

Pattern-Based Genome Mining Guides Discovery of the Antibiotic Indanopyrrole A from a Marine Streptomyces

Douglas Sweeney,[▽] Alexander Bogdanov,[▽] Alexander B. Chase, Gabriel Castro-Falcón, Alma Trinidad-Javier, Samira Dahesh, Victor Nizet, and Paul R. Jensen*



Cite This: *J. Nat. Prod.* 2024, 87, 2768–2778



Read Online

ACCESS |



Metrics & More

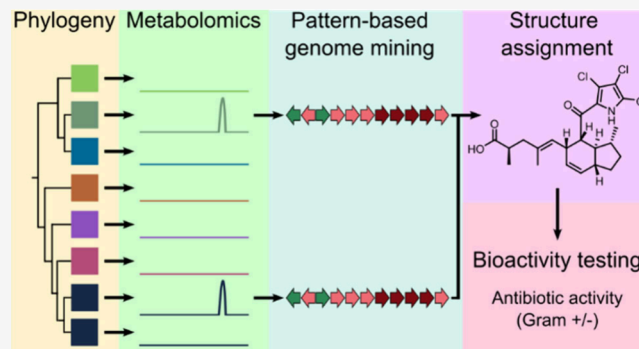


Article Recommendations



Supporting Information

ABSTRACT: Terrestrial actinomycetes in the genus *Streptomyces* have long been recognized as prolific producers of small-molecule natural products, including many clinically important antibiotics and cytotoxic agents. Although *Streptomyces* can also be isolated from marine environments, their potential for natural product biosynthesis remains underexplored. The MAR4 clade of largely marine-derived *Streptomyces* has been a rich source of novel halogenated natural products of diverse structural classes. To further explore the biosynthetic potential of this group, we applied pattern-based genome mining, leading to the discovery of the first halogenated pyrroloketoidane natural products, indanopyrrole A (1) and B (2), and the bioinformatic linkage of these compounds to an orphan biosynthetic gene cluster (BCG) in 20 MAR4 genomes. Indanopyrrole A displays potent broad-spectrum antibiotic activity against clinically relevant pathogens. A comparison of the putative indanopyrrole BGC with that of the related compound indanomycin provides new insights into the terminal cyclization and offloading mechanisms in pyrroloketoidane biosynthesis. Broader searches of public databases reveal the rarity of this BGC while also highlighting opportunities for discovering additional compounds in this uncommon class.



INTRODUCTION

Twenty-eight percent of drugs approved by the FDA in 2021 contained at least one halogen atom.¹ The presence of these electronegative atoms can enhance compound bioactivity,² as observed in the antibiotic vancomycin³ and the anticancer drug salinosporamide A.⁴ The biological activity of halogenated natural products makes them attractive targets for drug discovery, and the abundance of halogen atoms in the marine environment offers a unique opportunity for the discovery of new halogenated metabolites. In 2015, it was estimated that more than 5000 halogenated natural products had been discovered, with the majority being produced by marine organisms, including bacteria.⁵ Among marine bacteria, the MAR4 group of marine-derived *Streptomyces* have yielded a wide diversity of halogenated natural products,⁶ including phenazines,⁷ nitropyrroles,^{8,9} and tetrahydroxynaphthalene (THN)-derived molecules.^{10–14} In addition to these diverse halogenated compounds, genome mining has revealed halogenases in orphan MAR4 biosynthetic gene clusters (BGCs), suggesting that additional halogenated metabolites await discovery.^{15,16}

One approach to facilitate natural product discovery involves the pairing of genomic and metabolomic data collected from closely related strains; this is a process known as pattern-based genome mining or metabologenomics.^{17,18} This method

correlates gene cluster families with ions detected by mass spectrometry (MS), enabling the connection of metabolites to orphan BGCs, as shown in the marine actinomycete *Salinispora*¹⁷ and other bacteria.¹⁸ The technique is also applicable to BGC subclusters, as shown for the pyrrole-containing compounds chlorizidine,¹⁹ armeniaspirol,²⁰ and marinopyrrole.²¹ Recent advances in pattern-based genome mining include automated tools such as NPLinker²² and NPOmix,²³ which use correlation-based statistics and machine learning to computationally link metabolites to their cognate BGCs. Halogenated natural products are especially well-suited for pattern-based genome mining due to their distinct isotopic signatures and biosynthetic genes that can be used as “hooks” to detect candidate BGCs. Linking metabolomic and biosynthetic sequence data in this way can facilitate the discovery of novel halogenated metabolites, with deeper BGC analyses facilitating structure assignments.²⁴

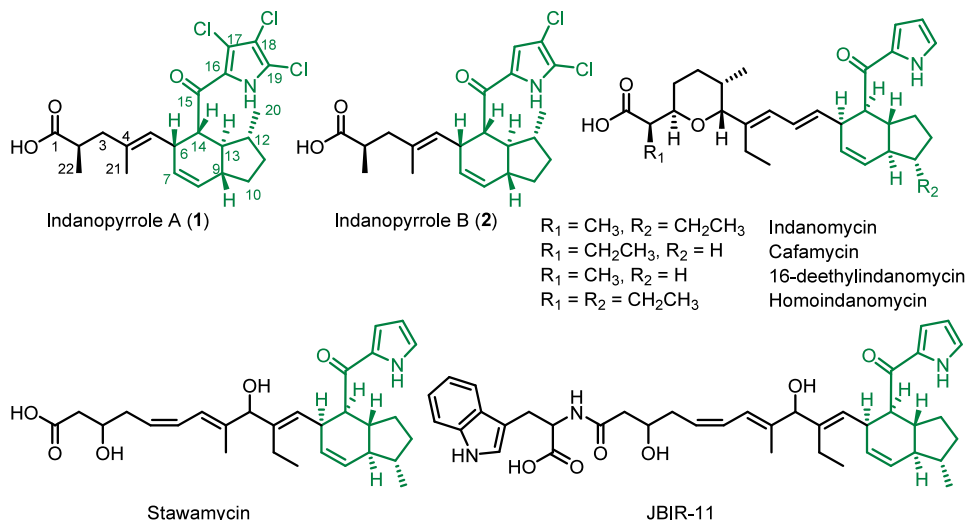
Received: August 13, 2024

Revised: November 15, 2024

Accepted: November 17, 2024

Published: November 22, 2024



Chart 1. Indanopyrroles A and B (1-2) and Previously Reported Pyrroloketoidane Natural Products^a

^aPyrroloketoidane moieties are colored green. Absolute configurations for 1 and 2 are based on bioinformatic prediction.

Table 1. NMR Data of Indanopyrrole A (1)

assignment ^a	δ_C^b	δ_H (J, in Hz)	COSY	HMBC	NOESY
1	181.5, C				
2	39.6, CH	2.44, ddd (9.0, 6.8, 5.7)	H ₂ -3, H ₃ -22	C-1, C-3, C-4, C-22	
3	44.7, CH ₂	a 1.89, dd (13.7, 9.0) b 2.28, dd (13.7, 5.7)	H-2, H-5 (w) H-2, H-5 (w)	C-1, C-2, C-4, C-5, C-22 C-1, C-2, C-4, C-5, C-22	
4	133.7, C				
5	128.3, CH	5.06, d (10.3)	H-3b, H-6, H ₃ -21	C-3, C-6, C-21	H-3a, H-6, H-13
6	38.7, CH	3.65, m	H-5, H-7, H-14	C-7	H-5, H-14, H ₃ -21
7	130.0, CH	5.33, dt (9.7, 3.4)	H-6, H-8, H-9 (w)	C-6, C-9, C-14	
8	130.4, CH	5.84, dt (9.7, 1.8)	H-7	C-6, C-7, C-9, C-10	
9	46.7, CH	2.10, m	H-7 (w), H ₂ -10		H-10b, H-12, H-14
10	28.6, CH ₂	a 1.29 m b 1.81, ddd (16.3, 9.2, 2.1)	H-9, H ₂ -11 H-9, H ₂ -11	C-9, C-11 C-12, C-13	
11	34.4, CH ₂	a 1.34, m b 2.01, ddd (16.9, 12.8, 8.8)	H ₂ -10, H-12 H ₂ -10, H-12	C-20 C-13	
12	39.3, CH	1.61, m	H ₂ -11, H-13, H ₃ -20	C-11, C-13	
13	47.8, CH	1.58, m	H-9, H-12, H-14	C-9, C-12, C-14, C-20	H-5, H ₃ -20
14	53.6, CH	3.88, dd (10.5, 6.7)	H-6, H-13	C-6, C-13, C-15, C-16	H-6, H-9, H-12
15	189.8, C				
16	128.0, C				
17	116.1, C				
18	111.5, C				
19	120.6, C				
20	21.7, CH ₃	0.99, d (6.1)	H-12	C-11, C-12, C-13	H-13
21	15.8, CH ₃	1.31, br s	H-5	C-3, C-4, C-5	H-2, H-6, H-7
22	16.8, CH ₃	0.98, d (6.8)	H-2	C-1, C-2, C-3	

^aAll assignments are based on extensive 1D and 2D NMR measurements (COSY, HSQC, HMBC). ^bMultiplicities determined by HSQC.

Pyrroloketoidanes are characterized by indane and pyrrole systems bridged via a ketone. Compounds containing these ring systems are rare and are often biologically active, making them logical targets for natural product discovery. Prior to this study, only six pyrroloketoidanes had been reported, none of which were halogenated (Chart 1). These compounds, all of bacterial origin, are the products of polyketide synthase (PKS) BGCs. The first pyrroloketoidane described was indanomycin, which contains a tetrahydropyran and was discovered in 1979 from *Streptomyces antibioticus* NRRL 8167.²⁵ Indanomycin has activity against Gram-positive bacteria²⁶ and acts as an

ionophore.^{27,28} Later work has shown that indanomycin has insecticidal²⁹ and antiviral activities.³⁰ The indanomycin analogue cafamycin was identified from a *Streptomyces* species in 1987.³¹ Shortly thereafter, 16-deethylindanomycin was isolated from *Streptomyces setonii* A80394A and reported to be active against Gram-positive bacteria and protist parasites.³² A compound with the same structure as 16-deethylindanomycin was published in 1990 under the name "omomycin" and reported to elevate cyclic guanosine monophosphate levels in rat heart cells.³³ The pyrroloketoidane containing compound homoidanomycin was described in a patent from 1989

Table 2. NMR Data of Indanopyrrole B (2)

assignment ^a	δ_C^b	δ_H (J in Hz)	COSY	HMBC
1	181.3, C			
2	39.2, CH	2.45, ddd (5.4, 6.8, 9.1)	H ₂ -3, H ₃ -22	C-1, C-3, C-4, C-22
3	44.7, CH ₂	<i>a</i> 1.92, dd (9.1, 13.4) <i>b</i> 2.29, dd (5.4, 13.4)	H-2, H-5 (w) H-2, H-5 (w)	C-1, C-2, C-4, C-5, C-21, C-22 C-1, C-2, C-4, C-5, C-21, C-22
4	133.1, C			
5	128.2, CH	5.10, br d (9.3)	H-3b, H-6, H ₃ -21	C-3, C-6 (w), C-21
6	40.2, CH	3.50 m	H-5, H-7, H-14	C-5
7	129.9, CH	5.33, dt (3.0, 3.2, 9.8)	H-6, H-8, H-9 (w)	C-6 (w), C-9
8	130.0, CH	5.83, br d (9.8)	H-7	C-6, C-9, C-10, C-13
9	46.1, CH	2.12, m	H-7 (w), H ₂ -10	C-8 (w)
10	27.9, CH ₂	<i>a</i> 1.27, m <i>b</i> 1.81, m	H-9, H ₂ -11 H-9, H ₂ -11	C-8, C-9, C-11 C-9, C-12, C-13
11	34.4, CH ₂	<i>a</i> 1.32, m <i>b</i> 2.01, m	H ₂ -10, H-12 H ₂ -10, H-12	C-10, C-20 C-10, C-13
12	38.9, CH	1.59, m	H ₂ -11, H-13, H ₃ -20	C-11, C-13
13	47.7, CH	1.51, dt (10.4, 10.6)	H-9, H-12, H-14	C-6, C-9, C-12, C-14, C-20
14	53.2, CH	3.49, m	H-6, H-13	C-6, C-13, C-15, C-16
15	189.8, C			
16	131.5, C			
17	115.8, CH	7.04, s		C-15, C-16, C-19
18	n.o.			
19	121.2, C			
20	20.8, CH ₃	0.92, d (6.6)	H-12	C-11, C-12, C-13
21	15.5, CH ₃	1.33, br s	H-5	C-3, C-4, C-5
22	16.2, CH ₃	1.00, d (6.8)	H-2	C-1, C-2, C-3

^aAll assignments are based on extensive 1D and 2D NMR measurements (COSY, HSQC, HMBC). n.o. = not observed. ^bMultiplicities determined by HSQC.

without any information about activity.³⁴ Stawamycin inhibits Epstein–Barr viral transcription factor BZLF1 binding to DNA and was the first indanopyrrole analogue described to lack a tetrahydropyran.³⁵ Finally, JBIR-11 was reported in 2008 from *Streptomyces viridochromogenes* subsp. *sulfomycin* NBRC 13830. This unusual analogue of stawamycin contains a tryptophan moiety and was shown to be cytotoxic to HT-1080 fibrosarcoma cells.³⁶

Here, we describe the discovery of novel dichlorinated and trichlorinated pyrroloketoidane antibiotics using paired metabolomic and genomic data sets. The candidate BGC provides new insights into pyrroloketoidane biosynthesis, while its distribution in bacterial genomes reveals opportunities for additional compound discovery.

RESULTS AND DISCUSSION

Isolation and Structure Elucidation. A search for novel, halogenated natural products from *Streptomyces* strains in the MAR4 clade led to the detection of a compound with an m/z of 458.1060 $[M + H]^+$ and an isotopic pattern characteristic of a trichlorinated molecule in the culture extract of strain CNX-425 (Figure S1). A Dictionary of Natural Products database search retrieved no matches, suggesting that it represented a new compound. The compound, which we have named indanopyrrole A (1), was isolated and structurally characterized as described below. During the isolation of 1, we obtained and characterized a minor related compound (2) that shared the same chromophore and was named indanopyrrole B (Chart 1).

The molecular formula (MF) of 1 $C_{22}H_{26}Cl_3NO_3$ with nine degrees of unsaturation was determined from accurate mass (HR-ESI-TOF-MS) measurement. NMR experiments (¹H,

¹³C, COSY, HSQC and HMBC) in CD₃OD led to the unambiguous elucidation of the planar structure (Table 1). The proton spectrum showed the presence of two methyl doublets (CH₃-20 and CH₃-22; δ_H 0.99, J = 6.1 Hz and 0.98, J = 6.8 Hz) and one methyl singlet (CH₃-21; δ_H 1.31) attached to an olefin. Additionally, three sp^2 resonances at 5.06 (H-5, d, J = 10.3 Hz), 5.33 (H-7, dt, J = 9.7, 3.4 Hz), and 5.84 ppm (H-8, dt, J = 9.7, 1.8 Hz) accounted for two olefins. In the ¹H–¹H COSY spectrum, an extensive spin system from H-5 to H-14 established the tetrahydroindane core of the molecule (Figure S2A), the position of methyl C-12, and the two substituents at C-6 and C-14. The side chain at C-6 was established using ¹H–¹³C HMBC, which showed long-range correlations between the methyl protons H₃-21 to C-3, C-4, and C-5 and from H₃-22 to the carboxylic acid carbonyl C-1 (δ_C 182.7 ppm), C-2, and C-3. The substructure composed of the tetrahydroindane core and the C-6 substituent accounted for 17 carbons, 25 protons, two oxygens, and five degrees of unsaturation. The remaining C₅HCl₃NO was suggested to be a trichlorinated pyrrole-2-carbonyl substituent on C-14 based on a fragment of 195.9127 m/z in the MS/MS spectrum of 1 (Figure S3).^{32,35–37} The UV maximum absorbance at 295 nm provided further support for the pyrrole moiety. HMBC correlations from H-14 to carbonyl C-15 and sp^2 hybridized carbon C-16 (δ_C 128.0 ppm) established the connection of the trichloropyrrole moiety to the indane core via a ketone bridge. Analysis of the ¹³C NMR spectrum revealed the chemical shifts for the remaining chlorinated carbons (C-17, C-18, and C-19) that were not observed by indirect detection experiments (HSQC and HMBC) and secured the planar structure of indanopyrrole A (1, Chart 1). The configuration of the C-4, C-5 olefin was assigned as *trans* (*E*) based on the

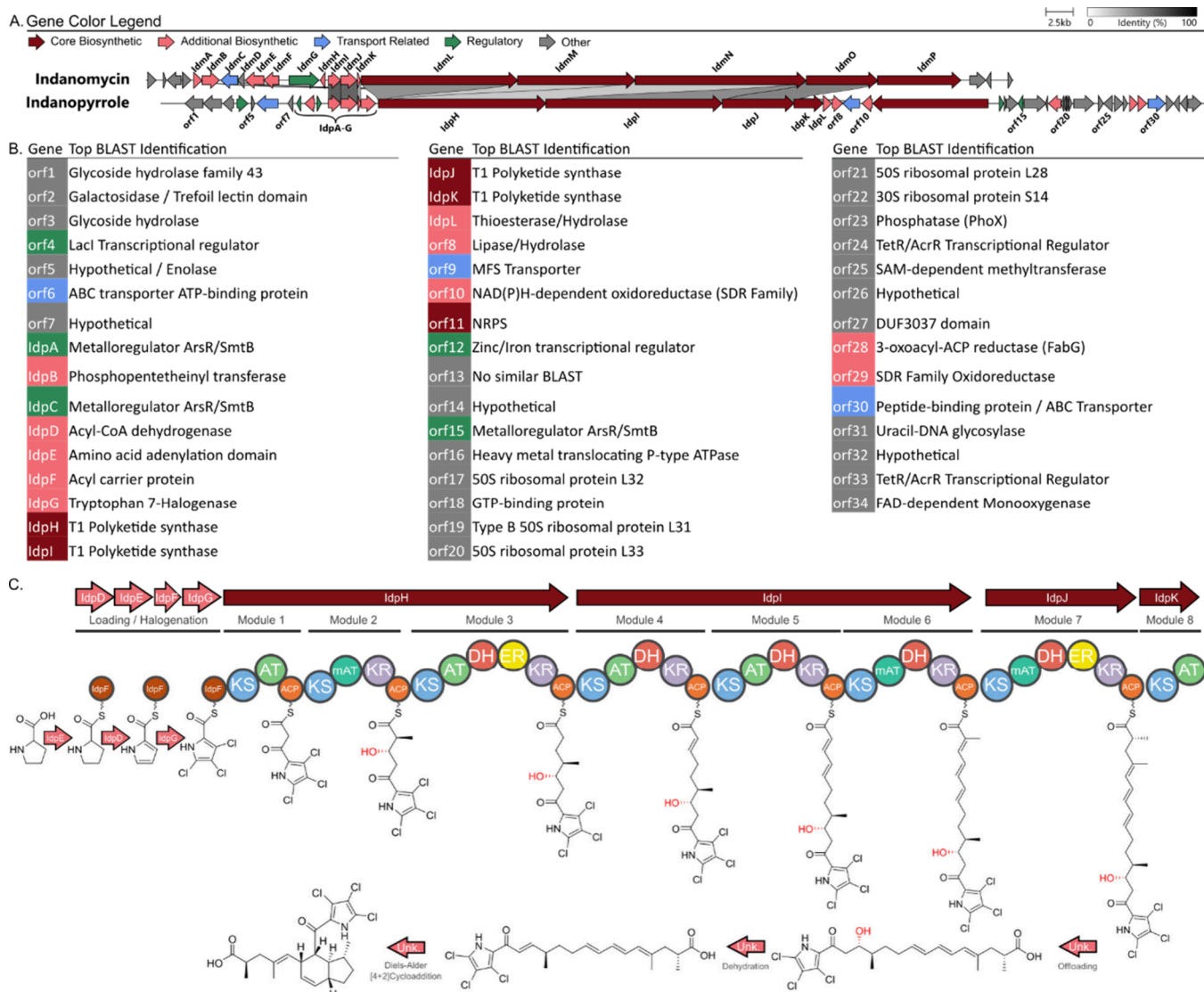


Figure 1. Candidate indanopyrrole BGC and the proposed biosynthesis. (A) Comparison of the candidate indanopyrrole BGC (*idp*) from *Streptomyces* sp. CNX-425 with the indanomycin BGC (*indm*, MIBiG No. BGC0000079). (B) *idp* gene annotations based on top BLAST matches to the NCBI NR/NT database. (C) Proposed indanopyrrole biosynthesis in concordance with the *idp* BGC. Cryptic dehydration of the hydroxy group (in red) is predicted to afford an alkene dienophile intermediate facilitating an intramolecular Diels–Alder ([4 + 2] cycloaddition) reaction to form the indane.

shielded ^{13}C resonance of the methyl substituent $\text{CH}_3\text{-21}$ (δC 15.7 ppm) and on a clear NOE correlation between H-5 and H-3a. The relative configuration of the five stereocenters on the indane was established using a NOESY experiment (NOE correlations observed between H-6, H-9, and H-14, as well as from H-13 to H₃-20) and is in accordance with previously reported pyrroloketoidanes (Chart 1). The absolute configurations at C-2 and C-12 in the linear biosynthetic precursor were predicted bioinformatically as 2R and 12S (12R after indane formation), based on an analysis of the candidate biosynthetic gene cluster (see below). This, in turn, led to the absolute configuration of all stereocenters in **1** (2R, 6R, 9R, 12R, 13S, and 14R; see Figures S3B and S3C) and differences in the indane core, compared to other pyrroloketoidanes (Chart 1).

HR-ESI-TOFMS/MS analyses of **2** revealed the molecular formula $\text{C}_{22}\text{H}_{27}\text{Cl}_2\text{NO}_3$ and a prominent dichloropyrrole containing MS/MS fragment (m/z 161.9527 calcd for

$\text{C}_5\text{H}_2\text{Cl}_2\text{NO}^+$, 161.9513, -8.65 ppm; see Figures S4 and S5). These data indicated that compound **2** is a dichlorinated analogue of compound **1**. To assign the positions of the two chlorines on the pyrrole, we recorded ^1H , HSQC, and HMBC NMR spectra using a 600 MHz NMR instrument equipped with a 5 mm cryoprobe (Table 2). As expected, the ^1H spectrum was almost identical with that of **1**, except for an additional sharp singlet at 7.04 ppm, the shielded H-6 and H-14 resonances (from 3.65 to 3.50 and 3.88 to 3.49 ppm, respectively), and the slightly shielded H₃-20 resonance from 0.99 to 0.92 ppm. In the HMBC spectrum of **2**, we observed a long-range correlation from the proton resonance at 7.04 ppm to the C-15 carbonyl, which established the structure of indanopyrrole B (**2**) as a 17-deschloro derivative of **1**.

Pattern-Based Identification of the Candidate Indanopyrrole (*idp*) BGC. Based on biosynthetic precedent for the pyrrole containing natural products indanomycin,²⁷ chlorizidine,¹⁹ armeniaspirol,²⁰ and marinopyrrole,²¹ we

anticipated that indanopyrrole biosynthesis is initiated using a proline-derived dichloropyrrole or trichloropyrrole starter unit. As such, a DIAMOND BLAST³⁸ database containing 42 MAR4 genomes was created using cblaster³⁹ and queried for the proline adenylyltransferase (*idmJ*) and prolyl-carrier protein dehydrogenase (*idmI*) gene sequences from the indanomycin (*idm*) BGC (MIBiG No. BGC0000079).⁴⁰ This analysis returned 33 hits in 25 MAR4 genomes, all within the bounds of BGCs called by AntiSMASH 5.0.⁴¹

Surprisingly, three distinct BGCs containing the pyrrole biosynthetic hooks were identified in the indanopyrrole producing strain CNX-425. To narrow down the candidates, we queried extracts of the same 42 strains for the indanopyrrole A molecular ion (m/z of 458.1060 ± 0.01) and detected it in 7. These 7 strains shared one BGC with the pyrrole biosynthetic hooks, and they had the highest antiSMASH similarity score (39%) with *idm*. Using this pattern-based genome mining approach,¹⁷ we identified this conserved BGC as the top candidate for indanopyrrole biosynthesis and named it *idp* (Figure 1A, S6). In total, 15 of the 42 MAR4 genomes contained full-length *idp* BGCs, while five additional strains contained partial BGCs located on contig edges (Figure S6). Among the seven strains found to produce indanopyrroles, two contained partial BGCs, suggesting that the truncations are artifacts of the sequencing and assembly process. While some strain-level variation among *idp* BGCs was observed (Figure S6), there is little evidence to suggest that some may be nonfunctional.

Noting low production in CNX-425, we scaled up the cultivation of additional *idp* containing MAR4 strains and found the highest production (5 mg of indanopyrrole A and 0.5 mg of indanopyrrole B from a 3 L culture) in strain CNY-716. Interestingly, indanopyrrole was not originally observed in this strain (Figure S6), but was instead observed in a scale-up-fermentation following the addition of XAD-7 resin (Figure S7; see the Methods section). The additional compound obtained from this strain facilitated the structure elucidation and was used for expanded antibiotic screening.

Biosynthesis of the Indanopyrroles. A comparison of the *idp* and *idm* BGCs provides insight into the structural differences between their small molecule products (Figure 1A–B). As expected, both BGCs contain homologous genes encoding proline adenylyltransferases (*idpE/idmJ*), proline carrier proteins (*idpF/idmK*), and dehydrogenases (*idpD/idmI*) to account for the generation of the pyrrole moieties from proline (Figure 1C). *idp* then diverges by the presence of the halogenase gene *idpG*, which was annotated as a tryptophan halogenase. Unlike other BGCs linked to halogenated MAR4 natural products, *idp* does not contain a vanadium haloperoxidase encoding gene. Instead, *idpG* shares homology with halogenase genes in the BGCs of other chlorinated pyrrole containing natural products^{19,20,42} and likely accounts for the trichlorinated and dichlorinated pyrrole moieties observed in **1** and **2**, respectively. The chlorinated pyrrole generated from *idpD–G* then serves as the starter unit for seven polyketide extensions encoded by the three T1PKS genes (*idpH–J*). Notably, a NaPDos2^{43,44} analysis of the module 1 (loading module) ketosynthase (KS) domain within *IdpH* places it in a clade with other pyrrole-accepting KSs, further supporting the functional prediction for the starter unit (Figure S8).

Following starter unit selection, the acyltransferase (AT) domains associated with modules 2, 6, and 7 are predicted to

select for methylmalonyl-CoA. Based on conserved tryptophan and histidine residues, the ketoreductase (KR) domain within module 2 is assigned to the A2-type, which produce 2S-methyl-3S-hydroxy intermediates (Figure S9).⁴⁵ Module 3 (*idpH*) contains the full suite of domains (KR, DH, and ER) to generate the alkane, while modules 4–6 (*idpI*) contain KR and dehydratase (DH) domains that would afford a conjugated triene. Module 7 contains the full suite of domains to generate a branched alkane that, based on the lack of a conserved tyrosine residue in the active site of the *idpJ* enoylreductase (ER) domain, is predicted to install R configuration (Figure S10).⁴⁵

The terminal T1PKS module (module 8, *idpK*) is comprised of a KS and AT domain and was observed in all *idp* BGCs except for one that appears to be truncated before this gene. This unusual domain organization resembles the terminal PKS module reported for *idm*,²⁷ which differs only by the presence of a terminal cyclase domain. In both *idp* and *idm*, the AT domain within the terminal module lacks the active site residues required for selection and loading of the extender unit and is therefore predicted to be nonfunctional (Figure S11). However, *idp* further differs from *idm* in that the KS domain of the terminal module lacks the active site residues required for decarboxylative condensation and is therefore also predicted to be inactive (Figure S12). The inability of this module to support chain extension is supported by the structures of **1** and **2**. A stand-alone thioesterase (TE) domain (*idpL*) located immediately after *idpK* is homologous to *idmA*, which was presumed to be associated with chain release from the megasynthase during indanomycin biosynthesis.⁴⁶ However, an analysis using the THYME thioesterase database of the *idpL* and *idmA* genes places them in the TE18 family of “editing” type II TEs, which remove prematurely decarboxylated extender units, stalled intermediates, or improperly edited CoA-bound starter units.^{47,48} Thus, it does not appear that *idpL* is involved in chain release. In the case of indanomycin, offloading has been proposed to involve the terminal cyclase domain within *idmP*,^{27,49} however, the lack of this domain in *idpK* suggests either a different mechanism for **1** and **2** or that the cyclase is not involved (Figure S13). Finally, while antiSMASH calls for a larger BGC, manual analyses have led us to propose that the *idp* BGC is best represented by *idpA–idpL*. As in indanomycin biosynthesis, questions remain about how the linear precursor predicted from the BGC is offloaded and cyclized to yield the final pyrrolotoketoidane products.

Indane Formation. Further comparison of the *idm* and *idp* BGCs and their products provides insight into the formation of the indane. In the case of indanomycin, it was suggested that *idmH* is an indane cyclase catalyzing a Diels–Alder [4 + 2] cycloaddition reaction. However, genetic knockout experiments lowered but did not abolish compound production⁵⁰ and molecular modeling of the crystal structure did not reveal the potential for enzymatic activity.⁵¹ Notably, only one gene with similarity to *idmH* was detected among the indanopyrrole producing MAR4 strains (CNY-716, 36% amino acid similarity), and it was not located within any of the BGCs identified in that genome. Thus, the *idmH* mediated mechanism of indane ring formation proposed for indanomycin does not appear to apply to indanopyrrole. While several unannotated *idp* open reading frames could account for indane formation, it is intriguing to consider that the conserved KS domain associated with the non-elongating, terminal PKS module observed in *idm* and all *idp* gene clusters may be

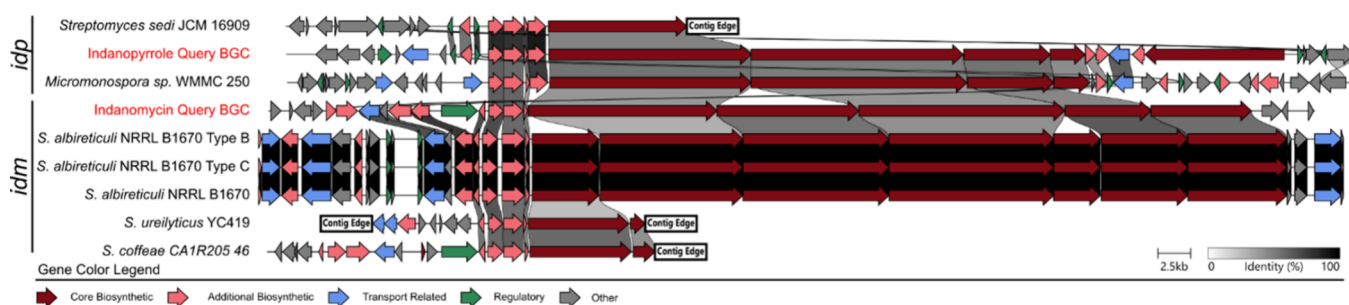


Figure 2. Distribution of *idp* and *idm*. The seven closest *idp* and *idm* homologues identified from GenBank searches highlight the rarity of these BGCs and present opportunities for discovery. A full list of BGC hits is provided in the [Supporting Information \(Figure S15\)](#).

involved. KS functional diversification is well-documented and includes the non-elongating, terminal KS domain in salinosporamide A biosynthesis, which was recently shown to catalyze the formation of an alkene which then facilitates intramolecular carbon–carbon bond formation and the cyclized end product.⁵² Notably, both *idm* and *idp* lack a dehydratase domain in the second PKS module, which would provide the alkene dienophile required for indane ring formation via a Diels–Alder [4 + 2] cycloaddition reaction.^{27,53–56} It is intriguing to speculate that the terminal, non-elongating KS domains observed in both *idm* and *idp* catalyze this cryptic dehydration step.

To address the potential neofunctionalization of the *idpK* KS domain, we modeled the 3D protein structure using the open-source implementation of AlphaFold 2⁵⁷ in ColabFold.⁵⁸ The model showed strong structural homology with the crystal structure of the KS domain from 6-deoxyerythronolide B synthase (DEBS), with the active site residues collocated within the substrate binding pocket (Figure S14A). Molecular docking studies using the AMDock suite⁵⁹ showed that indanopyrrole A (**1**) fits within the predicted substrate binding pocket of the *idpK* KS domain. In the model, the carboxylic acid moiety of **1** is adjacent to the active site cysteine (Cys-167). This position is homologous to DEBS Cys-211, which forms a thioester bond with the polyketide intermediate generated by the preceding PKS module (Figure S14A, inset). This spatial orientation aligns the active site histidine-302 above C-14 in the indanopyrrole structure. Positioning this residue above the carbon alpha to the carbonyl group could allow the active site histidine to abstract a hydrogen and initiate a cascade reaction that ends with the elimination of water from the C-13 hydroxy group thus accounting for the cryptic dehydration step and the formation of the alkene dienophile (Figure S14B). Finally, the shape of the substrate binding pocket conforms to the shape of indanopyrrole A, suggesting that it could promote a conformational change in the linear precursor that facilitates a spontaneous Diels–Alder reaction following dehydration (Figure S14A, inset). However, we were unable to model the proposed linear precursor within the active site to address this hypothesis. It should be noted that the absolute configurations of the indane ring systems in **1** and **2** are opposite from the indanomycins, which could account for the low KS domain sequence similarity and the differences in active site residues. While speculative, assigning dehydratase activity to the terminal KS domains in *idm* and *idp* would indicate a new mechanism for indane formation in indanopyrroles and other pyrroloketoidanes.

Gene Cluster and Metabolite Distribution. To more broadly explore *idp* and *idm* distributions, we used cblaster⁶⁰ to

remotely query the NCBI nr database, using the respective adenylyltransferase (*idpE/idmJ*) and proline dehydrogenase (*idpD/idmJ*) genes, which returned 3283 (*idp*) and 3198 (*idm*) hits. After filtering to include only those that contained at least one gene with >70% homology to a PKS gene within *idp* or *idm* and duplicate removal, 40 BGCs were identified and manually assigned to seven BGC groups and nine singletons based on gene synteny (Figure S15). AntiSMASH analyses revealed top matches to the indanomycin, nargenicin, and calcimycin BGCs along with three with no known product (Figure S15). Three complete *idm* BGCs were identified in *Streptomyces albireticuli* strains, NRRL B1670 (JAJQQQ010000001.1), NRRL B1670 Type B (JAJQQR010000001.1), and NRRL B1670 Type C (JAJQSS010000003.1) (Figure 2), all of which lacked the terminal cyclase domain observed in *idmP* (Figure S13). Additionally, two fragmented *idm* BGCs were detected on contig edges in *S. ureilyticus* (NZ_JAAKZX010000175.1) and *S. coffeae* (NZ_JAERRF010000046.1). One putative full-length *idp* BGC was observed in *Micromonospora* sp. WMMC 250 (NZ_JAPZBJ010000002.1) (Figure 2) and one fragmented BGC was identified on a contig edge in the genome of *Streptomyces sedii* JCM 16909 (VDGT01000017.1). These results indicate that despite being detected in genetically distant actinobacterial lineages *Streptomyces* and *Micromonospora*, both *idm* and *idp* are rare in publicly available sequence data.

We next used the Mass Spectrometry Search Tool⁶¹ (MASST) within the Global Natural Products Social Molecular Networking⁶² (GNPS) platform to search for the MS/MS spectrum of indanopyrrole A within public datasets. After detecting no hits using the default parameters, we lowered the minimum cosine score to 0.6, the minimum matched fragments to 2 and detected 71 hits when allowing for analogue searching. Upon manual inspection, none of the *m/z* values or fragmentation spectra matched indanopyrrole A. The lack of any significant matches among 2709 public datasets highlights the rarity of these natural products.

Bioactivity. Indanopyrrole A demonstrated significant antibacterial activity, with a minimum inhibitory concentration (MIC) of 4 µg/mL against membrane-deficient *Escherichia coli* lptD4213⁶³ (Table 3). Broader testing revealed potent activity against several clinically relevant Gram-positive pathogens, including methicillin-resistant *Staphylococcus aureus* TCH1516 (MIC = 2 µg/mL), group A *Streptococcus* M1T1 (MIC = 4 µg/mL), vancomycin-resistant *Enterococcus faecium* DAPS (MIC = 2 µg/mL), and methicillin-resistant *Staphylococcus epidermidis* (MIC = 4 µg/mL) (Table 1). To the best of our knowledge, **1** is the first pyrroloketoidane natural product to show activity

Table 3. Antibacterial Activities of Indanopyrroles A and B Reported as Minimum Inhibitory Activity^a

strains tested	Gram ±	Antibacterial Activity (μg/mL)	
		indanopyrrole A	indanopyrrole B
MRSA TCH1516	+	2	>16
Group A <i>Streptococcus</i> MIT1	+	4	>16
vancomycin-resistant <i>Enterococcus faecium</i> DAPS	+	2	>16
MRSE (<i>Staphylococcus epidermidis</i>)	+	4	NT ^a
<i>Haemophilus influenzae</i> 1	−	1	NT ^a
<i>Haemophilus influenzae</i> 2	−	2	NT ^a
<i>Escherichia coli</i> LptD4213	−	4	NT ^a
<i>Escherichia coli</i> K1 RS218	−	>32	>16
<i>Acinetobacter baumannii</i> 5075	−	>32	NT ^a

^aNT = not tested. Positive control (colistin) MICs: *Escherichia coli* K1 RS218 (0.5 μg/mL) and *Acinetobacter baumannii* 5075:1 μg/mL.

against Gram-negative organisms, showing MIC values of 1–2 μg/mL against *Haemophilus influenzae* (Table 1). In contrast, indanopyrrole B (2) was inactive at all concentrations tested, likely due to differences in chlorination between the two compounds. The cytotoxicity of 1 was measured at 16 μg/mL after 24 h of exposure in an A549 cell line viability assay (Figure S16), indicating an antibiotic therapeutic index of 4–8 (ratio of antibiotic to cytotoxic activity).

Given the antibiotic activity of indanopyrrole A, it is unclear how producing strains maintain resistance. However, a drug-resistant transporter related to the *Burkholderia thailandensis* EmrB multidrug efflux pump is conserved in all *idp* BGCs, suggesting a possible resistance mechanism. Similar mechanisms of drug resistance in Gram-negative organisms such as *Acinetobacter baumannii* and *Escherichia coli* are mediated by upregulation or genomic expansion of efflux pumps and may explain the lack of activity against these strains.^{64,65}

CONCLUSIONS

The MAR4 group of marine actinomycetes has been a prolific source of halogenated natural products.⁶ To further explore their biosynthetic potential, we applied pattern-based genome mining to 42 strains, focusing on identifying new halogenated metabolites and their candidate BGCs. This effort led to the discovery of the pyrroloketoindane natural products indanopyrroles A and B (1 and 2), which are the seventh and eighth pyrroloketoindane natural products described and the only halogenated members of this rare compound class. Comprehensive 1D and 2D NMR spectroscopy facilitated the structure elucidation of 1–2 including their relative configuration, while their absolute configuration was proposed based on bioinformatic prediction of the stereoselective KR and ER domains in the candidate BGC.

Indanopyrrole production was linked to a BGC (*idp*) containing genes associated with chloropyrrole biosynthesis. As in the biosynthesis of the related compound indanomycin, a cryptic dehydration step is predicted to facilitate an intramolecular Diels–Alder cycloaddition that yields the final cyclized product. Comparisons between the *idp* and *idm* BGCs suggest that the terminal, non-elongating KS domains observed in these BGCs could account for this cryptic dehydration. While *in-silico* docking studies support this hypothesis, experimental verification is still needed. A search of the

NCBI nr database revealed seven BGCs with a similarity to *idp* or *idm*, supporting the rarity of pyrroloketoindane natural products. All seven of these BGCs, present in diverse actinomycetes, contain terminal KS domains homologous to *idpK* and *idmP*, supporting the functional role of this domain in the pathway.

Indanopyrrole A (1) exhibits potent, broad-spectrum antibiotic activity and, to the best of our knowledge, is the first compound in this class with reported activity against Gram-negative bacteria. The inactivity of indanopyrrole B (2) emphasizes the importance of the trichlorinated pyrrole moiety for the activity of 1. Unlike the pyrroloketoindane antibiotic indanomycin, which acts as an ionophore for divalent cations, 1 lacks the tetrahydrofuran moiety known to coordinate metals,²⁷ indicating that it may function via a different mechanism. The promising therapeutic index of 1 supports further structure–activity studies and further antibiotic lead development.

This work showcases the value of pattern-based genome mining for discovering new antibiotics and their BGCs within the biosynthetically gifted MAR4 group of marine bacteria. It also expands the breadth of halogenated natural products reported from the MAR4 actinomycetes and provides new insights into potential KS functional diversification and indane ring formation in the rare pyrroloketoindane compound class.

EXPERIMENTAL SECTION

General Experimental Procedures. Optical rotations were recorded on a polarimeter (Jasco, Model P-2000). UV spectra were measured on a spectrophotometer (Jasco, Model V-630). IR spectra were acquired on a spectrometer (Model JASCO FTIR-4100, Jasco Corp., Tokyo, Japan). NMR spectra (1D and 2D) were measured at 23 °C on a JEOL ECZ spectrometer (500 MHz) equipped with a 3 mm ¹H{¹³C} room-temperature probe (JEOL, Akishima, Tokyo, Japan) or on a Bruker Avance III (600 MHz) NMR spectrometer with a 5 mm ¹H{¹³C/¹⁵N} room temperature or cryo probe (Billerica, MA). ¹³C NMR spectrum was recorded on a Varian 500 MHz spectrometer equipped with a 5 mm ¹H{¹³C} XSENS cold probe (Varian, Inc., Palo Alto, CA, USA (now Agilent Technologies)). NMR spectra were referenced to the solvent signals (CHD₂OD, δ_H 3.31 and CD₃OD, δ_C 49.00 ppm). LC-HR-ESIMS was performed on an HPLC system (Agilent, Model 1260 Infinity) that was equipped with a degasser, binary pump, autosampler, and DAD detector, coupled to a QToF device (Agilent, Model 6530 Accurate-Mass QToF) with an ESI source that was coupled with an HPLC system (Agilent, Model 1260 Infinity) and calibrated using the Agilent Reference Calibration Mix. Compounds were isolated on an Agilent HPLC system with a binary pump (Model 1100 G1312A), UV/vis detector (Model 1100 G1315A DAD), autosampler (Model 1100 G1313A), and degasser (Model 1100 G1322A) (all obtained from Agilent Technologies, Santa Clara, CA, USA).

Small-Scale Strain Cultivation and Metabolite Extraction.

All cultures were grown at 28 °C and shaken at 230 rpm with metal springs in A1 media containing potassium bromide (10 g/L starch, 4 g/L yeast extract, 2 g/L peptone, 22 g/L instant ocean, and 0.1 g/L KBr). Cryogenic stocks of 40 MAR4 strains as well as lyophilized material for the two MAR4-type strains purchased from the DSMZ (DSM 41644, DSM 41902) were inoculated as a preculture into 50 mL of media. After 7 days, 0.5 mL was transferred to 50 mL of the same medium for a second seed culture. After 5 days, 5 mL of the second seed culture was frozen for DNA analysis, and 0.5 mL was used to inoculate 50 mL of fresh media. After 4 days, HP-20 resin (1 g, wet weight) was added to each flask. After 3 days of incubation with resin, all cultures were extracted with 50 mL of EtOAc. Organic extracts were separated, dried with anhydrous Na₂SO₄, filtered, concentrated by rotary evaporation, and stored at −20 °C until further analysis.

Large-Scale Cultivation and Extraction of Strain CNY-716. Aliquots (10 mL) of a 50 mL preculture of *Streptomyces* sp. CNY-716 were inoculated into three 2.8 L Fernbach flasks containing 1 L of A1 medium and incubated at 28 °C and 120 rpm. On day 10, activated sterile XAD-7 resin (20 g, ThermoFisher Scientific) was added to each flask, and the cultures were incubated for an additional 4 days. After 14 days, the resin and the cellular material were filtered through cheesecloth, washed with deionized H₂O, and extracted with MeOH (4 × 200 mL). The solvent was removed from the pooled extracts under reduced pressure to yield a black, oily material (2.0 g).

Fractionation and Isolation. The CNY-716 crude extract was fractionated using vacuum liquid chromatography (15 g of C₁₈ silica gel) and a step gradient of MeOH:H₂O (50 mL each; 25:75, 50:50, 60:40, 70:30, 80:20, 90:10, and 2 × 100:0) into eight fractions (LC1–LC8). Fractions containing indanopyrroles (LC6 and LC7) were further purified using HPLC (Phenomenex Kinetex C₁₈, 5 μm, 150 × 4.6 mm column, isocratic ACN:H₂O 68:32 with 0.05% formic acid mobile phase, 1.3 mL/min flow rate) to yield 5 mg of indanopyrrole A (**1**, retention time *t_R* = 12 min) and 0.5 mg of indanopyrrole B (**2**, *t_R* = 6 min).

Compound Characterization. *Indanopyrrole A (1).* Indanopyrrole A (**1**) is an amorphous white solid; $[\alpha]_D^{22}$ –214 (c 0.26, MeOH); UV (MeOH) $\lambda_{\max}$ (log ϵ) 200 (3.59), 261 (sh) (2.97) 295 ((3.34); IR (neat) $\nu_{\max}$ 3186 cm^{–1}, 2938 cm^{–1}, 2866 cm^{–1}, 1707 cm^{–1}, 1639 cm^{–1}, 1445 cm^{–1}, 1396 cm^{–1}, 1372 cm^{–1}, 1226 cm^{–1}, 1018 cm^{–1}; ¹H, ¹³C and 2D NMR (see Table S1); HR-ESI-TOF-MS *m/z* 458.1060 [*M* + *H*]⁺ (calcd for C₂₂H₂₇Cl₃NO₃⁺, 458.1052, 1.75 ppm).

Indanopyrrole B (2). Indanopyrrole B (**2**) is an amorphous white solid; $[\alpha]_D^{22}$ –111 (c 0.26, MeOH); UV (MeOH) $\lambda_{\max}$ (log ϵ) 200 (3.51), 245 (3.01) 296 (3.51); IR (neat) $\nu_{\max}$ 3326 cm^{–1}, 2952 cm^{–1}, 2866 cm^{–1}, 1644 cm^{–1}, 1410 cm^{–1}, 1027 cm^{–1}; ¹H, ¹³C and 2D NMR (see Table S2); HR-ESI-TOF-MS *m/z* [*M* + *H*]⁺ 424.1451 (calcd for C₂₂H₂₈Cl₂NO₃⁺, 424.1441, 2.36 ppm).

Genomic DNA Extraction. DNA was extracted from frozen aliquots (5 mL) of the 42 MAR4 strains using the Promega Wizard Genomic DNA Purification Kit with modifications suggested for Gram-positive bacteria. DNA purity, concentrations, and size were assessed using NanoDrop, Qubit, and gel electrophoresis. Short-read, paired-end Illumina sequencing (PE150) was performed at SeqCenter (Pittsburgh, PA). Initial genome assembly was performed by quality filtering raw reads with the BBMap Toolkit,⁶⁶ followed by a preliminary assembly using SPAdes.⁶⁷ All assemblies were compared for whole-genome average nucleotide identity (ANI) using fastANI⁶⁸ to identify strains sharing 95% ANI. Ten representative strains from each 95% ANI clade were selected for long-read Nanopore sequencing (Oxford) and compiled with public SRA data from previously sequenced strains (*N* = 12 Illumina, *N* = 3 PACBIO). Data were combined for each strain to perform a hybrid assembly with a unicycler,⁶⁹ using a kmer count of 31, 41, 51, 61, 71, 81, 91, 95, 101, 105, 111, and “mode” determined by ANI similarity to a reference strain (i.e., “bold” with ANI = 100%, “normal” with ANI >99%, “conservative” with ANI >97%). All assemblies were checked for quality, completeness, and contamination with checkM.⁷⁰

Biosynthetic gene clusters (BGCs) were predicted using antiSMASH v5⁴¹ and clustered into gene cluster families using BiG-SCAPE⁷¹ with GCFs defined at 0.4 dissimilarity based on a combined metric of gene synteny, protein domain structure, and homology. MAR4 BGCs were annotated using the MIBiG 2.7 database.⁷² Network files from BiG-SCAPE were visualized in Cytoscape 3.10.⁷³ Synteny plots were generated from GBK files created by AntiSMASH using the clinker package with the default parameters.

BGC amino acid sequences were sourced from NCBI GenBank and the AntiSMASH output. For each gene alignment, MuscleS⁷⁴ was used to create a stratified ensemble of 16 alignments using the default parameters. The alignment with the highest column confidence value was extracted from the ensemble using the maxcc option and used for a comparison of active site residues.

Metabolomics and Mass Spectrometry. Dried crude extracts were resuspended in MeOH (1 mg/mL) and centrifuge filtered using 0.2 μm filters (American Chromatography Supply, Vineland, NJ).

Samples (5 μL) were injected into an Agilent 1290 HPLC instrument coupled to an Agilent 6530 quadrupole time-of-flight (QToF) spectrometer. Parameters were set to a flow rate of 0.75 mL/min through a Kinetex C₁₈ reversed-phase column (5 μm, 150 mm × 4.6 mm) under the following conditions: 0–4 min, 5% acetonitrile (0.1% TFA) in water (0.1% TFA) with this first 4 min diverted to waste; 4–34 min, 10%–100% acetonitrile (0.1% TFA) in water (0.1% TFA); 34–36 min, 100% acetonitrile; 36–36.5 min, 100%–5% acetonitrile; 36.5–40 min, 5% acetonitrile. MS1 data were collected in positive and negative over two runs with a mass range of 80–1700 *m/z*, acquiring three spectra per second; MS2 fragmentation data were collected using two scans per second with a collision energy of 30 eV. The source gas temperature was 300 °C at a flow rate of 11 L/min at 35 psig.

Molecular Modeling and Docking Studies. A 3D model of indanopyrrole A (**1**) was built in Spartan'24 V1.1.0 and minimal energy conformers searched using molecular mechanics (CorrMMFF, ΔE < 25 kJ/mol) and further optimized using denisty functional theory (DFT) (Est. Density Functional ω B97X-D/6-31G*). The online version of ColabFold (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>) was used to generate a 3D structure from the amino acid sequence of the *IdpK* KS domain. Default parameters were used except the “template_mode” was changed to include the pdb70 reference library. Both 3D models were used as input for docking analysis using AMDock.⁷⁵ The default search parameters were used in an Autodock Vina model.^{76,77}

Antibacterial Testing. **1** was 10-fold serially diluted (from 128 μg/mL to 0.25 μg/mL) in a 96-well plate using sterile medium (Muller–Hinton broth (MHB); 50 μL of each solution per well). An *E. coli* LptD 4213 inoculum (50 μL) was added to reach a final concentration of 2e5 CFU/mL (determined via OD₆₀₀). The final test concentrations ranged from 64 μg/mL to 0.125 μg/mL. After incubation for 18 h at 37 °C, the well with the lowest concentration of compound that did not exhibit microbial growth was determined to be the MIC.

Additional MIC values were determined using broth microdilution in accordance with the Clinical Laboratory Standards Institute (CLSI) guidelines, using cation-adjusted MHB with minor modifications. Briefly, bacteria were grown to mid log phase (OD_{600 nm} = 0.4) at 37 °C while shaking, except for GAS, which was grown under static conditions. Bacterial cells were then centrifuged, washed, and diluted in PBS to obtain 2 × 10⁶ cfu/mL with 10 μL added to individual wells of a 96-well plate containing 170 μL MHB. Serial dilutions of indanopyrroles A and B starting at 32 or 16 μg/mL, respectively, were made in a separate plate, and 20 μL of the compound was then added to the test plate. The plates were sealed with parafilm and incubated at 37 °C for 24 h. Turbidity was measured at OD_{600 nm}, using an EnSpire Alpha plate reader. MIC was defined as the lowest concentration of the test compounds that inhibited bacterial growth. The positive control was colistin, the negative control was media only, and the growth control was the test organism alone with no drug.

Cell Viability Assay. A549 cells were seeded in 24 well plates (Corning, USA) at 2 × 10⁵ cells/well. Cells were left untreated or treated with 16, 32, and 128 μg/mL indanopyrrole A, or with corresponding amounts of the H₂O:DMSO (1:2) solvent vehicle as a negative control. As a positive control, A549 cells were lysed with Triton X-100. Cell culture supernatants were collected at two time points: 2 and 24 h. Cellular cytotoxicity was assessed by measuring the levels of lactate dehydrogenase (LDH; Promega, USA) released by the host cell into the supernatant. The percentage of cell death was calculated after subtracting the levels found in untreated control cells and dividing by the levels in a positive control of cells treated with a lysis solution (Triton X-100).

■ ASSOCIATED CONTENT

Data Availability Statement

Raw NMR data files for indanopyrroles A and B have been deposited in the Natural Products Magnetic Resonance

Database (NP-MRD, www.np-mrd.org) and can be accessed under Identifier Nos. NP0341895 and NP0341896, respectively. The LC-HRMS files have been uploaded to the MassIVE repository under MassIVE ID No. MSV000096392. Genome sequences are available for strain CNY-716 (NCBI BioSample ID No. SAMN44453297) and CNX-425 (NCBI BioSample ID No. SAMN44453231).

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jnatprod.4c00934>.

Supplemental figures including MS and NMR spectra for all compounds, genetic analyses, and 3D modeling (PDF)

AUTHOR INFORMATION

Corresponding Author

Paul R. Jensen – Scripps Institution of Oceanography, University of California–San Diego, La Jolla, California 92093, United States; orcid.org/0000-0003-2349-1888; Email: pjensen@ucsd.edu

Authors

Douglas Sweeney – Scripps Institution of Oceanography, University of California–San Diego, La Jolla, California 92093, United States

Alexander Bogdanov – Scripps Institution of Oceanography, University of California–San Diego, La Jolla, California 92093, United States

Alexander B. Chase – Department of Earth Sciences, Southern Methodist University, Dallas, Texas 75275, United States

Gabriel Castro-Falcón – Scripps Institution of Oceanography, University of California–San Diego, La Jolla, California 92093, United States; orcid.org/0000-0002-3504-0980

Alma Trinidad-Javier – Scripps Institution of Oceanography, University of California–San Diego, La Jolla, California 92093, United States

Samira Dahesh – Department of Pediatrics, University of California–San Diego, La Jolla, California 92093, United States; orcid.org/0009-0006-4451-2034

Victor Nizet – Department of Pediatrics and Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California–San Diego, La Jolla, California 92093, United States; orcid.org/0000-0003-3847-0422

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jnatprod.4c00934>

Author Contributions

[†]Authors D.S. and A.B. contributed equally to this work. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge assistance from Tadeusz Molinski (UCSD Chemistry and Biochemistry Department, Skaggs School of Pharmacy) for providing access to optical rotation, UV and IR instruments and expert guidance, as well as Brendan Duggan (Skaggs School of Pharmacy) and Anthony Mrse (Department of Chemistry, UCSD) for expert guidance and help in the NMR recording. This work was supported by

the National Institutes of Health (Grant No. R01GM085770 to P.R.J.).

REFERENCES

- (1) Benedetto Tiz, D.; Bagnoli, L.; Rosati, O.; Marini, F.; Sancineto, L.; Santi, C. *Molecules* **2022**, 27 (5), 1643.
- (2) Elsocht, M.; Giron, P.; Maes, L.; Versées, W.; Gutierrez, G. J.; De Grève, J.; Ballet, S. *Int. J. Mol. Sci.* **2021**, 22 (2), 635.
- (3) Harris, C. M.; Kannan, R.; Kopecka, H.; Harris, T. M. *J. Am. Chem. Soc.* **1985**, 107 (23), 6652–6658.
- (4) Groll, M.; Huber, R.; Potts, B. C. M. *J. Am. Chem. Soc.* **2006**, 128 (15), 5136–5141.
- (5) Gribble, G. *Mar. Drugs* **2015**, 13 (7), 4044–4136.
- (6) Sweeney, D.; Chase, A. B.; Bogdanov, A.; Jensen, P. R. *J. Nat. Prod.* **2024**, 87 (2), 439–452.
- (7) Asolkar, R. N.; Singh, A.; Jensen, P. R.; Aalbersberg, W.; Carté, B. K.; Feussner, K.-D.; Subramani, R.; DiPasquale, A.; Rheingold, A. L.; Fenical, W. *Tetrahedron* **2017**, 73 (16), 2234–2241.
- (8) Kwon, H. C.; Espindola, A. P. D. M.; Park, J.-S.; Prieto-Davó, A.; Rose, M.; Jensen, P. R.; Fenical, W. *J. Nat. Prod.* **2010**, 73 (12), 2047–2052.
- (9) Raju, R.; Piggott, A. M.; Barrientos Diaz, L. X.; Khalil, Z.; Capon, R. J. *Org. Lett.* **2010**, 12 (22), 5158–5161.
- (10) Carretero-Molina, D.; Ortiz-López, F. J.; Martín, J.; Oves-Costales, D.; Díaz, C.; de la Cruz, M.; Cautain, B.; Vicente, F.; Genilloud, O.; Reyes, F. *Mar. Drugs* **2020**, 18 (1), 22.
- (11) Pereira, F.; Almeida, J. R.; Paulino, M.; Grilo, I. R.; Macedo, H.; Cunha, I.; Sobral, R. G.; Vasconcelos, V.; Gaudêncio, S. P. *Mar. Drugs* **2020**, 18 (1), 63.
- (12) Cheng, Y.-B.; Jensen, P. R.; Fenical, W. *Eur. J. Org. Chem.* **2013**, 2013 (18), 3751–3757.
- (13) Pathirana, C.; Jensen, P. R.; Fenical, W. *Tetrahedron Lett.* **1992**, 33 (50), 7663–7666.
- (14) Hardt, I. H.; Jensen, P. R.; Fenical, W. *Tetrahedron Lett.* **2000**, 41 (13), 2073–2076.
- (15) Bauermeister, A.; Pereira, F.; Grilo, I. R.; Godinho, C. C.; Paulino, M.; Almeida, V.; Gobbo-Neto, L.; Prieto-Davó, A.; Sobral, R. G.; Lopes, N. P.; Gaudêncio, S. P. *Environ. Microbiol.* **2019**, 21 (3), 1099–1112.
- (16) Gallagher, K. A.; Jensen, P. R. *BMC Genomics* **2015**, 16 (1), 960.
- (17) Duncan, K. R.; Crüsemann, M.; Lechner, A.; Sarkar, A.; Li, J.; Ziemert, N.; Wang, M.; Bandeira, N.; Moore, B. S.; Dorrestein, P. C.; Jensen, P. R. *Chem. Biol.* **2015**, 22 (4), 460–471.
- (18) Goering, A. W.; McClure, R. A.; Doroghazi, J. R.; Albright, J. C.; Haverland, N. A.; Zhang, Y.; Ju, K. S.; Thomson, R. J.; Metcalf, W. W.; Kelleher, N. L. *ACS Cent. Sci.* **2016**, 2 (2), 99–108.
- (19) Mantovani, S. M.; Moore, B. S. *J. Am. Chem. Soc.* **2013**, 135 (48), 18032–18035.
- (20) Fu, C.; Xie, F.; Hoffmann, J.; Wang, Q.; Bauer, A.; Brönstrup, M.; Mahmud, T.; Müller, R. *ChemBioChem* **2019**, 20 (6), 764–769.
- (21) Yamanaka, K.; Ryan, K. S.; Gulder, T. A. M.; Hughes, C. C.; Moore, B. S. *J. Am. Chem. Soc.* **2012**, 134 (30), 12434–12437.
- (22) Hjörleifsson Eldjárn, G.; Ramsay, A.; van der Hooft, J. J. J.; Duncan, K. R.; Soldatou, S.; Rousu, J.; Daly, R.; Wandy, J.; Rogers, S. *PLoS Comput. Biol.* **2021**, 17 (5), No. e1008920.
- (23) Leão, T. F.; Wang, M.; da Silva, R.; Gurevich, A.; Bauermeister, A.; Gomes, P. W. P.; Brejnrod, A.; Glukhov, E.; Aron, A. T.; Louwen, J. J. R.; Kim, H. W.; Reher, R.; Fiore, M. F.; van der Hooft, J. J. J.; Gerwick, L.; Gerwick, W. H.; Bandeira, N.; Dorrestein, P. C. *PNAS Nexus* **2022**, 1 (5), 1–15.
- (24) Shin, Y.-H.; Im, J. H.; Kang, I.; Kim, E.; Jang, S. C.; Cho, E.; Shin, D.; Hwang, S.; Du, Y. E.; Huynh, T.-H.; Ko, K.; Ko, Y.-J.; Nam, S.-J.; Awakawa, T.; Lee, J.; Hong, S.; Abe, I.; Moore, B. S.; Fenical, W.; Yoon, Y. J.; Cho, J.-C.; Lee, S. K.; Oh, K.-B.; Oh, D.-C. *J. Am. Chem. Soc.* **2023**, 145 (3), 1886–1896.
- (25) Westley, J. W.; Evans, R. H., Jr.; Sello, L. H.; Troupe, N. J. *Antibiot. (Tokyo)* **1979**, 32 (2), 100–107.

- (26) Liu, C.-M.; Hermann, T. E.; Liu, M.; Bull, D. N.; Palleroni, N. J.; Prosser, B. L. T.; Westley, O. W.; Miller, P. A. *J. Antibiot. (Tokyo)* **1979**, *32* (2), 95–99.
- (27) Li, C.; Roege, K. E.; Kelly, W. L. *ChemBioChem* **2009**, *10* (6), 1064–1072.
- (28) Roege, K. E.; Kelly, W. L. *Org. Lett.* **2009**, *11* (2), 297–300.
- (29) Zhang, D.; Nair, M. G.; Murry, M.; Zhang, Z. *J. Antibiot. (Tokyo)* **1997**, *50* (7), 617–620.
- (30) Aherfi, S.; Pradines, B.; Devaux, C.; Honore, S.; Colson, P.; La Scola, B.; Raoult, D. *Future Microbiol.* **2021**, *16* (17), 1341–1370.
- (31) Murenets, N. V.; Kudina, M. K.; Korobkova, T. P.; Drobysheva, T. N.; Kliuev, N. A. Kafamycin—A New Pyrrol Ether Antibiotic (in Russ.) *Antibiot. Med. Biotechnol.* **1987**, *32* [11], 811–814.
- (32) Larsen, S. H.; Boeck, L. V. D.; Mertz, F. P.; Paschal, J. W.; Ocolowicz, J. L. *J. Antibiot. (Tokyo)* **1988**, *41* (9), 1170–1177.
- (33) Gerasimova, T. I.; Stvolinskaia, N. S. The Effect of the Polyether Antibiotic Omomycin on the Activity of Cyclic Nucleotides in the Primary Culture of Newborn Rat Heart Cells (in Russ.) *Vopr. Med. Khim.* **1990**, *36* [5], 39–41.
- (34) Toth, P.; Szell, V.; Koczka, I.; Horvath, G.; Szabo, I. M.; Szayly, M.; Ambrus, G.; Bedo, J.; Berdy, J. Producing Indanomycin and Homindanomycin Using *Streptomyces Galbus*. Hung. Patent No. HU49909 A2, 1989.
- (35) Miao, S.; Anstee, M. R.; Baichwal, V.; Park, A. *Tetrahedron Lett.* **1995**, *36* (32), 5699–5702.
- (36) Izumikawa, M.; Komaki, H.; Hashimoto, J.; Takagi, M.; Shinya, K. *J. Antibiot.* **2008**, *61* (5), 326–329.
- (37) Beloeil, J. C.; Delsuc, M. A.; Lallemand, J. Y.; Dauphin, G.; Jeminet, G. *J. Org. Chem.* **1984**, *49* (10), 1797–1800.
- (38) Buchfink, B.; Reuter, K.; Drost, H.-G. *Nat. Methods* **2021**, *18* (4), 366–368.
- (39) Gilchrist, C. L. M.; Booth, T. J.; van Wersch, B.; van Grieken, L.; Medema, M. H.; Chooi, Y. *Bioinform. Adv.* **2021**, *1* (1), 1–10.
- (40) Terlouw, B. R.; Blin, K.; Navarro Muñoz, J. C.; Avalon, N. E.; Chevetter, M. G.; Egbert, S.; Lee, S.; Meijer, D.; Recchia, M. J. J.; Reitz, Z. L.; et al. *Nucleic Acids Res.* **2023**, *51* (D1), D603–D610.
- (41) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2019**, *47* (W1), W81–W87.
- (42) Purdy, T. N.; Kim, M. C.; Cullum, R.; Fenical, W.; Moore, B. S. *J. Am. Chem. Soc.* **2021**, *143* (10), 3682–3686.
- (43) Ziemert, N.; Podell, S.; Penn, K.; Badger, J. H.; Allen, E.; Jensen, P. R. *PLoS One* **2012**, *7* (3), No. e34064.
- (44) Klau, L. J.; Podell, S.; Creamer, K. E.; Demko, A. M.; Singh, H. W.; Allen, E. E.; Moore, B. S.; Ziemert, N.; Letzel, A. C.; Jensen, P. R. *J. Biol. Chem.* **2022**, *298* (10), No. 102480.
- (45) Kwan, D. H.; Schulz, F. *Molecules* **2011**, *16* (7), 6092–6115.
- (46) Keatinge-Clay, A. T. *Chem. Rev.* **2017**, *117* (8), 5334–5366.
- (47) Cantu, D. C.; Chen, Y.; Reilly, P. J. *Protein Sci.* **2010**, *19* (7), 1281–1295.
- (48) Cantu, D. C.; Chen, Y.; Lemons, M. L.; Reilly, P. J. *Nucleic Acids Res.* **2011**, *39*, D342–D346.
- (49) Ichikawa, N.; Sasagawa, M.; Yamamoto, M.; Komaki, H.; Yoshida, Y.; Yamazaki, S.; Fujita, N. *Nucleic Acids Res.* **2012**, *41* (D1), D408–D414.
- (50) Rommel, K. R.; Li, C.; Kelly, W. L. *Org. Lett.* **2011**, *13* (10), 2536–2539.
- (51) Drulyte, I.; Obajdin, J.; Trinh, C. H.; Kalverda, A. P.; van der Kamp, M. W.; Hemsworth, G. R.; Berry, A. *IUCrJ.* **2019**, *6* (6), 1120–1133.
- (52) Bauman, K. D.; Shende, V. V.; Chen, P. Y.-T.; Trivella, D. B. B.; Gulder, T. A. M.; Vellalath, S.; Romo, D.; Moore, B. S. *Nat. Chem. Biol.* **2022**, *18* (5), 538–546.
- (53) Roush, W. R.; Myers, A. G. *J. Org. Chem.* **1981**, *46* (7), 1509–1511.
- (54) Roush, W. R.; Peseckis, S. M. *Tetrahedron Lett.* **1982**, *23* (47), 4879–4882.
- (55) Roush, W. R.; Peseckis, S. M.; Walts, A. E. *J. Org. Chem.* **1984**, *49* (18), 3429–3432.
- (56) Stocking, E. M.; Williams, R. M. *Angew. Chem.* **2003**, *115* (27), 3186–3223.
- (57) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. *Nature* **2021**, *596* (7873), 583–589.
- (58) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. *Nat. Methods* **2022**, *19* (6), 679–682.
- (59) Valdés-Tresanco, M. S.; Valdés-Tresanco, M. E.; Valiente, P. A.; Moreno, E. *Biol. Direct* **2020**, *15* (1), 12.
- (60) Sayers, E. W.; Bolton, E. E.; Brister, J. R.; Canese, K.; Chan, J.; Comeau, D. C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; Madej, T.; Marchler-Bauer, A.; Lanczycki, C.; Lathrop, S.; Lu, Z.; Thibaud-Nissen, F.; Murphy, T.; Phan, L.; Skripchenko, Y.; Tse, T.; Wang, J.; Williams, R.; Trawick, B. W.; Pruitt, K. D.; Sherry, S. T. *Nucleic Acids Res.* **2022**, *50* (D1), D20–D26.
- (61) Wang, M.; Jarmusch, A. K.; Vargas, F.; Aksenov, A. A.; Gauglitz, J. M.; Weldon, K.; Petras, D.; da Silva, R.; Quinn, R.; Melnik, A. V.; van der Hooft, J. J. J.; Caraballo-Rodríguez, A. M.; Nothias, L. F.; Aceves, C. M.; Panitchpakdi, M.; Brown, E.; Di Ottavio, F.; Sikora, N.; Elijah, E. O.; Labarta-Bajo, L.; Gentry, E. C.; Shalpour, S.; Kyle, K. E.; Puckett, S. P.; Watrous, J. D.; Carpenter, C. S.; Bouslimani, A.; Ernst, M.; Swofford, A. D.; Zúñiga, E. I.; Balunas, M. J.; Klassen, J. L.; Loomba, R.; Knight, R.; Bandeira, N.; Dorrestein, P. C. *Nat. Biotechnol.* **2020**, *38* (1), 23–26.
- (62) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapon, C. A.; Luzzatto-Knaan, T.; et al. *Nat. Biotechnol.* **2016**, *34* (8), 828–837.
- (63) Nonejuie, P.; Burkart, M.; Pogliano, K.; Pogliano, J. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (40), 16169–16174.
- (64) Vila, J.; Martí, S.; Sánchez-Céspedes, J. *J. Antimicrob. Chemother.* **2007**, *59* (6), 1210–1215.
- (65) Soto, S. M. *Virulence* **2013**, *4* (3), 223–229.
- (66) Bushnell, B. BBMap: A Fast, Accurate Splice-Aware Aligner. Presented at the 9th Annual Genomics of Energy & Environment Meeting, Walnut Creek, CA, March 17–20, 2014, pp 3–5. sourceforge.net/projects/bbmap/
- (67) Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A. A.; Dvorkin, M.; Kulikov, A. S.; Lesin, V. M.; Nikolenko, S. I.; Pham, S.; Pribelski, A. D.; Pyshkin, A. V.; Sirotkin, A. V.; Vyahhi, N.; Tesler, G.; Alekseyev, M. A.; Pevzner, P. A. *J. Comput. Biol.* **2012**, *19* (5), 455–477.
- (68) Jain, C.; Rodriguez-R, L. M.; Phillippy, A. M.; Konstantinidis, K. T.; Aluru, S. *Nat. Commun.* **2018**, *9* (1), 5114.
- (69) Wick, R. R.; Judd, L. M.; Gorrie, C. L.; Holt, K. E. *PLoS Comput. Biol.* **2017**, *13* (6), e1005595.
- (70) Parks, D. H.; Imelfort, M.; Skennerton, C. T.; Hugenholtz, P.; Tyson, G. W. *Genome Res.* **2015**, *25* (7), 1043–1055.
- (71) Navarro-Muñoz, J. C.; Selem-Mojica, N.; Mullowney, M. W.; Kautsar, S. A.; Tryon, J. H.; Parkinson, E. I.; De Los Santos, E. L. C.; Yeong, M.; Cruz-Morales, P.; Abubucker, S.; Roeters, A.; Lokhorst, W.; Fernandez-Guerra, A.; Cappelini, L. T. D.; Goering, A. W.; Thomson, R. J.; Metcalf, W. W.; Kelleher, N. L.; Barona-Gomez, F.; Medema, M. H. *Nat. Chem. Biol.* **2020**, *16* (1), 60–68.
- (72) Kautsar, S. A.; Blin, K.; Shaw, S.; Navarro-Muñoz, J. C.; Terlouw, B. R.; Van Der Hooft, J. J. J.; Van Santen, J. A.; Tracanna, V.; Suarez Duran, H. G.; Pascal Andreu, V.; Selem-Mojica, N.; Alanjary, M.; Robinson, S. L.; Lund, G.; Epstein, S. C.; Sisto, A. C.; Charkoudian, L. K.; Collemare, J.; Linington, R. G.; Weber, T.; Medema, M. H. *Nucleic Acids Res.* **2019**, *48* (D1), D454–D458.
- (73) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res.* **2003**, *13* (11), 2498–2504.

- (74) Edgar, R. C. *Nat. Commun.* **2022**, *13* (1), 6968.
- (75) Valdés-Tresanco, M. S.; Valdés-Tresanco, M. E.; Valiente, P. A.; Moreno, E. *Biol. Direct* **2020**, *15* (1), 12.
- (76) Trott, O.; Olson, A. J. *J. Comput. Chem.* **2010**, *31* (2), 455–461.
- (77) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. *J. Chem. Inf. Model.* **2021**, *61* (8), 3891–3898.