Pattern-based genome mining guides discovery of the antibiotic indanopyrrole A from a marine streptomycete

Douglas Sweeney[*†], Alexander Bogdanov*[*†], Alexander B. Chase[‡], Gabriel Castro-Falcón[†], Alma Trinidad-Javier[†], Samira Dahesh[§], Victor Nizet[§,#], Paul R. Jensen[¶†]*

[†] Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, USA

[‡] Department of Earth Sciences, Southern Methodist University, Dallas, TX 75275, USA

[§] Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA

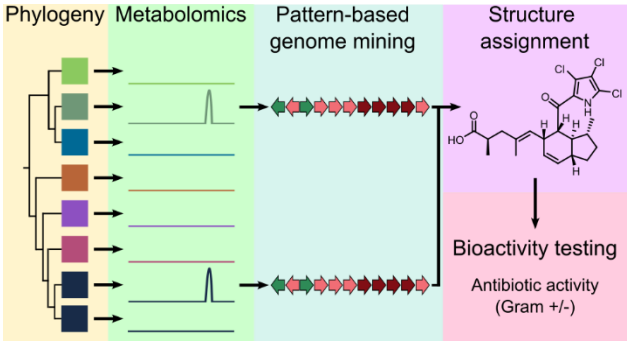
[#] Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA

*These authors contributed equally

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 pyrroloketoindane 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 pyrroloketoindane biosynthesis. Broader searches of public databases reveal the rarity of this BGC while also highlighting opportunities for discovering additional compounds in this uncommon class.

GRAPHICAL ABSTRACT



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 vancomycin[3] 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 discovering new halogenated metabolites. In 2015, it was estimated that more than 5,000 halogenated natural products had been discovered, with the majority 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, 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 NPLinker[22] 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]

Pyrroloketoindanes are characterized by indane and pyrrole systems bridged via a ketone. Compounds containing these ring systems are rare and often biologically active, making them logical targets for natural product discovery. Prior to this study, only six pyrroloketoindanes 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 pyrroloketoindane described was indanomycin, which contains a tetrahydropyran and was discovered in 1979 from *Streptomyces antibioticus* NRRL 8167[25]. Indanomycin has activity against Gram-positive bacteria[26] and acts as an ionophore. [27,28] Later work has shown that indanomycin has insecticidal[29] 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 pyrroloketoindane containing compound homoindanomycin was described in a patent from 1989 without any information about activity.[34] Stawamycin inhibits Epstein-Barr viral transcription factor BZLF1 binding to DNA and

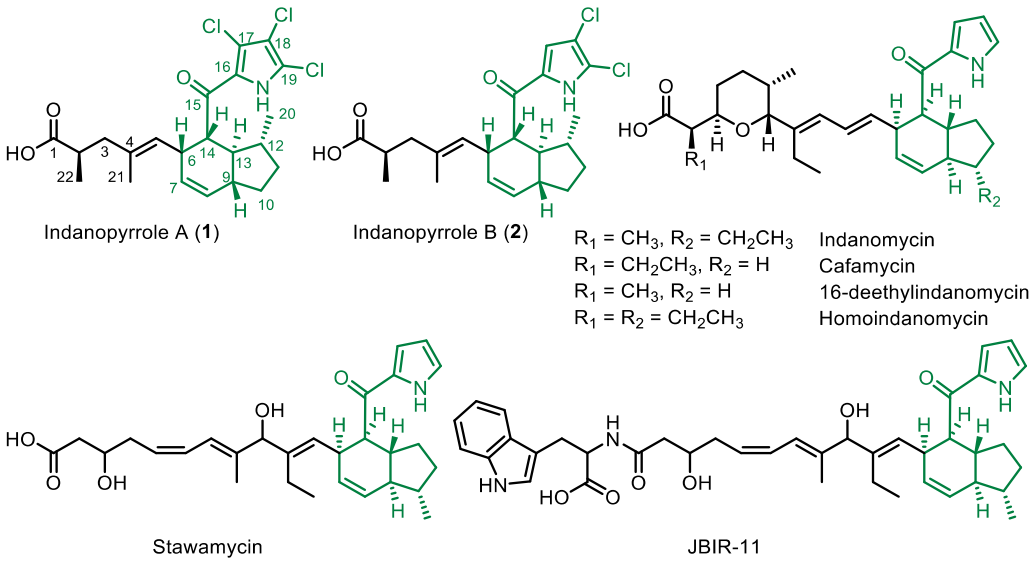
was the first indanomycin analogue described to lack a tetrahydropyran.[35] Finally, JBIR-11 was reported in 2008 from *Streptomyces viridochromogenes* subsp. *sulfomycini* NBRC 13830. This unusual analogue of stawamycin contains a tryptophan moiety and was shown to be cytotoxic to HT-1080 fibrosarcoma cells.[36]

Here we describe the discovery of novel di- and trichlorinated pyrroloketoindane antibiotics using paired metabolomic and genomic datasets. The candidate BGC provides new insights into pyrroloketoindane 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 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).

Chart 1. Indanopyrroles A and B (**1**-**2**) and previously reported pyrroloketoindane natural products. Pyrroloketoindane moieties are colored green. Absolute configurations for **1-2** are based on bioinformatic prediction.



The molecular formula (MF) of **1** $C_{22}H_{26}Cl_3NO_3$ with nine degrees of unsaturation was determined from the accurate mass (HR-ESI-TOF-MS) measurement. NMR experiments (^{1}H, ^{13}C, COSY, HSQC and HMBC) in CD_3OD led to the unambiguous elucidation of the planar structure (Table S1). The proton spectrum showed the presence of two methyl doublets (CH_3-20 and CH_3-22; δH 0.99, *J* = 6.1 Hz and 0.98, *J* = 6.8 Hz) and one methyl singlet (CH_3-21; δH 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 ^{1}H-^{1}H 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 two substituents at C-6 and C-14. The sidechain at C-6 was established using ^{1}H-^{13}C HMBC, which showed long range correlations between the methyl protons H_3-21 to C-3, C-4, and C-5 and from H_3-22 to the carboxylic acid

carbonyl C-1 (δC 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 C_5HCl_3NO 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 the carbonyl C-15 and a sp^2 hybridized carbon C-16 (δC 128.0 ppm) established the connection of the trichloropyrrole moiety to the indane core via a ketone bridge. Analysis of the ^{13}C 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 shielded ^{13}C resonance of the methyl substituent CH_3-21 (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 NOESY experiment (NOE correlations observed between H-6, H-9 and H-14, as well as from H-13 to H_3-20) and is in accordance with previously reported pyrroloketoindanes (Chart 1). The absolute configurations at C-2 and C-12 in the linear biosynthetic precursor were predicted bioinformatically as 2*R* and 12*S* (12*R* 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** (2*R*, 6*R*, 9*R*, 12*R*, 13*S*, 14*R* Figure S3B-C) and differences in the indane core compared to other pyrroloketoindanes (Chart 1). Raw NMR data files can be accessed from the Natural Product Magnetic Resonance Database Project (NP-MRD https://np-mrd.org/) under identifiers NP0341895 (**1**) and NP0341896 (**2**).

HR-ESI-TOFMS/MS analyses of **2** revealed the molecular formula $C_{22}H_{27}Cl_2NO_3$ and a prominent dichloropyrrole containing MS/MS fragment (*m/z* 161.9527 calcd for $C_5H_2Cl_2NO^+$, 161.9513, -8.65 ppm, Figures S4-5). These data indicated that compound **2** is a dichlorinated analogue of **1**. To assign the positions of the two chlorines on the pyrrole, we recorded ^{1}H, HSQC, and HMBC NMR spectra using a 600 MHz NMR instrument equipped with a 5 mm cryoprobe (Table S2). As expected, the ^{1}H spectrum was almost identical with that of **1**, except for an additional sharp singlet at 7.04 ppm, the shielded of H-6 and H-14 resonances (from 3.65 to 3.50 and 3.88 to 3.49 ppm, respectively), and the slightly shielded H_3-20 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 di- or trichloropyrrole starter unit. As such, a DIAMOND BLAST[38] database containing 42 MAR4 genomes was created using cblaster[39] and queried for the proline adenyltransferase (*idmJ*) and prolyl-carrier protein dehydrogenase (*idmI*) gene sequences from the indanomycin (*idm*) BGC (MIBiG #: 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

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

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

174 **Biosynthesis of the indanopyrroles.** A comparison of the *idp* and *idm* BGCs provides insight into
175 the structural differences between their small molecule products (Figure 1A-B). As expected,
176 both BGCs contain homologous genes encoding proline adenyltransferases (*idpE/idmJ*), proline
177 carrier proteins (*idpF/idmK*), and dehydrogenases (*idpD/idmI*) to account for the generation of

178 the pyrrole moieties from proline (Figure 1C). *Idp* then diverges by the presence of the
179 halogenase gene *idpG*, which was annotated as a tryptophan halogenase. Unlike other BGCs
180 linked to halogenated MAR4 natural products, *idp* does not contain a vanadium haloperoxidase
181 encoding gene. Instead, *IdpG* shares homology with halogenase genes in the BGCs of other
182 chlorinated pyrrole containing natural products[19,20,42] and likely accounts for the tri- and
183 dichlorinated pyrrole moieties observed in **1** and **2**, respectively. The chlorinated pyrrole
184 generated from *idpD-G* then serves as the starter unit for seven polyketide extensions encoded
185 by the three T1PKS genes (*idpH-J*). Notably, a NaPDoS2[43,44] analysis of the module 1 (loading
186 module) ketosynthase (KS) domain within *IdpH* places it in a clade with other pyrrole accepting
187 KSs, further supporting the functional prediction for the starter unit (Figure S8).
188

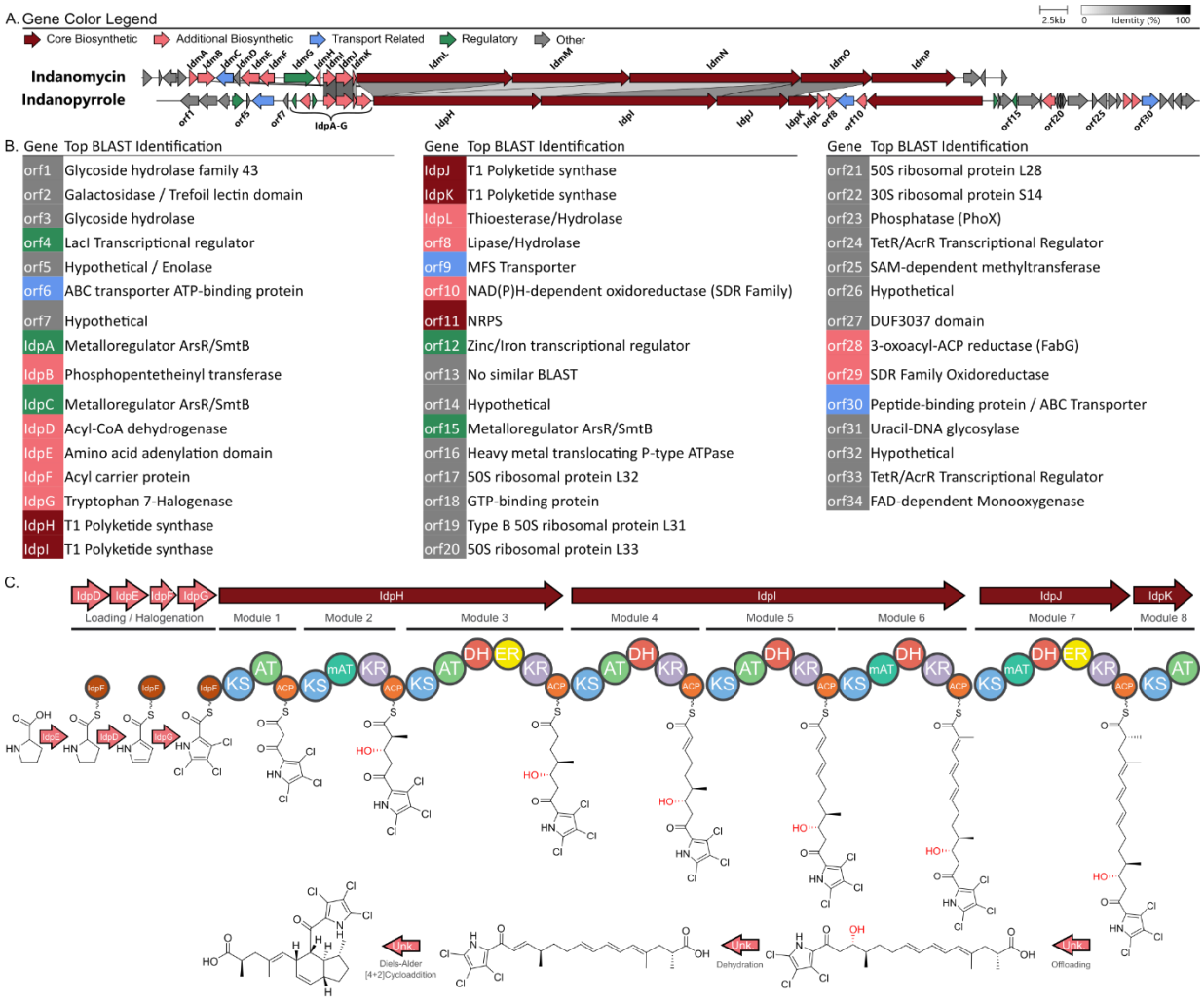


Figure 1. Candidate indanopyrrole BGC and proposed biosynthesis. A. Comparison of the candidate indanopyrrole BGC (*idp*) from *Streptomyces* sp. CNX-425 with the indanomycin BGC (*idm*, MIBiG #: BGC0000079). B. *Idp* gene annotations based on top BLAST matches to the NCBI NR/NT database. C. Proposed 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.

Following starter unit selection, the acyltransferase (AT) domains associated with modules two, six, and seven 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 2*S*-methyl-3*S*-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**-**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**-**2** or that the cyclase is not involved (Figure S13). Finally, while antiSMASH calls 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 pyrroloketoindane 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 production[50] and molecular modelling 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 involved. KS functional diversification is well documented and includes the non-elongating, terminal KS domain in salinosporamide A

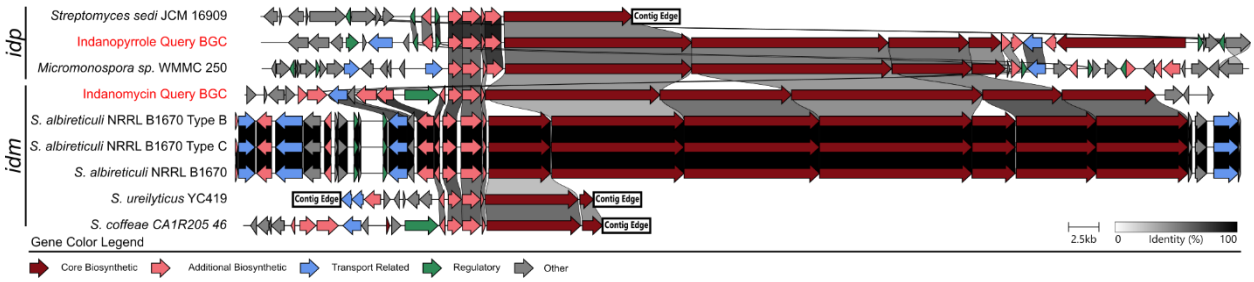
biosynthesis, which was recently shown to catalyze the formation of an alkene which then facilitates intramolecular carbon-carbon bond formation and the cyclized end product.[52] Notably, both *idm* and *idp* lack a dehydratase domain in the second PKS module, which would provide the alkene dienophile 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 modelled the 3D protein structure using the open-source implementation of AlphaFold 2[57] 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 suite[59] 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 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 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 configuration 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 the indanopyrroles and other pyrroloketoindanes.

Gene cluster and metabolite distribution. To more broadly explore *idp* and *idm* distributions, we used cblaster[60] to remotely query the NCBI nr database using the respective adenyltransferase (*idpE/idmJ)* and proline dehydrogenase (*idpD/idmJ)* genes, which returned 3,283 (*idp*) and 3,198 (*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 (JAJQQS010000003.1) (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

285 *Streptomyces sedi* JCM 16909 (VDGT01000017.1). These results indicate that, despite being
286 detected in genetically distant actinobacterial lineages (*Streptomyces* and *Micromonospora*),
287 both *idm* and *idp* are rare in publicly available sequence data.



288

289 **Figure 2.** Distribution of *idp* and *idm*. The seven closest *idp* and *idm* homologs identified from
290 GenBank searches highlight the rarity of these BGCs and presents opportunities for discovery. A
291 full list of BGC hits is provided in Supplementary Figure S15.

292 We next used the Mass Spectrometry Search Tool[61] (MASST) within the Global Natural Products
293 Social Molecular Networking[62] (GNPS) platform to search for the MS/MS spectrum of
294 indanopyrrole A within public datasets. After detecting no hits using the default parameters, we
295 lowered the minimum cosine score to 0.6 and the minimum matched fragments to 2 and
296 detected 71 hits when allowing for analogue searching. Upon manual inspection, none of the
297 *m/z* values or fragmentation spectra matched indanopyrrole A. The lack of any significant
298 matches among 2,709 public datasets highlights the rarity of these natural products.

299 **Bioactivity.** Indanopyrrole A demonstrated significant antibacterial activity, with a minimum
300 inhibitory concentration (MIC) of 4 µg/mL against membrane-deficient *Escherichia coli*
301 lptD4213[63] (Table 1). Broader testing revealed potent activity against several clinically relevant
302 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 epidermis* (MIC = 4 µg/mL) (Table 1). To our best knowledge, **1** is the first pyrroloketoindane natural product to show activity 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 hours 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).

Table 1. Antibacterial activities of indanopyrroles A and B reported as minimum inhibitory activity. N.T. = Not Tested

Strains tested	Gram +/-	Indanopyrrole A	Indanopyrrole B
MRSA TCH1516	+	2 µg/mL	>16 µg/mL
Group A *Streptococcus* M1T1	+	4 µg/mL	>16 µg/mL
Vancomycin-resistant *Enterococcus faecium* DAPS	+	2 µg/mL	>16 µg/mL
MRSE (*Staphylococcus epidermidis*)	+	4 µg/mL	N.T.
Haemophilus influenzae 1	-	1 µg/mL	N.T.
Haemophilus influenzae 2	-	2 µg/mL	N.T.
Escherichia coli LptD4213	-	4 µg/mL	N.T.
Escherichia coli K1 RS218	-	>32 µg/mL	>16 µg/mL
Acinetobacter baumannii 5075	-	>32 µg/mL	N.T.

Given the antibiotic activity of indanopyrrole A, it is unclear how producing strains maintain resistance. However, a drug resistance 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 baumanii* 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**-**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 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 a functional role for 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 tri-chlorinated 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 it may function via a different mechanism. The promising therapeutic index of indanopyrrole A (**1**) supports further structure-activity studies and 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 actinomyces 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 Jasco P-2000 polarimeter. UV spectra were measured on a Jasco V-630 spectrophotometer. IR spectra were acquired on a JASCO FTIR-4100 spectrometer (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 $^{1}H\{^{13}C\}$ room temperature probe (JEOL, Akishima, Tokyo, Japan) or on a Bruker Avance III (600

MHz) NMR spectrometer with a 5 mm $^{1}H\{^{13}C/^{15}N\}$ room temperature or cryo probe (Billerica, MA). ^{13}C NMR spectrum was recorded on a Varian 500 MHz spectrometer equipped with a 5 mm $^{1}H\{^{13}C\}$ XSens cold probe (Varian Inc., Palo Alto, CA, USA, now Agilent Technologies). NMR spectra were referenced to the solvent signals (CHD_2OD, δ_H 3.31 and CD_3OD, δ_C 49.00 ppm). LC-HR-ESIMS was performed on an Agilent 1260 Infinity HPLC system equipped with a degasser, binary pump, autosampler, DAD detector, coupled to an Agilent 6530 Accurate-Mass QToF with ESI-source coupled with an Agilent 1260 Infinity HPLC and calibrated using the Agilent Reference Calibration Mix. Compounds were isolated on an Agilent HPLC system with 1100 G1312A binary pump, 1100 G1315A DAD UV/Vis detector, 1100 G1313A autosampler, and 1100 G1322A degasser (Agilent Technologies, Santa Clara, CA).

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, 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 into 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 four days, HP-20 resin (1 g, wet weight) was added to each flask. After three days of incubation with resin, all cultures were extracted with 50 mL EtOAc. Organic extracts were separated, dried with anhydrous Na_2SO_4, 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 incubated for an additional 4 days. After 14 days, the resin and the cellular material were filtered through cheesecloth, washed with deionized H_2O, and extracted with MeOH (4 x 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 C_{18} silica gel) and a step gradient of MeOH:H_2O (50 mL each; 25:75, 50:50, 60:40, 70:30, 80:20, 90:10, and 2x 100:0) into eight fractions (LC1-8). Fractions containing indanopyrroles (LC6-7) were further purified using HPLC (Phenomenex Kinetex $C_{18,}$ 5 µm, 150x4.6 mm column, isocratic ACN:H_2O 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 indanopyrrole B (**2**, t_R = 6 min).

Compound characterization. *Indanopyrrole A (**1**)*: amorphous white solid; $[\alpha]^{22}_D$ –214 (*c* 0.26, MeOH); UV (MeOH) λ_{max} (log ε) 200 (3.59), 261 (sh) (2.97) 295 ((3.34); IR (neat) ν_{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, ^{13}C and 2D NMR, Table S1; HR-ESI-TOF-MS *m/z* 458.1060 $[M+H]^+$ (calcd for $C_{22}H_{27}Cl_3NO_3^+$, 458.1052, 1.75 ppm).

*Indanopyrrole B (**2**)*: amorphous white solid; $[\alpha]^{22}_{D}$ -111 (*c* 0.26, MeOH); UV (MeOH) λ_{max} (log ε) 200 (3.51), 245 (3.01) 296 (3.51); IR (neat) ν_{max} 3326 cm^{-1}, 2952 cm^{-1},2866 cm^{-1}, 1644 cm^{-1}, 1410 cm^{-1}, 1027 cm^{-1}; ^{1}H, ^{13}C and 2D NMR, Table S2; HR-ESI-TOF-MS *m/z* $[M+H]^{+}$ 424.1451 (calcd for $C_{22}H_{28}Cl_2NO_3^{+}$, 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 suggested modifications 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 fastANI[68] 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 unicycler[69] using a kmer count 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 v5[41] and clustered into gene cluster families using BiG-SCAPE[71] 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 default parameters.

BGC amino acid sequences were sourced from NCBI GenBank and the AntiSMASH output. For each gene alignment Muscle5[74] 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 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 coupled to an Agilent 6530 quadrupole time-of-fight (QToF) spectrometer. Parameters were set to a flow rate of 0.75 mL/min through a Kinetex C_{18} reversed-phase column (5 µm, 150x4.6 mm) under the following conditions: 0-4 min 5% acetonitrile (0.1% TFA) in water (0.1% TFA) with this first 4 minutes 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 was 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 30eV. The source gas temperature was 300 °C at a flow rate of 11 L per minute at 35 psig.

Molecular modelling 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 DFT (Est. Density Functional ωB97X-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 ten-fold serially diluted (from 128 µg/mL to 0.25 µg/mL) in a 96-well plate using sterile medium (Muller-Hinton broth; 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 OD_{600}). The final test concentrations ranged from 64 µg/mL to 0.125 µg/mL. After incubation for 18 hours 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 Mueller Hinton Broth (MHB) with minor modifications. Briefly, bacteria were grown to mid-log phase (OD_{600nm} = 0.4) at 37 °C while shaking except for GAS which was grown under static condition. Bacterial cells were then centrifuged, washed, and diluted in PBS to obtain 2×10^6 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 µg/mL or 16 µg/mL, respectively, were made in a separate plate, 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_{600nm} using an EnSpire Alpha plate reader. MIC was defined as the lowest concentration of the test compounds that inhibited bacterial growth.

Cell Viability Assay. A549 cells were seeded in 24 well plates (Corning, United States) at $2x10^5$ cells/well. Cells were left untreated or treated with 16, 32, and 128 µg/mL of 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 hrs. Cellular cytotoxicity was assessed by measuring the levels of lactate dehydrogenase (LDH; Promega, United States) released by the host cell into 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

Supporting Information:

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

AUTHOR INFORMATION

Corresponding Author

¶ Email: pjensen@ucsd.edu

Author Notes

*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.

ACKNOWLEDGEMENT

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 R01GM085770 to PRJ.

REFERENCES

(1) Benedetto Tiz, D.; Bagnoli, L.; Rosati, O.; Marini, F.; Sancineto, L.; Santi, C. New Halogen-Containing Drugs Approved by FDA in 2021: An Overview on Their Syntheses and Pharmaceutical Use. *Molecules* **2022**, *27* (5), 1643. https://doi.org/10.3390/molecules27051643.

(2) Elsocht, M.; Giron, P.; Maes, L.; Versées, W.; Gutierrez, G. J.; De Grève, J.; Ballet, S. Structure–Activity Relationship (SAR) Study of Spautin-1 to Entail the Discovery of Novel NEK4 Inhibitors. *Int J Mol Sci* **2021**, *22* (2), 635. https://doi.org/10.3390/ijms22020635.

(3) Harris, C. M.; Kannan, R.; Kopecka, H.; Harris, T. M. The Role of the Chlorine Substituents in the Antibiotic Vancomycin: Preparation and Characterization of Mono- and Didechlorovancomycin. *J Am Chem Soc* **1985**, *107* (23), 6652–6658. https://doi.org/10.1021/ja00309a038.

(4) Groll, M.; Huber, R.; Potts, B. C. M. Crystal Structures of Salinosporamide A (NPI-0052) and B (NPI-0047) in Complex with the 20S Proteasome Reveal Important Consequences of β-Lactone Ring Opening and a Mechanism for Irreversible Binding. *J Am Chem Soc* **2006**, *128* (15), 5136–5141. https://doi.org/10.1021/ja058320b.

(5) Gribble, G. Biological Activity of Recently Discovered Halogenated Marine Natural Products. *Mar Drugs* **2015**, *13* (7), 4044–4136. https://doi.org/10.3390/md13074044.

(6) Sweeney, D.; Chase, A. B.; Bogdanov, A.; Jensen, P. R. MAR4 *Streptomyces* : A Unique Resource for Natural Product Discovery. *J Nat Prod* **2024**, *87* (2), 439–452. https://doi.org/10.1021/acs.jnatprod.3c01007.

(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. Marinocyanins, Cytotoxic Bromo-Phenazinone Meroterpenoids from a Marine Bacterium from the Streptomycete Clade MAR4. *Tetrahedron* **2017**, *73* (16), 2234–2241. https://doi.org/10.1016/j.tet.2017.03.003.

(8) Kwon, H. C.; Espindola, A. P. D. M.; Park, J.-S.; Prieto-Davó, A.; Rose, M.; Jensen, P. R.; Fenical, W. Nitropyrrolins A–E, Cytotoxic Farnesyl-α-Nitropyrroles from a Marine-Derived Bacterium within

the Actinomycete Family Streptomycetaceae. *J Nat Prod* **2010**, *73* (12), 2047–2052. https://doi.org/10.1021/np1006229.

(9) Raju, R.; Piggott, A. M.; Barrientos Diaz, L. X.; Khalil, Z.; Capon, R. J. Heronapyrroles A–C: Farnesylated 2-Nitropyrroles from an Australian Marine-Derived Streptomyces Sp. *Org Lett* **2010**, *12* (22), 5158–5161. https://doi.org/10.1021/ol102162d.

(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. New Napyradiomycin Analogues from Streptomyces Sp. Strain CA-271078. *Mar Drugs* **2019**, *18* (1), 22. https://doi.org/10.3390/md18010022.

(11) Pereira, F.; Almeida, J. R.; Paulino, M.; Grilo, I. R.; Macedo, H.; Cunha, I.; Sobral, R. G.; Vasconcelos, V.; Gaudêncio, S. P. Antifouling Napyradiomycins from Marine-Derived Actinomycetes Streptomyces Aculeolatus. *Mar Drugs* **2020**, *18* (1), 63. https://doi.org/10.3390/md18010063.

(12) Cheng, Y. Bin; Jensen, P. R.; Fenical, W. Cytotoxic and Antimicrobial Napyradiomycins from Two Marine-Derived Streptomyces Strains. *European J Org Chem* **2013**, *2013* (18), 3751–3757. https://doi.org/10.1002/ejoc.201300349.

(13) Pathirana, C.; Jensen, P. R.; Fenical, W. Marinone and Debromomarinone: Antibiotic Sesquiterpenoid Naphthoquinones of a New Structure Class from a Marine Bacterium. *Tetrahedron Lett* **1992**, *33* (50), 7663–7666. https://doi.org/10.1016/0040-4039(93)88010-G.

(14) Hardt, I. H.; Jensen, P. R.; Fenical, W. Neomarinone, and New Cytotoxic Marinone Derivatives, Produced by a Marine Filamentous Bacterium (Actinomycetales). *Tetrahedron Lett* **2000**, *41* (13), 2073–2076. https://doi.org/10.1016/S0040-4039(00)00117-9.

(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. Intra-clade Metabolomic Profiling of MAR4 *Streptomyces* from the Macaronesia Atlantic Region Reveals a Source of Anti-biofilm Metabolites. *Environ Microbiol* **2019**, *21* (3), 1099–1112. https://doi.org/10.1111/1462-2920.14529.

(16) Gallagher, K. A.; Jensen, P. R. Genomic Insights into the Evolution of Hybrid Isoprenoid Biosynthetic Gene Clusters in the MAR4 Marine Streptomycete Clade. *BMC Genomics* **2015**, *16* (1), 960. https://doi.org/10.1186/s12864-015-2110-3.

(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. Molecular Networking and Pattern-Based Genome Mining Improves Discovery of Biosynthetic Gene Clusters and Their Products from Salinispora Species. *Chem Biol* **2015**, *22* (4), 460–471. https://doi.org/10.1016/j.chembiol.2015.03.010.

(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. Metabologenomics: Correlation of Microbial Gene Clusters with Metabolites Drives Discovery of a Nonribosomal Peptide with an Unusual Amino Acid Monomer. *ACS Cent Sci* **2016**, *2* (2), 99–108. https://doi.org/10.1021/acscentsci.5b00331.

(19) Mantovani, S. M.; Moore, B. S. Flavin-Linked Oxidase Catalyzes Pyrrolizine Formation of Dichloropyrrole-Containing Polyketide Extender Unit in Chlorizidine a. *J Am Chem Soc* **2013**, *135* (48), 18032–18035. https://doi.org/10.1021/ja409520v.

(20) Fu, C.; Xie, F.; Hoffmann, J.; Wang, Q.; Bauer, A.; Brönstrup, M.; Mahmud, T.; Müller, R. Armeniaspirol Antibiotic Biosynthesis: Chlorination and Oxidative Dechlorination Steps Affording Spiro[4.4]Non-8-Ene. *ChemBioChem* **2019**, *20* (6), 764–769. https://doi.org/10.1002/cbic.201800791.

(21) Yamanaka, K.; Ryan, K. S.; Gulder, T. A. M.; Hughes, C. C.; Moore, B. S. Flavoenzyme-Catalyzed Atropo-Selective N,C -Bipyrrole Homocoupling in Marinopyrrole Biosynthesis. *J Am Chem Soc* **2012**, *134* (30), 12434–12437. https://doi.org/10.1021/ja305670f.

(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. Ranking Microbial Metabolomic and Genomic Links in the NPLinker Framework Using Complementary Scoring Functions. *PLoS Comput Biol* **2021**, *17* (5), e1008920. https://doi.org/10.1371/journal.pcbi.1008920.

(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. NPOmix: A Machine Learning Classifier to Connect Mass Spectrometry Fragmentation Data to Biosynthetic Gene Clusters. *PNAS Nexus* **2022**, *1* (5), 1–15. https://doi.org/10.1093/pnasnexus/pgac257.

(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. Genomic and Spectroscopic Signature-Based Discovery of Natural Macrolactams. *J Am Chem Soc* **2023**, *145* (3), 1886–1896. https://doi.org/10.1021/jacs.2c11527.

(25) WESTLEY, J. W.; EVANS Jr., R. H.; SELLO, L. H.; TROUPE, N. Isolation and Characterization of Antibiotic X-14547A, a Novel Monocarboxylic Acid Ionophore Produced by Streptomyces Antibioticus NRRL 8167. *J Antibiot (Tokyo)* **1979**, *32* (2), 100–107. https://doi.org/10.7164/antibiotics.32.100.

(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. X-14547A, a New Ionophorous Antibiotic Produced by Streptomyces Antibioticus NRRL 8167. Discovery, Fermentation, Biological Properties and Taxonomy of the Producing Culture. *J Antibiot (Tokyo)* **1979**, *32* (2), 95–99. https://doi.org/10.7164/antibiotics.32.95.

(27) Li, C.; Roege, K. E.; Kelly, W. L. Analysis of the Indanomycin Biosynthetic Gene Cluster from Streptomyces Antibioticus NRRL 8167. *ChemBioChem* **2009**, *10* (6), 1064–1072. https://doi.org/10.1002/cbic.200800822.

(28) Roege, K. E.; Kelly, W. L. Biosynthetic Origins of the Ionophore Antibiotic Indanomycin. *Org Lett* **2009**, *11* (2), 297–300. https://doi.org/10.1021/ol802422n.

(29) ZHANG, D.; G. NAIR, M.; MURRY, M.; ZHANG, Z. Insecticidal Activity of Indanomycin. *J Antibiot (Tokyo)* **1997**, *50* (7), 617–620. https://doi.org/10.7164/antibiotics.50.617.

(30) Aherfi, S.; Pradines, B.; Devaux, C.; Honore, S.; Colson, P.; Scola, B. La; Raoult, D. Drug Repurposing against SARS-CoV-1, SARS-CoV-2 and MERS-CoV. *Future Microbiol* **2021**, *16* (17), 1341–1370. https://doi.org/10.2217/fmb-2021-0019.

(31) Murenets, N. V; Kudinova, M. K.; Korobkova, T. P.; Drobysheva, T. N.; Kliuev, N. A. [Kafamycin--a New Pyrrol Ether Antibiotic]. *Antibiot Med Biotekhnol* **1987**, *32* (11), 811–814.

(32) LARSEN, S. H.; BOECK, L. V. D.; MERTZ, F. P.; PASCHAL, J. W.; OCCOLOWITZ, J. L. 16-Deethylindanomycin (A83094A), a Novel Pyrrole-Ether Antibiotic Produced by a Strain of Streptomyces Setonii. Taxonomy, Fermentation, Isolation and Characterization. *J Antibiot (Tokyo)* **1988**, *41* (9), 1170–1177. https://doi.org/10.7164/antibiotics.41.1170.

(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]. *Vopr Med Khim* **1990**, *36* (5), 39–41.

(34) Toth, Piroska; Szell, Valeria; Koczka, Istvan; Horvath, Gyula; Szabo, Istvan Mihaly; Szayly, Marta; Ambrus, Gabor; Bedo, Julianna; Berdy, J. Producing Indanomycin and Homoindanomycin Using Streptomyces Galbus. HU49909 A2, 1989.

(35) Miao, S.; Anstee, M. R.; Baichwal, V.; Park, A. Stawamycin, a New Pyrroloketoindane Natural Product from the Cultures of Streptomyces Sp. *Tetrahedron Lett* **1995**, *36* (32), 5699–5702. https://doi.org/10.1016/0040-4039(95)01127-4.

(36) Izumikawa, M.; Komaki, H.; Hashimoto, J.; Takagi, M.; Shin-ya, K. Stawamycin Analog, JBIR-11 from Streptomyces Viridochromogenes Subsp. Sulfomycini NBRC 13830. *Journal of Antibiotics* **2008**, *61* (5), 326–329. https://doi.org/10.1038/ja.2008.47.

(37) Beloeil, J. C.; Delsuc, M. A.; Lallemand, J. Y.; Dauphin, G.; Jeminet, G. Application of Homonuclear and Heteronuclear Two-Dimensional Chemical-Shift Correlation NMR Spectroscopy to the Complete Assignment of Proton and Carbon-13 NMR Spectra of Ionophorous Antibiotic X.14547 A. *J Org Chem* **1984**, *49* (10), 1797–1800. https://doi.org/10.1021/jo00184a025.

(38) Buchfink, B.; Reuter, K.; Drost, H.-G. Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND. *Nat Methods* **2021**, *18* (4), 366–368. https://doi.org/10.1038/s41592-021-01101-x.

(39) Gilchrist, C. L. M.; Booth, T. J.; van Wersch, B.; van Grieken, L.; Medema, M. H.; Chooi, Y. Cblaster: A Remote Search Tool for Rapid Identification and Visualization of Homologous Gene Clusters. *Bioinformatics Advances* **2021**, *1* (1), 1–10. https://doi.org/10.1093/bioadv/vbab016.

(40) Terlouw, B. R.; Blin, K.; Navarro-Muñoz, J. C.; Avalon, N. E.; Chevrette, M. G.; Egbert, S.; Lee, S.; Meijer, D.; Recchia, M. J. J.; Reitz, Z. L.; van Santen, J. A.; Selem-Mojica, N.; Tørring, T.; Zaroubi, L.; Alanjary, M.; Aleti, G.; Aguilar, C.; Al-Salihi, S. A. A.; Augustijn, H. E.; Avelar-Rivas, J. A.; Avitia-Domínguez, L. A.; Barona-Gómez, F.; Bernaldo-Agüero, J.; Bielinski, V. A.; Biermann, F.; Booth, T. J.; Carrion Bravo, V. J.; Castelo-Branco, R.; Chagas, F. O.; Cruz-Morales, P.; Du, C.; Duncan, K. R.; Gavriilidou, A.; Gayrard, D.; Gutiérrez-García, K.; Haslinger, K.; Helfrich, E. J. N.; van der Hooft, J. J.

J.; Jati, A. P.; Kalkreuter, E.; Kalyvas, N.; Kang, K. Bin; Kautsar, S.; Kim, W.; Kunjapur, A. M.; Li, Y.-X.; Lin, G.-M.; Loureiro, C.; Louwen, J. J. R.; Louwen, N. L. L.; Lund, G.; Parra, J.; Philmus, B.; Pourmohsenin, B.; Pronk, L. J. U.; Rego, A.; Rex, D. A. B.; Robinson, S.; Rosas-Becerra, L. R.; Roxborough, E. T.; Schorn, M. A.; Scobie, D. J.; Singh, K. S.; Sokolova, N.; Tang, X.; Udwary, D.; Vigneshwari, A.; Vind, K.; Vromans, S. P. J. M.; Waschulin, V.; Williams, S. E.; Winter, J. M.; Witte, T. E.; Xie, H.; Yang, D.; Yu, J.; Zdouc, M.; Zhong, Z.; Collemare, J.; Linington, R. G.; Weber, T.; Medema, M. H. MIBiG 3.0: A Community-Driven Effort to Annotate Experimentally Validated Biosynthetic Gene Clusters. *Nucleic Acids Res* **2023**, *51* (D1), D603–D610. https://doi.org/10.1093/nar/gkac1049.

(41) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. AntiSMASH 5.0: Updates to the Secondary Metabolite Genome Mining Pipeline. *Nucleic Acids Res* **2019**, *47* (W1), W81–W87. https://doi.org/10.1093/nar/gkz310.

(42) Purdy, T. N.; Kim, M. C.; Cullum, R.; Fenical, W.; Moore, B. S. Discovery and Biosynthesis of Tetrachlorizine Reveals Enzymatic Benzylic Dehydrogenation via an Ortho -Quinone Methide. *J Am Chem Soc* **2021**, *143* (10), 3682–3686. https://doi.org/10.1021/jacs.0c12415.

(43) Ziemert, N.; Podell, S.; Penn, K.; Badger, J. H.; Allen, E.; Jensen, P. R. The Natural Product Domain Seeker NaPDoS: A Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. *PLoS One* **2012**, *7* (3), e34064. https://doi.org/10.1371/journal.pone.0034064.

(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. The Natural Product Domain Seeker Version 2 (NaPDoS2) Webtool Relates Ketosynthase Phylogeny to Biosynthetic Function. *Journal of Biological Chemistry* **2022**, *298* (10), 102480. https://doi.org/10.1016/j.jbc.2022.102480.

(45) Kwan, D. H.; Schulz, F. The Stereochemistry of Complex Polyketide Biosynthesis by Modular Polyketide Synthases. *Molecules* **2011**, *16* (7), 6092–6115. https://doi.org/10.3390/molecules16076092.

(46) Keatinge-Clay, A. T. The Uncommon Enzymology of Cis-Acyltransferase Assembly Lines. *Chem Rev* **2017**, *117* (8), 5334–5366. https://doi.org/10.1021/acs.chemrev.6b00683.

(47) Cantu, D. C.; Chen, Y.; Reilly, P. J. Thioesterases: A New Perspective Based on Their Primary and Tertiary Structures. *Protein Science* **2010**, *19* (7), 1281–1295. https://doi.org/10.1002/pro.417.

(48) Cantu, D. C.; Chen, Y.; Lemons, M. L.; Reilly, P. J. ThYme: A Database for Thioester-Active Enzymes. *Nucleic Acids Res* **2011**, *39* (Database), D342–D346. https://doi.org/10.1093/nar/gkq1072.

(49) Ichikawa, N.; Sasagawa, M.; Yamamoto, M.; Komaki, H.; Yoshida, Y.; Yamazaki, S.; Fujita, N. DoBISCUIT: A Database of Secondary Metabolite Biosynthetic Gene Clusters. *Nucleic Acids Res* **2013**, *41* (D1), 408–414. https://doi.org/10.1093/nar/gks1177.

(50) Rommel, K. R.; Li, C.; Kelly, W. L. Identification of a Tetraene-Containing Product of the Indanomycin Biosynthetic Pathway. *Org Lett* **2011**, *13* (10), 2536–2539. https://doi.org/10.1021/ol200570u.

(51) Drulyte, I.; Obajdin, J.; Trinh, C. H.; Kalverda, A. P.; van der Kamp, M. W.; Hemsworth, G. R.; Berry, A. Crystal Structure of the Putative Cyclase IdmH from the Indanomycin Nonribosomal Peptide Synthase/Polyketide Synthase. *IUCrJ* **2019**, *6* (6), 1120–1133. https://doi.org/10.1107/S2052252519012399.

(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. Enzymatic Assembly of the Salinosporamide γ-Lactam-β-Lactone Anticancer Warhead. *Nat Chem Biol* **2022**, *18* (5), 538–546. https://doi.org/10.1038/s41589-022-00993-w.

(53) Roush, W. R.; Myers, A. G. Antibiotic X-14547A: Total Synthesis of the Right-Hand Half. *J Org Chem* **1981**, *46* (7), 1509–1511. https://doi.org/10.1021/jo00320a060.

(54) Rousch, W. R.; Peseckis, S. M. Studies on the Total Synthesis of Antibiotic X-14547A: The Pentaene Approach. *Tetrahedron Lett* **1982**, *23* (47), 4879–4882. https://doi.org/10.1016/S0040-4039(00)85737-8.

(55) Roush, W. R.; Peseckis, S. M.; Walts, A. E. Synthesis of Antibiotic X-14547A. *J Org Chem* **1984**, *49* (18), 3429–3432. https://doi.org/10.1021/jo00192a052.

(56) Stocking, E. M.; Williams, R. M. Chemie Und Biologie Natürlicher Diels-Alder-Reaktionen. *Angewandte Chemie* **2003**, *115* (27), 3186–3223. https://doi.org/10.1002/ange.200200534.

(57) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, 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. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(58) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making Protein Folding Accessible to All. *Nat Methods* **2022**, *19* (6), 679–682. https://doi.org/10.1038/s41592-022-01488-1.

(59) Valdés-Tresanco, M. S.; Valdés-Tresanco, M. E.; Valiente, P. A.; Moreno, E. AMDock: A Versatile Graphical Tool for Assisting Molecular Docking with Autodock Vina and Autodock4. *Biol Direct* **2020**, *15* (1), 12. https://doi.org/10.1186/s13062-020-00267-2.

(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. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **2022**, *50* (D1), D20–D26. https://doi.org/10.1093/nar/gkab1112.

(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.; Shalapour, S.; Kyle, K. E.; Puckett, S. P.; Watrous, J. D.; Carpenter, C. S.; Bouslimani, A.; Ernst, M.; Swafford, A. D.; Zúñiga, E. I.; Balunas, M. J.; Klassen, J. L.; Loomba, R.; Knight, R.;

Bandeira, N.; Dorrestein, P. C. Mass Spectrometry Searches Using MASST. *Nat Biotechnol* **2020**, *38* (1), 23–26. https://doi.org/10.1038/s41587-019-0375-9.

(62) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. Ø.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **2016**, *34* (8), 828–837. https://doi.org/10.1038/nbt.3597.

(63) Nonejuie, P.; Burkart, M.; Pogliano, K.; Pogliano, J. Bacterial Cytological Profiling Rapidly Identifies the Cellular Pathways Targeted by Antibacterial Molecules. *Proceedings of the National Academy of Sciences* **2013**, *110* (40), 16169–16174. https://doi.org/10.1073/pnas.1311066110.

(64) Vila, J.; Martí, S.; Sánchez-Céspedes, J. Porins, Efflux Pumps and Multidrug Resistance in Acinetobacter Baumannii. *Journal of Antimicrobial Chemotherapy* **2007**, *59* (6), 1210–1215. https://doi.org/10.1093/jac/dkl509.

(65) Soto, S. M. Role of Efflux Pumps in the Antibiotic Resistance of Bacteria Embedded in a Biofilm. *Virulence* **2013**, *4* (3), 223–229. https://doi.org/10.4161/viru.23724.

(66) Work, R. BBMap : A Fast , Accurate , Splice-Aware Aligner. **2014**, 3–5.

(67) Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A. A.; Dvorkin, M.; Kulikov, A. S.; Lesin, V. M.; Nikolenko, S. I.; Pham, S.; Prjibelski, A. D.; Pyshkin, A. V.; Sirotkin, A. V.; Vyahhi, N.; Tesler, G.; Alekseyev, M. A.; Pevzner, P. A. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **2012**, *19* (5), 455–477. https://doi.org/10.1089/cmb.2012.0021.

(68) Jain, C.; Rodriguez-R, L. M.; Phillippy, A. M.; Konstantinidis, K. T.; Aluru, S. High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries. *Nat Commun* **2018**, *9* (1), 5114. https://doi.org/10.1038/s41467-018-07641-9.

(69) Wick, R. R.; Judd, L. M.; Gorrie, C. L.; Holt, K. E. Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads. *PLoS Comput Biol* **2017**, *13* (6), 1–22. https://doi.org/10.1371/journal.pcbi.1005595.

(70) Parks, D. H.; Imelfort, M.; Skennerton, C. T.; Hugenholtz, P.; Tyson, G. W. CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. *Genome Res* **2015**, *25* (7), 1043–1055. https://doi.org/10.1101/gr.186072.114.

(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. A Computational Framework to Explore Large-Scale Biosynthetic Diversity. *Nat Chem Biol* **2020**, *16* (1), 60–68. https://doi.org/10.1038/s41589-019-0400-9.

(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. MIBiG 2.0: A Repository for Biosynthetic Gene Clusters of Known Function. *Nucleic Acids Res* **2020**, *48* (D1), D454–D458. https://doi.org/10.1093/nar/gkz882.

(73) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **2003**, *13* (11), 2498–2504. https://doi.org/10.1101/gr.1239303.

(74) Edgar, R. C. Muscle5: High-Accuracy Alignment Ensembles Enable Unbiased Assessments of Sequence Homology and Phylogeny. *Nat Commun* **2022**, *13* (1), 6968. https://doi.org/10.1038/s41467-022-34630-w.

(75) Valdés-Tresanco, M. S.; Valdés-Tresanco, M. E.; Valiente, P. A.; Moreno, E. AMDock: A Versatile Graphical Tool for Assisting Molecular Docking with Autodock Vina and Autodock4. *Biol Direct* **2020**, *15* (1), 12. https://doi.org/10.1186/s13062-020-00267-2.

(76) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.

(77) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model* **2021**, *61* (8), 3891–3898. https://doi.org/10.1021/acs.jcim.1c00203.