* mutant strains, based on the underlying hypothesis that ClbP inactivation would result in the accumulation of precolibactin. These *clbP* mutant strains were often further modified by the knockdown of other genes. This approach has led to the identification of >40 *clb* products to date.[28] These isolates have provided substantial insight into colibactin's structure and the

https://doi.org/10.1021/acs.accounts.3c00183
Acc. Chem. Res. 2023, 56, 1656−1668

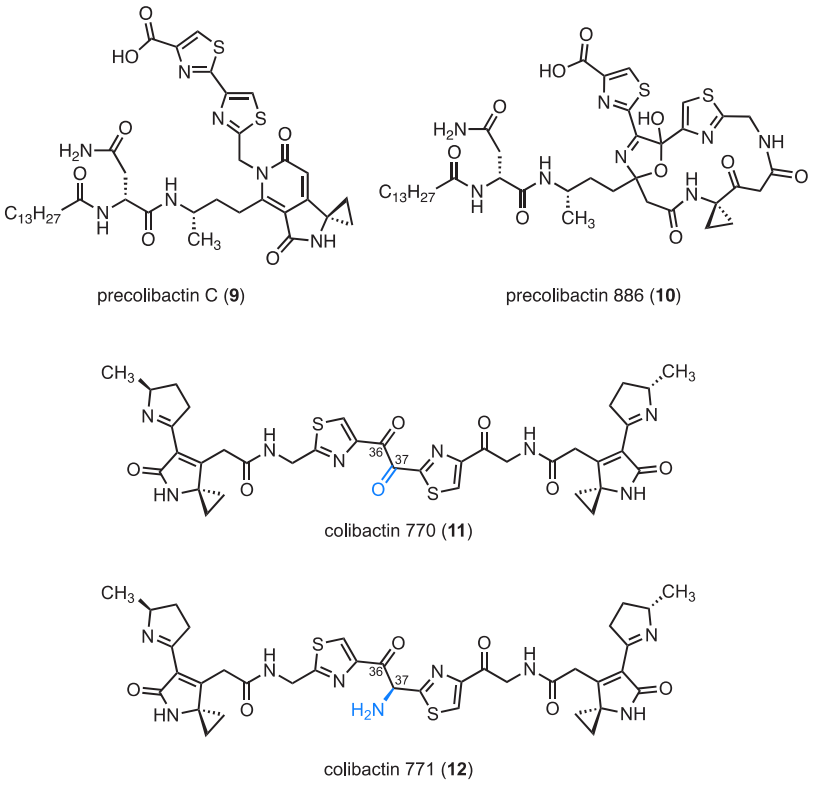


Figure 5. Structures of selected products derived from mutant (for **9** and **10**) or wild-type (for **11** and **12**) *clb*$^+$*E. coli*.

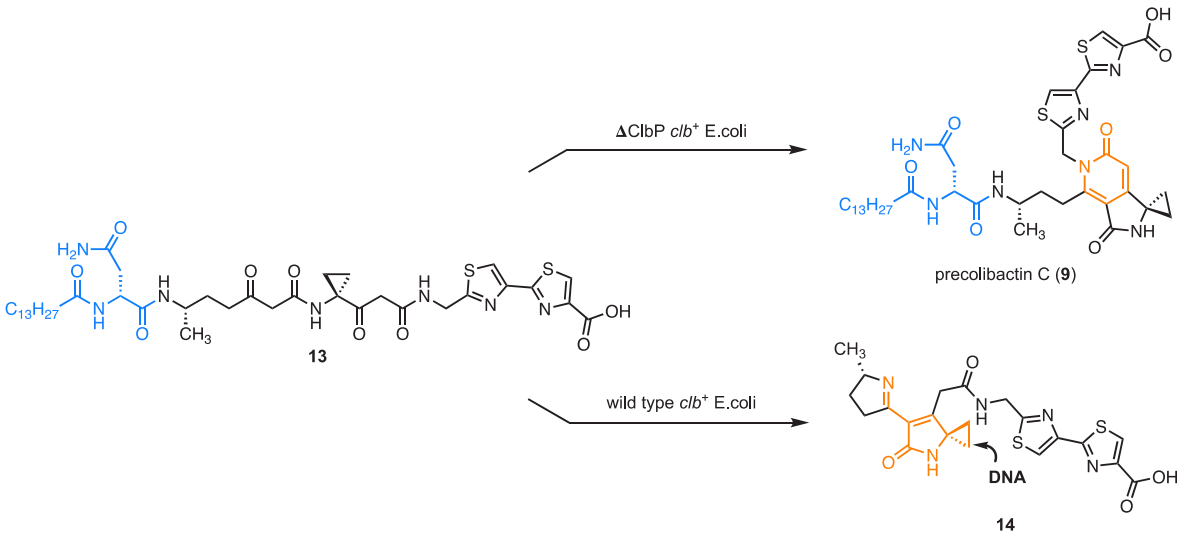


Figure 6. In *clbP* mutant *clb*$^+$ *E. coli*, the linear biosynthetic intermediate **13** undergoes a 2-fold cyclodehydration to yield nongenotoxic pyridones, such as precolibactin C (**9**). Synthetic studies revealed that in wild-type *clb*$^+$ *E. coli*, ClbP-mediated deacylation of linear precolibactins, such as **13**, provides unsaturated imines, such as **14**, which are potent DNA alkylating agents.

function(s) of the modified enzymes.[29] However, while the mutation of *clbP* was thought to simply promote the accumulation of stable *clb* products, we discovered that this genetic modification derails the biosynthetic pathway.[30] Herein, we focus on the *clb* products precolibactin C (**9**), precolibactin 886 (**10**), colibactin 770 (**11**), and colibactin 771 (**12**, Figure 5). We outline how the study of these metabolites contributed to our understanding of colibactin's biosynthesis, its molecular mechanism of action, and the structure of colibactin 771 (**12**) itself, which is thought to be the final *clb* product but which has still eluded direct isolation. Recently, a degradation product of colibactin 772 (**12**) was isolated from wild-type *clb*$^+$ *E. coli*.[31]

Precolibactin C (**9**) was predicted[32] in 2015 and later isolated from a mutant *clb*$^+$ *E. coli* strain (Figure 5).[33] Biosynthetic studies indicated that precolibactin C (**9**) was off-loaded from the assembly line as its linear precursor before undergoing nonenzymatic cyclization to generate the pyridone residue.[34] The presence of a cyclopropane led to speculation that colibactin's genotoxicity results from the alkylation of DNA by ring-opening addition,[35] a mechanism of action established for other classes of metabolites.[36] However, we

https://doi.org/10.1021/acs.accounts.3c00183
Acc. Chem. Res. 2023, 56, 1656−1668

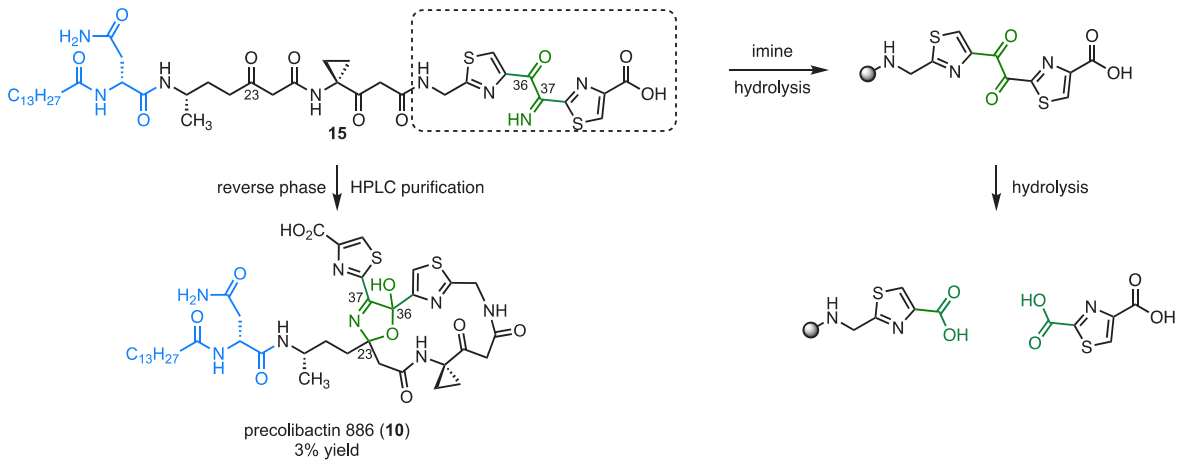


Figure 7. Studies suggest that the macrocyclic precolibactin 886 (**10**) is an artifact resulting from the cyclization of a linear ketoimine precursor during the analytical and purification processes. Advanced precolibactins containing the ketoimine residue are susceptible to C36−C37 bond cleavage.

posited that these pyridones would be poor DNA alkylating agents, as cyclopropane opening would require disruption of aromaticity. We hypothesized that these isolates are biosynthetic derailment products derived from the persistence of the *N*-myrisotyl-D-Asn residue (Figure 6). We proposed that in wild-type strains, ClbP-mediated deacylation of linear precolibactins, such as **13**, would trigger an alternative pathway to generate an unsaturated imine residue (**14**), a scaffold that had been previously proposed[35] but which had eluded isolation.

We used a chemical synthetic approach to show that both cyclization pathways were viable from the same intermediate, providing access to compounds containing either pyridone or unsaturated imine residues.[37] Mechanism-of-action studies established that the unsaturated imines, but not pyridine-containing analogs, alkylate DNA by ring-opening addition.[30] This model was further supported by the observation that synthetic unsaturated imines such as **14** are substrates of the resistance enzyme ClbS, a cyclopropane hydrolase that protects *clb*$^+$ bacteria from autotoxicity.[38] The mutation of *clbP* promotes the accumulation of metabolites containing *N*-myrisytoyl-D-Asn, but the persistence of this residue diverts the biosynthetic pathway toward unnatural, pyridone-containing products such as precolibactin C (**9**).

Efforts to characterize more advance precolibactin metabolites resulted in the isolation of the macrocyclic precolibactin 886 (**10**; Figure 5).[39] Precolibactin 886 (**10**) is one of the most biosynthetically advanced precolibactins, with only three *clb* enzymes unaccounted for in its biosynthesis. The unusual structure of precolibactin 886 (**10**) and a low titer (2.8 mg isolated from a 1000 L fermentation) raised questions about its origin. We synthesized ketoimine **15**, the putative precursor to precolibactin 886 (**10**, Figure 7).[2] All attempts to achieve the macrocyclization of **15** were unsuccessful. Instead, products resulting from the hydrolysis of the ketoimine and, surprisingly, cleavage of the C36−C37 bond were formed. In light of this, we began to suspect that the cyclization may occur during analytical or preparative HPLC, which had been used to both monitor the cultures and isolate precolibactin 886 (**10**).[39] Subjecting **15** to semipreparative HPLC purification provided precolibactin 886 (**10**; 3%).[2] The low mass balance likely derives from decomposition by C36−C37 bond cleavage during the purification (discussed above). Our findings make clear that the oxidized two-carbon spacer renders advanced *clb* products unstable. We cannot rigorously exclude the possibility that macrocyclization does occur during fermentation (and we were simply unable to recapitulate these conditions in the laboratory), but our studies would seem to suggest that macrocyclic precolibactins are artifacts generated during the analytical and purification processes. A more complex derivative of precolibactin 886 (**10**) was subsequently isolated.[40] This isolate was not completely characterized, and the structure assignment is likely incorrect.[41]

While the isolation studies discussed above provided advanced precolibactins and the mechanism-of-action studies provided the first link between metabolite structure and *clb* genotoxicity, the structure assignment of colibactin remained incomplete. In 2018, Nougayrède and co-workers disclosed that HeLa cells infected with *clb*$^+$ *E. coli* accumulated DNA interstrand cross-links (ICLs).[42] At first glance, this observation appears to be inconsistent with the DSB phenotype reported earlier.[22] However, the transient production of DSBs is an obligate step in the repair of ICLs by the Fanconi anemia (FA) pathway.[41a] Separately, it was known that all of the biosynthetic enzymes in the *clb* gene cluster are required for the genotoxic phenotype.[22,42] Collectively, these observations suggested to us that the vestiges of colibactin may be entrained in the ICLs first observed by Nougayrède. Characterization of the colibactin-derived ICL might provide a means to infer its structure.

To achieve this, we carried out high-resolution LC-MS/MS analysis of DNA−colibactin adducts arising from the digestion of exogenous DNA that had been added to cultures of wild-type and auxotrophic *clb*$^+$ *E. coli* (the latter were supplemented with carbon-13-labeled amino acids).[43,3] The structure of the bis(adenine) adduct **16** was established by this approach (Figure 8). Though the location of the adenine base could be confidently assigned, we could not assign the site of adenine alkylation. The N3-linked structure shown is based on studies by Balskus and co-workers, who established that synthetic colibactin fragments[44,37] form monoadenine adducts linked through N3.[45]

The structure **16** was further supported by the observation of two mono(adenine) adducts arising from oxidative hydrolysis of the C36−C37 bond of **16** (Figure 7). Based on biosynthetic logic and the analysis of the DNA adduct **16**, we

https://doi.org/10.1021/acs.accounts.3c00183
Acc. Chem. Res. 2023, 56, 1656−1668

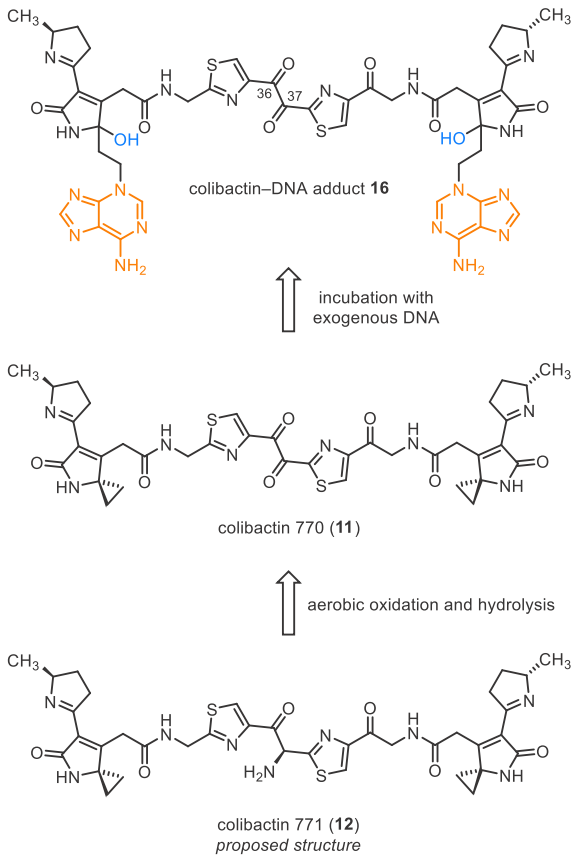


Figure 8. Isotope labeling and tandem MS analysis of *clb*$^+$ *E. coli* induced-DNA ICLs led to the identification of colibactin–DNA adduct **16**. α-Aminoketone colibactin 771 (**12**) is proposed as the structure of colibactin based on biosynthetic logic. Colibactin 770 (**11**), which arises from aerobic oxidation and the hydrolysis of **12**, was observed in bacterial extracts; its structure was confirmed by chemical synthesis. DNA ICLs induced by synthetic samples of **11** were indistinguishable (by tandem MS analysis) from those induced by *clb*$^+$ *E. coli*.

proposed the α-aminoketone **12** (colibactin 771) as the structure of colibactin. While undetectable in bacterial extracts, we did observe and characterize the diketone colibactin 770 (**11**), which derives from aerobic oxidation and hydrolysis of the α-aminoketone residue of **12**. DNA adduct **16** arises from opening of the cyclopropane of **11** or **12** by adenine. We synthesized colibactin 770 (**11**) and demonstrated that it was identical to natural material by LC/MS co-injection. DNA cross-links induced by synthetic colibactin 770 (**11**) were indistinguishable (by tandem MS analysis) from those induced by *clb*$^+$ *E. coli*.[3] A similar structure assignment was advanced contemporaneously by Balskus and co-workers.[46] Notably, the bis(adenine) adduct **16** was observed as its doubly charged cation in the MS spectra of DNA digests. The singly charged cation was ~10^3 less abundant, as expected for a small molecule that contains two basic residues.

OCIMICIDES

In 2010, the structures and antimalarial activities of ocimicide A_1 (**17**), ocimicide B_1 (**18**), and ocimicide C_1 (**19**) and semisynthetic derivatives (+)-ocimicide A_2 (**20**), (−)-ocimicide B_2 (**21**), (−)-ocimicide C_2 (**22**), and (+)-ocimicide C_3 (**23**) were disclosed in the patent literature (Figure 9).[47] The natural ocimicides were obtained from extracts of *Ocimum sanctum* root bark. The ocimicides reportedly demonstrated nM antimalarial activity against chloroquine-sensitive and -resistant strains of *Plasmodium falciparum* and high selectivities toward malarial parasites over adult human epithelial cells. The semisynthetic derivatives **20**–**23** provided mice with prophylactic protection from *Plasmodium berghei* infection[47] and effected a radical cure in rhesus monkeys,[47b] without detectable toxicity.

The ocimicide alkaloids (**17**–**23**) were reported to contain a rigid hexacyclic core comprising a tetrasubstituted amino-cyclopropane, a pyrrolidine ring, and a quinoline ring with apparently alternating orientations (compare **17** and **18**). Notably, the central cyclopropane ring is part of a reactive donor–acceptor system.[48] The structures of **17**–**23** were established by HRMS, IR, UV–vis, and NMR, but no data was

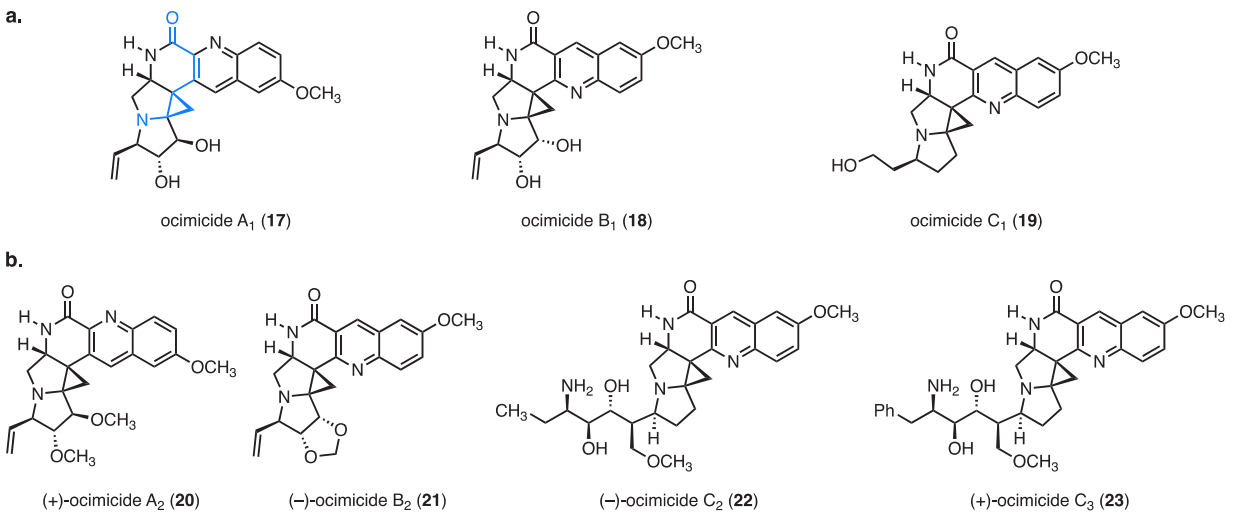


Figure 9. Reported structures of a. ocimicide A_1 (**17**), ocimicide B_1 (**18**), and ocimicide C_1 (**19**) and b. semisynthetic derivatives (+)-ocimicide A_2 (**20**), (−)-ocimicide B_2 (**21**), (−)-ocimicide C_2 (**22**), and (+)-ocimicide C_3 (**23**). The donor–acceptor cyclopropane system is highlighted in blue in **17**.

https://doi.org/10.1021/acs.accounts.3c00183
Acc. Chem. Res. 2023, 56, 1656−1668

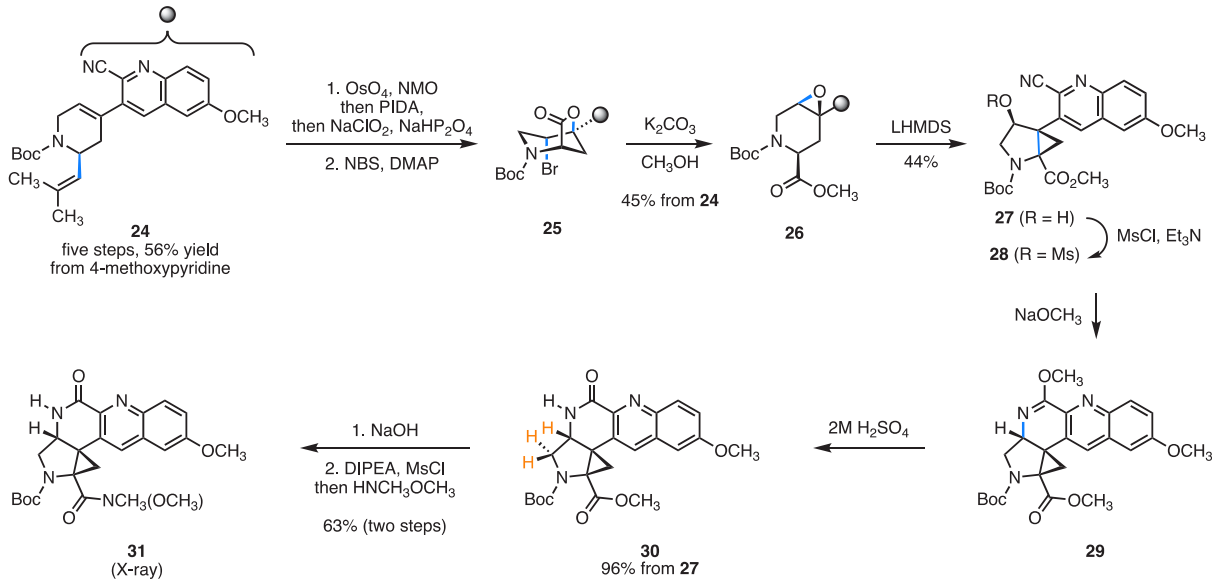


Figure 10. Synthesis of lactam **30** and Weinreb amide **31**. To track the stereochemical relay, newly formed bonds in each step are highlighted in blue. The $^{3}J_{H-H}$ coupling constants between the protons shown in orange in **30** did not agree with those reported for ocimicide A_1 (**17**).

provided for the natural isolates.[47] Tabulated spectroscopic data were reported for the semisynthetic derivatives **20**−**23**.

We developed a synthetic approach to ocimicide A_1 (**17**) that employed a stereochemical relay to establish the azabicyclo[3.1.0]hexane core, which contains a tetrasubstituted cyclopropane (Figure 10).[1] The alkene within **24** (prepared in five steps and 56% yield from 4-methoxypyridine) was cleaved and converted to a carboxylic acid. Stereoselective bromolactonization generated the bromolactone **25**. Exposure of the unpurified bromolactone **25** to potassium carbonate in methanol provided the epoxy ester **26** (45% from **24**). Deprotonation of **26**, followed by heating, induced an epoxide-opening−ring-contraction reaction to establish the azabicyclo[3.1.0]hexane **27** (44%). Mesylation of **27**, followed by treatment with sodium methoxide, provided the methyl imidate **29**. Acid-catalyzed hydrolysis of **29** then generated the lactam **30** (96% from **27**). By this approach, the relative configuration of the azabicyclo[3.1.0]hexane residue was derived from the bromolactonization step. Saponification of the methyl ester **30**, followed by Weinreb amide formation, provided the crystalline Weinreb amide **31**, whose structure was confirmed by X-ray analysis.

The *tert*-butyl carbamate of **31** could be removed under acidic conditions, and the resulting ammonium salt **32** was stable (Figure 11). However, all attempts to form the free base of **32** resulted in uncharacterized decomposition products. It

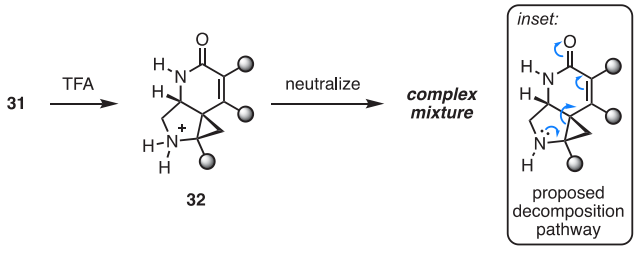


Figure 11. Attempts to form the free base of ammonium salt **32** resulted in decomposition, potentially through opening of the donor−acceptor cyclopropane (see inset).

seemed plausible that ring opening of the donor−acceptor cyclopropane occurred (inset of Figure 11). The ocimicides themselves were reportedly stable isolates.[47] Analysis of the coupling constants within the azabicyclo[3.1.0]hexane core revealed inconsistencies between our intermediate **30** (Figure 10) and the natural and semisynthetic isolates. The $^{3}J_{H-H}$ coupling constants within **30** (5.2 and 0 Hz) did not match those reported for (+)-ocimicide A_2 (**20**; 9.2 and 4.4 Hz).

We employed DFT calculations using the protocol of Hoye and co-workers[49] to determine the theoretical carbon-13 chemical shifts for all possible diastereomers of (+)-ocimicide A_2 at carbons 12, 13, 14, and 17 and N15 (Figure 12). Each of the 32 diastereomers was subjected to a conformational search using BOSS,[50] and the geometry of all conformers within 5 kcal/mol of the lowest-energy conformer was optimized. The carbon-13 shifts were calculated[51] and Boltzmann-averaged. The accuracy of these calculations was benchmarked against 19-(*E*)-hunteracine (**33**)[52] and the ammonium salt of **32** (Figure 12b).[51,53]

A comparison of the reported and theoretical carbon-13 chemical shifts for (+)-ocimicide A_2 (**20**) revealed large inconsistencies, notably at the lactam−quinoline fusion [C9 and 20; absolute error (AE) = 7.3−13.1 ppm] and the cyclopropane (C10; AE = 4.2−9.8 ppm). These three positions were calculated for the salt **32** with high accuracy (AE = 0.3−1.8 ppm for C9 and 20; AE = 2.8 ppm for C10), suggesting a discrepancy between the actual structure and the assignment at this region of the molecule. Larger deviations were observed within the pyrrolidine ring of **20** (C13, 14, 22, and 23; AE = 8.3−20.9 ppm) and the diastereomers **20a** (AE > 6.0 ppm for C14, 22, and 23), indicating that the spectroscopic differences were not exclusively of a stereochemical nature.

In the absence of both original spectra and 2D NMR data or access to the original isolates, we attempted to reisolate the material from various strains of *Ocimum sanctum* following the reported protocol but were unable to observe (by LC/MS) the alkaloids in the plant extracts. In light of this and with no clear alternative structure to computationally examine or synthesize, our efforts toward the ocimicide alkaloids came to an end. The observed reactivity and spectroscopic differences led us to

https://doi.org/10.1021/acs.accounts.3c00183
Acc. Chem. Res. 2023, 56, 1656−1668

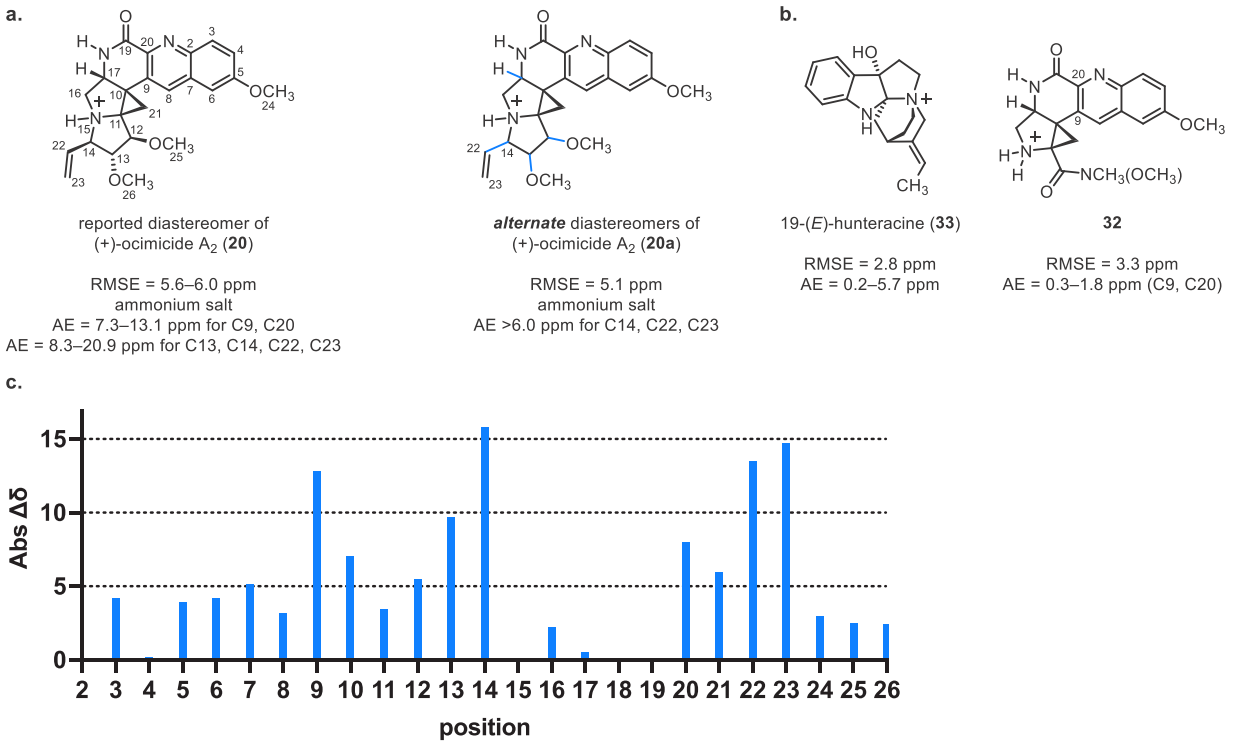


Figure 12. a. RMSE and absolute error (AE) of calculated carbon-13 chemical shifts for the reported diastereomer of ocimicide A_2 (**20**) and the average of all alternate diastereomers of (+)-ocimicide A_2 (**20a**) varied at positions 12, 13, 14, and 17. The numbering system employed corresponds to that used in the isolation reports. b. RMSE and AE of theoretical carbon-13 chemical shifts for 19-(*E*)-hunteracine (**33**) and Weinreb amide ammonium salt **32**. c. Absolute difference between theoretical and reported carbon-13 chemical shifts for the reported structure of ocimicide A_2 (**20**).

conclude that a structural revision of these natural products is likely necessary.

DISCUSSION AND CONCLUSIONS

We end with some thoughts on the limitations and future of secondary metabolite structure determination and standards for data reporting, and we advocate for the systematic incorporation of carbon-13 DFT calculations into structure determination and synthesis.

MicroED

Our lomaiviticin studies underscore the significant impact that microED stands to have on secondary metabolite structure determination. In the absence of microED, it would have been difficult to advance the structure revision with confidence due to the ambiguity of multiple-bond C−H correlations. Additionally, the C_2 symmetry of (−)-lomaiviticins A and B makes it impossible to distinguish transannular ROESY and HMBC correlations from interactions of atoms within the same ring. Though introduced only in 2018,[16] microED has already been employed in conjunction with state-of-the-art genomics methods to facilitate the identification of novel secondary metabolites and to correct additional structure misassignments.[54] The impact of this technique is certain to increase further as purpose-built microED instruments become widely available.

NMR Analysis

Recent advances in instrumentation allow NMR analysis to be carried out on <1 μmol quantities of sample.[55] However, the interpretation of these data can still be problematic, especially for highly oxidized metabolites such as the lomaiviticins. The original structure assignment relied heavily on HMBC correlations, and it can be difficult or impossible to distinguish $^2J_{C-H}$, $^3J_{C-H}$, or $^4J_{C-H}$ couplings. A common misconception is that longer-range correlations display lower intensity than shorter-range correlations. In fact, the intensity of H−C couplings is largely unpredictable.[56] New methods for differentiating two- and three-bond HMBC correlations (e.g., separate echo and antiecho XLOC or SEA XLOC experiments)[57] as well as directly delineating carbon−carbon correlations (adequate sensitivity double-quantum spectroscopy, ADEQUATE)[58] have been developed. However, extended instrumentation time and optimization of the experiment are required to properly implement these methods.

The construction of a global infrastructure to support the dissemination of NMR data has been advocated for elsewhere,[59] and we echo this sentiment here. Most researchers have access to NMR processing software, and a database that provides FID files would be of immense utility. An extension of this database to the patent literature is warranted as well, though this is a more challenging objective. Automated methods for structure determination based on NMR data are being developed.[60]

DFT NMR Calculations

We used the method of Hoye and co-workers[49] to calculate theoretical carbon-13 shifts in our ocimicide work. Though state of the art at the time, the approach still required a large amount of manual intervention. In our later lomaiviticin work, we employed Spartan,[61] which has a fully automated method to calculate proton and carbon-13 NMR chemical shifts.[18] The software uses iterative geometry optimizations and energy calculations to arrive at a small number of conformers for

https://doi.org/10.1021/acs.accounts.3c00183
Acc. Chem. Res. 2023, 56, 1656−1668

NMR calculations while minimizing computational time. As demonstrated in our lomaiviticin studies these methods provide accurate data for large structures. These computational methods are straightforward to execute, requiring the input of only a single candidate structure in Spartan. As such, we advocate for their systematic use in validating the assignments of novel secondary metabolites. We have also found them to be useful in synthetic studies to facilitate structure assignments of advanced intermediates when X-ray or microED data are not available.

Synthetic Chemistry in Conjunction with Functional Analysis

The natural product landscape has been transformed as a result of the rapid expansion of affordable next-generation sequencing technologies and data mining tools. These technologies have contributed to a surge in the discovery of novel BCGs and have assisted in the prediction of the metabolite chemical structure directly from sequencing data.[62] Nonetheless, the process of isolating and functionally characterizing the encoded natural products remains a challenge. In the colibactin story, we demonstrated that the engagement of synthetic chemistry in the structure elucidation process can help to overcome some of the limitations inherent in isolation. By undertaking the targeted synthesis of isolated and putative metabolites, we obtained key insights into the structural assignment, stability, mechanism of action, and biosynthesis of colibactin. We propose that this multidisciplinary approach may enable the structural and functional analysis of other elusive metabolites.

AUTHOR INFORMATION

Corresponding Author

Seth B. Herzon − *Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States; Departments of Pharmacology and Therapeutic Radiology, Yale School of Medicine, New Haven, Connecticut 06520, United States;* orcid.org/0000-0001-5940-9853; Email: seth.herzon@yale.edu

Authors

Mikaela DiBello − *Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States;* orcid.org/0000-0002-0948-5863

Alan R. Healy − *Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States;* Present Address: Chemistry Program, New York University Abu Dhabi (NYUAD), Saadiyat Island, United Arab Emirates (UAE); orcid.org/0000-0002-1231-8353

Herman Nikolayevskiy − *Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States;* Present Address: Department of Chemistry, University of San Francisco, San Francisco, CA 94117, United States.; orcid.org/0000-0002-4658-4182

Zhi Xu − *Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States;* orcid.org/0000-0002-1476-1367

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.accounts.3c00183

Author Contributions

CRediT: **Mikaela DiBello** writing-original draft (supporting), writing-review & editing (supporting); **Alan R. Healy** writing-original draft (supporting), writing-review & editing (supporting); **Herman Nikolayevskiy** writing-original draft (supporting), writing-review & editing (supporting); **Zhi Xu** writing-original draft (lead), writing-review & editing (lead); **Seth B Herzon** writing-original draft (lead), writing-review & editing (lead).

Notes

The authors declare no competing financial interest.

Biographies

Mikaela DiBello was born in Mahopac, NY (1997), received her B.Sc. in chemistry from Rensselaer Polytechnic Institute in 2019, and is currently pursuing her Ph.D. at Yale University, where she has worked on the synthesis of novel pleuromutilin analogs and the total synthesis of diazofluorene natural products.

Alan R. Healy was born in Co. Clare, Ireland in 1988, completed his undergraduate studies in medicinal chemistry at Trinity College Dublin (TCD), and obtained an M.Sc. in biomedical science from the University of Edinburgh and a Ph.D. from St. Andrews University in 2014. Healy was a postdoctoral fellow with Seth B. Herzon and Jason M. Crawford at Yale University and is currently an assistant professor of chemistry at New York University Abu Dhabi (NYUAD). Healy's research focuses on the development of novel methods to accelerate the discovery and study of dark matter metabolites.

Herman Nikolayevskiy was born in Tashkent, Uzbekistan in 1989, completed undergraduate studies in chemical engineering at The Cooper Union for the Advancement of Science and Art, obtained his Ph.D. in chemistry from Yale University 2017, and was an IRTA postdoctoral fellow at the NIH. He is currently an assistant professor and the program director for the MS in Chemistry Program at the University of San Francisco. Nikolayevskiy's research focuses on the development of safer chemotherapies through triggered intramolecular deactivation and the development of covalent inhibitors against bacterial Sortase A.

Zhi Xu was born in Harbin, China in 1997. He obtained his undergraduate degree in chemical biology from Peking University under the supervision of Prof. Xiaoguang Lei. In 2019, he began his graduate studies in the laboratory of Prof. Seth Herzon at Yale University, where he works on total syntheses of diazofluorene and terpenoid natural products.

Seth B. Herzon was born in Philadelphia, PA in 1979, completed undergraduate studies at Temple University, obtained a Ph.D. from Harvard University, and was a postdoctoral fellow at the University of Illinois, Urbana−Champaign. He is currently the Milton Harris '29 Ph.D. Professor of Chemistry and Professor of Pharmacology and Therapeutic Radiology at the Yale School of Medicine and a member of the Yale Comprehensive Cancer Center. Herzon's research focuses on synthetic and translational studies of DNA damage and microbiome-derived secondary metabolites and the development of novel therapeutics targeting tumor-associated DNA repair defects.

ACKNOWLEDGMENTS

Financial support from the National Institutes of Health (R01GM110506, R35-GM131913, and R01CA215553), the Charles H. Revson foundation (postdoctoral fellowship to A.R.H.), and Yale University is gratefully acknowledged.

REFERENCES

(1) Nikolayevskiy, H.; Moe Tun, M. K.; Rablen, P. R.; Ben Mamoun, C.; Herzon, S. B. A complex stereochemical relay approach to the antimalarial alkaloid ocimicide A1. Evidence for a structural revision. *Chem. Sci.* **2017**, *8*, 4867.

(2) Healy, A. R.; Wernke, K. M.; Kim, C. S.; Lees, N. R.; Crawford, J. M.; Herzon, S. B. Synthesis and reactivity of precolibactin 886. *Nat. Chem.* **2019**, *11*, 890.
(3) Xue, M.; Kim, C. S.; Healy, A. R.; Wernke, K. M.; Wang, Z.; Frischling, M. C.; Shine, E. E.; Wang, W.; Herzon, S. B.; Crawford, J. M. Structure elucidation of colibactin and its DNA cross-links. *Science* **2019**, *365*, eaax2685.
(4) Kim, L. J.; Xue, M.; Li, X.; Xu, Z.; Paulson, E.; Mercado, B.; Nelson, H. M.; Herzon, S. B. Structure revision of the lomaiviticins. *J. Am. Chem. Soc.* **2021**, *143*, 6578.
(5) For a review, see Reynolds, W. F. In *Pharmacognosy*; Badal, S.; Delgoda, R., Eds.; Academic Press: Boston, 2017; Chapter 29, pp *567*.
(6) For further discussion, see Miller, S. J.; Clardy, J. Beyond grind and find. *Nat. Chem.* **2009**, *1*, 261.
(7) He, H.; Ding, W. D.; Bernan, V. S.; Richardson, A. D.; Ireland, C. M.; Greenstein, M.; Ellestad, G. A.; Carter, G. T. Lomaiviticins A and B, potent antitumor antibiotics from *micromonospora lomaivitiensis*. *J. Am. Chem. Soc.* **2001**, *123*, 5362.
(8) *Personal communcation from J. Janso*, 2011.
(9) Colis, L. C.; Woo, C. M.; Hegan, D. C.; Li, Z.; Glazer, P. M.; Herzon, S. B. The cytotoxicity of (−)-lomaiviticin A arises from induction of double-strand breaks in DNA. *Nat. Chem.* **2014**, *6*, 504.
(10) For reviews of the kinamycins, see (a) Gould, S. J. Biosynthesis of the kinamycins. *Chem. Rev.* **1997**, *97*, 2499. (b) Marco-Contelles, J.; Molina, M. T. Naturally occurring diazo compounds: The kinamycins. *Curr. Org. Chem.* **2003**, *7*, 1433. (c) Arya, D. P. Diazo and diazonium DNA cleavage agents: Studies on model systems and natural product mechanisms of action. *Top. Heterocycl. Chem.* **2006**, *2*, 129. (d) Nawrat, C. C.; Moody, C. J. Natural products containing a diazo group. *Nat. Prod. Rep.* **2011**, *28*, 1426. (e) Herzon, S. B. The kinamycins. In *Total Synthesis of Natural Products. At the Frontiers of Organic Chemistry*; Li, J. J.; Corey, E. J., Eds.; Springer-Verlag: Berlin, 2012; pp 39. (f) Herzon, S. B.; Woo, C. M. The diazofluorene antitumor antibiotics: Structural elucidation, biosynthetic, synthetic, and chemical biological studies. *Nat. Prod. Rep.* **2012**, *29*, 87.
(11) (a) Williamson, K. L.; Howell, T.; Spencer, T. A. Nuclear magnetic resonance line widths of angular methyl groups in decalins, steroids, and *N*-methylquinolizidinium ions. Determination of ring fusion stereochemistry. *J. Am. Chem. Soc.* **1966**, *88*, 325. (b) Padwa, A.; Shefter, E.; Alexander, E. ″The correlation of the crystal and molecular structure with the nuclear magnetic resonance spectrum of a bicyclo[1.1.1]pentane derivative.″. *J. Am. Chem. Soc.* **1968**, *90*, 3717.
(12) Woo, C. M.; Beizer, N. E.; Janso, J. E.; Herzon, S. B. Isolation of lomaiviticins C−E, transformation of lomaiviticin c to lomaiviticin a, complete structure elucidation of lomaiviticin a, and structure−activity analyses. *J. Am. Chem. Soc.* **2012**, *134*, 15285.
(13) Kersten, R. D.; Lane, A. L.; Nett, M.; Richter, T. K. S.; Duggan, B. M.; Dorrestein, P. C.; Moore, B. S. Bioactivity-guided genome mining reveals the lomaiviticin biosynthetic gene cluster in *salinispora tropica*. *ChemBioChem.* **2013**, *14*, 955.
(14) Woo, C. M.; Lu, L.; Gholap, S. L.; Smith, D. R.; Herzon, S. B. Development of a convergent entry to the diazofluorene antitumor antibiotics: Enantioselective synthesis of kinamycin f. *J. Am. Chem. Soc.* **2010**, *132*, 2540.
(15) Woo, C. M.; Gholap, S. L.; Lu, L.; Kaneko, M.; Li, Z.; Ravikumar, P. C.; Herzon, S. B. Development of enantioselective synthetic routes to (−)-kinamycin F and (−)-lomaiviticin aglycon. *J. Am. Chem. Soc.* **2012**, *134*, 17262.
(16) (a) Jones, C. G.; Martynowycz, M. W.; Hattne, J.; Fulton, T. J.; Stoltz, B. M.; Rodriguez, J. A.; Nelson, H. M.; Gonen, T. The cryoEM method microed as a powerful tool for small molecule structure determination. *ACS Cent. Sci.* **2018**, *4*, 1587. (b) Gruene, T.; Wennmacher, J. T. C.; Zaubitzer, C.; Holstein, J. J.; Heidler, J.; Fecteau-Lefebvre, A.; De Carlo, S.; Müller, E.; Goldie, K. N.; Regeni, I.; Li, T.; Santiso-Quinones, G.; Steinfeld, G.; Handschin, S.; van Genderen, E.; van Bokhoven, J. A.; Clever, G. H.; Pantelic, R. Rapid structure determination of microcrystalline molecular compounds using electron diffraction. *Angew. Chem., Int. Ed. Engl.* **2018**, *57*, 16313. For a review, see (c) Gemmi, M.; Mugnaioli, E.; Gorelik, T. E.; Kolb, U.; Palatinus, L.; Boullay, P.; Hovmöller, S.; Abrahams, J. P. 3D electron diffraction: The nanocrystallography revolution. *ACS Cent. Sci.* **2019**, *5*, 1315.
(17) *Personal communication from Lee Joon Kim (UCLA) to Mengzhao Xue (Yale)*, Jan 11, 2021.
(18) Hehre, W.; Klunzinger, P.; Deppmeier, B.; Driessen, A.; Uchida, N.; Hashimoto, M.; Fukushi, E.; Takata, Y. Efficient protocol for accurately calculating 13C chemical shifts of conformationally flexible natural products: Scope, assessment, and limitations. *J. Nat. Prod.* **2019**, *82*, 2299.
(19) For a discussion, see Williamson, R. T.; Buevich, A. V.; Martin, G. E.; Parella, T. Lr-HSQCMBC: A sensitive NMR technique to probe very long-range heteronuclear coupling pathways. *J. Org. Chem.* **2014**, *79*, 3887.
(20) For a review, see Parella, T.; Espinosa, J. F. Long-range proton−carbon coupling constants: NMR methods and applications. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 17.
(21) For a recent review, see Shine, E. E.; Crawford, J. M. Molecules from the microbiome. *Annu. Rev. Biochem.* **2021**, *90*, 789.
(22) Nougayrède, J.-P.; Homburg, S.; Taieb, F.; Boury, M.; Brzuszkiewicz, E.; Gottschalk, G.; Buchrieser, C.; Hacker, J.; Dobrindt, U.; Oswald, E. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* **2006**, *313*, 848.
(23) For a recent review, see Dougherty, M. W.; Jobin, C. Shining a light on colibactin biology. *Toxins* **2021**, *13*, 346.
(24) Walsh, C. T.; Fischbach, M. A. Natural products version 2.0: Connecting genes to molecules. *J. Am. Chem. Soc.* **2010**, *132*, 2469.
(25) For a review of our work, see Williams, P. C.; Wernke, K. M.; Tirla, A.; Herzon, S. B. Employing chemical synthesis to study the structure and function of colibactin, a "dark matter" metabolite. *Nat. Prod. Rep.* **2020**, *37*, 1532.
(26) (a) Bian, X.; Fu, J.; Plaza, A.; Herrmann, J.; Pistorius, D.; Stewart, A. F.; Zhang, Y.; Muller, R. In vivo evidence for a prodrug activation mechanism during colibactin maturation. *ChemBioChem.* **2013**, *14*, 1194. (b) Brotherton, C. A.; Balskus, E. P. A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. *J. Am. Chem. Soc.* **2013**, *135*, 3359. (c) 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.* **2014**, *136*, 9244.
(27) Mousa, J. J.; Newsome, R. C.; Yang, Y.; Jobin, C.; Bruner, S. D. ClbM is a versatile, cation-promiscuous mate transporter found in the colibactin biosynthetic gene cluster. *Biochem. Biophys. Res. Commun.* **2017**, *482*, 1233.
(28) For a review, see Tang, J.-W.; Liu, X.; Ye, W.; Li, Z.-R.; Qian, P.-Y. Biosynthesis and bioactivities of microbial genotoxin colibactins. *Nat. Prod. Rep.* **2022**, *39*, 991.
(29) (a) Bode, H. B. The microbes inside us and the race for colibactin. *Angew. Chem., Int. Ed. Engl.* **2015**, *54*, 10408. (b) Faïs, T.; Delmas, J.; Barnich, N.; Bonnet, R.; Dalmasso, G. Colibactin: More than a new bacterial toxin. *Toxins* **2018**, *10*, 151. (c) Hirayama, Y.; Sato, M.; Watanabe, K. Advancing the biosynthetic and chemical understanding of the carcinogenic risk factor colibactin and its producers *Biochemistry* **2022**,.*612*782
(30) For a discussion, see Healy, A. R.; Herzon, S. B. Molecular basis of gut microbiome-associated colorectal cancer: A synthetic perspective. *J. Am. Chem. Soc.* **2017**, *139*, 14817.
(31) Zhou, T.; Hirayama, Y.; Tsunematsu, Y.; Suzuki, N.; Tanaka, S.; Uchiyama, N.; Goda, Y.; Yoshikawa, Y.; Iwashita, Y.; Sato, M.; Miyoshi, N.; Mutoh, M.; Ishikawa, H.; Sugimura, H.; Wakabayashi, K.; Watanabe, K. Isolation of new colibactin metabolites from wild-type escherichia coli and in situ trapping of a mature colibactin derivative. *J. Am. Chem. Soc.* **2021**, *143*, 5526.
(32) Li, Z. R.; Li, Y.; Lai, J. Y.; Tang, J.; Wang, B.; Lu, L.; Zhu, G.; Wu, X.; Xu, Y.; Qian, P. Y. Critical intermediates reveal new biosynthetic events in the enigmatic colibactin pathway. *ChemBioChem.* **2015**, *16*, 1715.
(33) Zha, L.; Wilson, M. R.; Brotherton, C. A.; Balskus, E. P. Characterization of polyketide synthase machinery from the *pks* island

https://doi.org/10.1021/acs.accounts.3c00183
Acc. Chem. Res. 2023, 56, 1656−1668

facilitates isolation of a candidate precolibactin. *ACS Chem. Biol.* **2016**, *11*, 1287.
(34) 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.* **2017**, *139*, 4195.
(35) (a) Vizcaino, M. I.; Crawford, J. M. The colibactin warhead crosslinks DNA. *Nat. Chem.* **2015**, *7*, 411. (b) Brotherton, C. A.; Wilson, M.; Byrd, G.; Balskus, E. P. Isolation of a metabolite from the *pks* island provides insights into colibactin biosynthesis and activity. *Org. Lett.* **2015**, *17*, 1545.
(36) See, for example, Boger, D. L.; Garbaccio, R. M. Shape-dependent catalysis: Insights into the source of catalysis for the CC-1065 and duocarmycin DNA alkylation reaction. *Acc. Chem. Res.* **1999**, *32*, 1043.
(37) 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.* **2016**, *138*, 15563.
(38) (a) Bossuet-Greif, N.; Dubois, D.; Petit, C.; Tronnet, S.; Martin, P.; Bonnet, R.; Oswald, E.; Nougayrede, J. P. Escherichia coli clbS is a colibactin resistance protein. *Mol. Microbiol.* **2016**, *99*, 897. (b) Tripathi, P.; Shine, E. E.; Healy, A. R.; Kim, C. S.; Herzon, S. B.; Bruner, S. D.; Crawford, J. M. Clbs is a cyclopropane hydrolase that confers colibactin resistance. *J. Am. Chem. Soc.* **2017**, *139*, 17719.
(39) Li, Z. R.; Li, J.; Gu, J. P.; Lai, J. Y.; Duggan, B. M.; Zhang, W. P.; Li, Z. L.; Li, Y. X.; Tong, R. B.; Xu, Y.; Lin, D. H.; Moore, B. S.; Qian, P. Y. Divergent biosynthesis yields a cytotoxic aminomalonate-containing precolibactin. *Nat. Chem. Biol.* **2016**, *12*, 773.
(40) Li, Z.-R.; Li, J.; Cai, W.; Lai, J. Y. H.; McKinnie, S. M. K.; Zhang, W.-P.; Moore, B. S.; Zhang, W.; Qian, P.-Y. Macrocyclic colibactin induces DNA double-strand breaks via copper-mediated oxidative cleavage. *Nat. Chem.* **2019**, *11*, 880.
(41) For a discussion, see (a) Herzon, S. B. Macrocyclic colibactins. *Nat. Chem.* **2020**, *12*, 1005. See also (b) Tirla, A.; Wernke, K. M.; Herzon, S. B. On the stability and spectroscopic properties of 5-hydroxyoxazole-4-carboxylic acid derivatives. *Org. Lett.* **2021**, *23*, 5457.
(42) Bossuet-Greif, N.; Vignard, J.; Taieb, F.; Mirey, G.; Dubois, D.; Petit, C.; Oswald, E.; Nougayrede, J. P. The colibactin genotoxin generates DNA interstrand cross-links in infected cells. *MBio* **2018**, *9*, No. e02393-17.
(43) 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* **2018**, *57*, 6391.
(44) 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.* **2016**, *138*, 5426.
(45) Wilson, M. R.; Jiang, Y.; Villalta, P. W.; Stornetta, A.; Boudreau, P. D.; Carra, A.; Brennan, C. A.; Chun, E.; Ngo, L.; Samson, L. D.; Engelward, B. P.; Garrett, W. S.; Balbo, S.; Balskus, E. P. The human gut bacterial genotoxin colibactin alkylates DNA. *Science* **2019**, *363*, No. eaar7785.
(46) (a) Jiang, Y.; Stornetta, A.; Villalta, P. W.; Wilson, M. R.; Boudreau, P. D.; Zha, L.; Balbo, S.; Balskus, E. P. Reactivity of an unusual amidase may explain colibactin's DNA cross-linking activity. *J. Am. Chem. Soc.* **2019**, *141*, 11489. (b) Jiang, Y.; Stornetta, A.; Villalta, P. W.; Wilson, M. R.; Boudreau, P. D.; Zha, L.; Balbo, S.; Balskus, E. P. Reactivity of an unusual amidase may explain colibactin's DNA cross-linking activity. *J. Am. Chem. Soc.* **2019**, *141*, 11489.
(47) (a) Zhu, S. Ocimum sanctum natural product derivatives with antimalarial activity. U.S. Patent 7,851,508, 2010. (b) Zhu, S. Small molecules with antiprotozoal activity. U.S. Patent 20100292264A1, 2010.
(48) For a review, see Reissig, H.-U.; Hirsch, E. Donor-acceptor substituted cyclopropanes: Synthesis and ring opening to 1,4-dicarbonyl compounds. *Angew. Chem., Int. Ed. Engl.* **1980**, *19*, 813.
(49) Willoughby, P. H.; Jansma, M. J.; Hoye, T. R. A guide to small-molecule structure assignment through computation of (1H and 13C) NMR chemical shifts. *Nat. Protoc.* **2014**, *9*, 643.
(50) Jorgensen, W. L.; Tirado-Rives, J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.* **2005**, *26*, 1689.
(51) Wiitala, K. W.; Hoye, T. R.; Cramer, C. J. Hybrid density functional methods empirically optimized for the computation of 13C and 1H chemical shifts in chloroform solution. *J. Chem. Theory Comput.* **2006**, *2*, 1085.
(52) dos Santos Torres, Z. E.; Silveira, E. R.; Rocha e Silva, L. F.; Lima, E. S.; de Vasconcellos, M. C.; de Andrade Uchoa, D. E.; Filho, R. B.; Pohlit, A. M. Chemical composition of aspidosperma ulei markgr. And antiplasmodial activity of selected indole alkaloids. *Molecules* **2013**, *18*, 6281.
(53) For a review, see Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. Computational prediction of 1H and 13C chemical shifts: A useful tool for natural product, mechanistic, and synthetic organic chemistry. *Chem. Rev.* **2012**, *112*, 1839.
(54) Kim, L. J.; Ohashi, M.; Zhang, Z.; Tan, D.; Asay, M.; Cascio, D.; Rodriguez, J. A.; Tang, Y.; Nelson, H. M. Prospecting for natural products by genome mining and microcrystal electron diffraction. *Nat. Chem. Biol.* **2021**, *17*, 872.
(55) For a review, see Molinski, T. F. Microscale methodology for structure elucidation of natural products. *Curr. Opin. Biotechnol.* **2010**, *21*, 819.
(56) Burns, D. C.; Reynolds, W. F. Minimizing the risk of deducing wrong natural product structures from NMR data. *Magn. Reson. Chem.* **2021**, *59*, 500.
(57) Gyöngyösi, T.; Nagy, T. M.; Kövér, K. E.; Sørensen, O. W. Distinguishing between two- and three-bond correlations for all 13C multiplicities in heteronuclear NMR spectroscopy. *Chem. Commun.* **2018**, *54*, 9781.
(58) Martin, G. E. In *Annual Reports on NMR Spectroscopy*; Webb, G. A., Ed.; Academic Press: 2011; Vol. *74*, Chapter 5, pp 215.
(59) McAlpine, J. B.; Chen, S.-N.; Kutateladze, A.; MacMillan, J. B.; Appendino, G.; Barison, A.; Beniddir, M. A.; Biavatti, M. W.; Bluml, S.; Boufridi, A.; Butler, M. S.; Capon, R. J.; Choi, Y. H.; Coppage, D.; Crews, P.; Crimmins, M. T.; Csete, M.; Dewapriya, P.; Egan, J. M.; Garson, M. J.; Genta-Jouve, G.; Gerwick, W. H.; Gross, H.; Harper, M. K.; Hermanto, P.; Hook, J. M.; Hunter, L.; Jeannerat, D.; Ji, N.-Y.; Johnson, T. A.; Kingston, D. G. I.; Koshino, H.; Lee, H.-W.; Lewin, G.; Li, J.; Linington, R. G.; Liu, M.; McPhail, K. L.; Molinski, T. F.; Moore, B. S.; Nam, J.-W.; Neupane, R. P.; Niemitz, M.; Nuzillard, J.-M.; Oberlies, N. H.; Ocampos, F. M. M.; Pan, G.; Quinn, R. J.; Reddy, D. S.; Renault, J.-H.; Rivera-Chávez, J.; Robien, W.; Saunders, C. M.; Schmidt, T. J.; Seger, C.; Shen, B.; Steinbeck, C.; Stuppner, H.; Sturm, S.; Taglialatela-Scafati, O.; Tantillo, D. J.; Verpoorte, R.; Wang, B.-G.; Williams, C. M.; Williams, P. G.; Wist, J.; Yue, J.-M.; Zhang, C.; Xu, Z.; Simmler, C.; Lankin, D. C.; Bisson, J.; Pauli, G. F. The value of universally available raw NMR data for transparency, reproducibility, and integrity in natural product research. *Nat. Prod. Rep.* **2019**, *36*, 35.
(60) Robien, W. The advantage of automatic peer-reviewing of 13C-NMR reference data using the csearch-protocol. *Molecules* **2021**, *26*, 3413.
(61) *Spartan 20*; Wavefunction, Inc.: Irvine, CA.
(62) For a recent review, see Panter, F.; Bader, C. D.; Müller, R. Synergizing the potential of bacterial genomics and metabolomics to find novel antibiotics. *Chem. Sci.* **2021**, *12*, 5994.

NOTE ADDED AFTER ASAP PUBLICATION

This paper originally published ASAP on May 23, 2023. Figure 12 was revised and a new version reposted on May 25, 2023.

https://doi.org/10.1021/acs.accounts.3c00183
Acc. Chem. Res. 2023, 56, 1656−1668