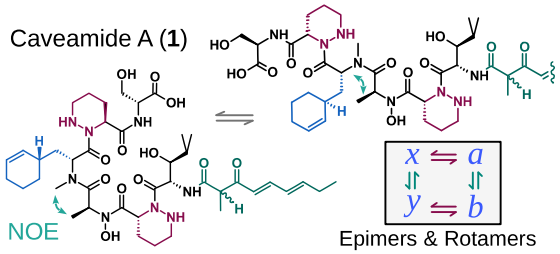
Piperazate-Guided Isolation of Caveamides A and B, Cyclohexenylalanine-Containing Non-ribosomal Peptides from a Cave Actinomycete

Mostafa Hagar[1], Kalindi D. Morgan[1,5], Spencer D. Stumpf[3], Maya Tsingos[3], Carmen A. Banuelos[4], Marianne D. Sadar[4], Joshua A. V. Blodgett[3], Raymond J. Andersen[1,2,*], and Katherine S. Ryan[1,*]

[1] Department of Chemistry, The University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4
[2] Department of Earth, Ocean, and Atmospheric Sciences, The University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4
[3] Department of Biology, Washington University in St Louis, St Louis MO 63122
[4] Department of Genome Sciences, BC Cancer, Vancouver, British Columbia, Canada V5Z 1L3
[5] Current address: Department of Chemistry, University of Northern British Columbia, Prince George, British Columbia V2N 4Z9

Supporting Information Placeholder

ABSTRACT: Hybrid genome-mining/^{15}N-NMR was used to target compounds containing piperazate (Piz) residues, leading to the discovery of caveamides A (**1**) and B (**2**) from *Streptomyces* sp. strain BE230, isolated from New Rankin Cave (Missouri). Caveamides are highly dynamic molecules containing an unprecedented β-ketoamide polyketide fragment, two Piz residues, and a new *N*-methyl-cyclohexenylalanine residue. Caveamide B (**2**) exhibited nanomolar cytotoxicity against several cancer cell lines and nanomolar anti-microbial activity against MRSA and *E. coli*.



Biologically active microbial natural products are an important source of inspiration for the development of new anti-microbial and anti-cancer drugs, but compound rediscovery is a frequent obstacle to finding new natural product drug leads. Contemporary isolation campaigns address this issue with a variety of strategies, including early upstream dereplication,[1] screening with novel bioassays,[2] sampling unusual organisms from underexplored habitats,[3] and mining genomes for novel biosynthetic gene clusters (BGCs).[4]

Our groups have previously reported a combined genome-mining/^{15}N-NMR approach for the targeted isolation of molecules containing piperazate (Piz), a non-proteinogenic amino acid comprising a cyclic hydrazine.[5,6] Piz is incorporated in a diverse set of non-ribosomal peptides (NRPs) and hybrid polyketide-NRPs (PK/NRPs) which frequently exhibit anti-microbial, cytotoxic, anti-parasitic, and antiviral activities.[7] Furthermore, most reported Piz-containing compounds have other uncommon structural features such as new polyketide fragments, dimeric aryl-aryl coupling, and other non-proteinogenic amino acids.[7] We previously proposed that Piz-containing natural products represent a pool of compounds enriched with unusual chemistry and with biological activity,[7] and that targeting Piz-containing compounds for isolation is likely to yield interesting molecules.[5–7]

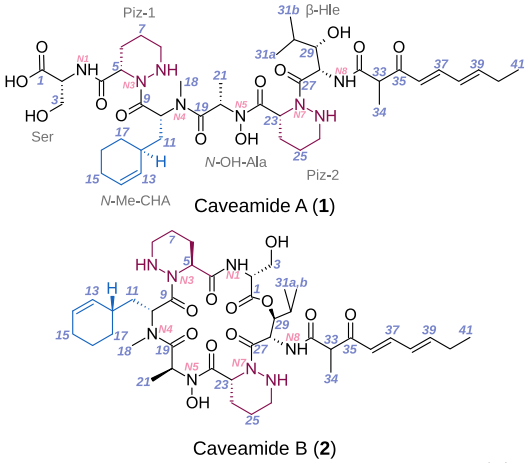


Figure 1. Enumerated structures of caveamides A (**1**) and B (**2**).

Here, we present a validation of this approach, and report the isolation of compounds which, in addition to exhibiting potent bioactivity, also feature a new cyclohexenylalanine (CHA) amino acid residue. We describe challenging structural elucidation, an approach for configurational assignment in cyclic alkene residues, and nanomolar anti-microbial and cytotoxic activity. Finally, we further discuss implications about CHA biosynthesis and set the stage for further investigation of the elusive family to which CHA belongs.

Streptomyces sp. strain BE230 was collected from New Rankin Cave in Eureka, MO as described elsewhere.[9] The strain was

identified as a potential Piz producer via an established PCR screen (using degenerate oligonucleotides LMPzbB FN2 and LMPzbB RN1)[10] that targets Piz synthases. Revealing BE230 as a candidate producer, this triggered genome-sequencing of the strain (described elsewhere[9]) plus LC/MS^2 scanning for Piz-associated mass features in growth extracts (essentially as described[11]). The production of Piz compounds by BE230 was finally confirmed by the detection of Piz-characteristic ^{15}N NMR resonances upon cultivation with $^{15}N_2$-L-ornithine (L-Orn). We have previously reported the use of ^{1}H-^{15}N-HSQC and ^{1}H-^{15}N-HSQC-TOCSY to screen for Piz residues, and used this approach to isolate incarnatapeptins A and B,[5] and dentigerumycins F and G.[6] Here, we also deployed 1D ^{1}H-^{15}N-HSQC experiments, acquired in a fraction of the time as 2D experiments, to pre-empt committing to longer data acquisitions (**S7**).

Production cultures of BE230 were grown as lawns on solid YMS agar supplemented with L-Orn. A 'tracer' extract, prepared from a batch supplemented with $^{15}N_2$-L-Orn, was combined with larger unlabeled crude extracts to amplify the characteristic Piz resonances depicted in **S7**. After several rounds of separation (**S8**), a final HPLC purification step yielded pure **1** and **2** (**Figure 1**) as glassy solids. HRMS (ESI) (**1**) *m/z*: $[M + H]^+$ Calcd for $C_{42}H_{67}N_8O_{12}$ 875.4873; Found 875.4875. (**2**) *m/z*: $[M + Na]^+$ Calcd for $C_{42}H_{64}N_8O_{11}Na$ 879.4592; Found 879.4588.

Compounds **1** and **2** each elute as a pair of HPLC peaks, and isolation of the separated peaks followed by reinjection gives the originating pair of chromatographic peaks (**Figure**

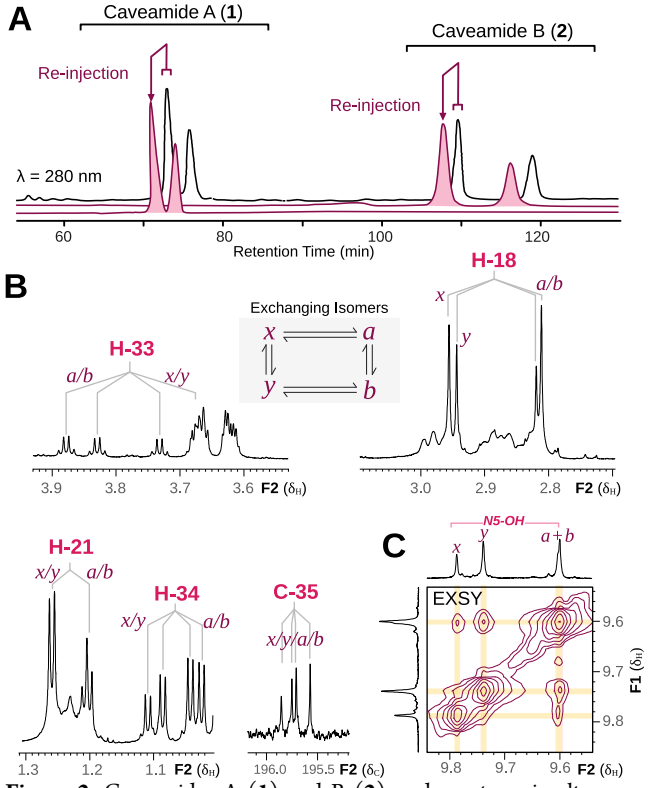


Figure 2. Caveamides A (**1**) and B (**2**) undergo two simultaneous exchange processes. (a) HPLC chromatograms depicting **1** and **2** before purification, and after isolation and subsequent re-injection of separated isomers. (b) Insets from ^{1}H and ^{13}C NMR data of **1** acquired in DMSO-d_6 at 850 MHz, showing exchanging sets of resonances. The isomers of **1** are designated as *x*, *y*, *a*, and *b*. (c) An Exchange Spectroscopy (EXSY) experiment inset of the hydroxamic acid OH resonances, where evidence of exchange is most clearly seen.

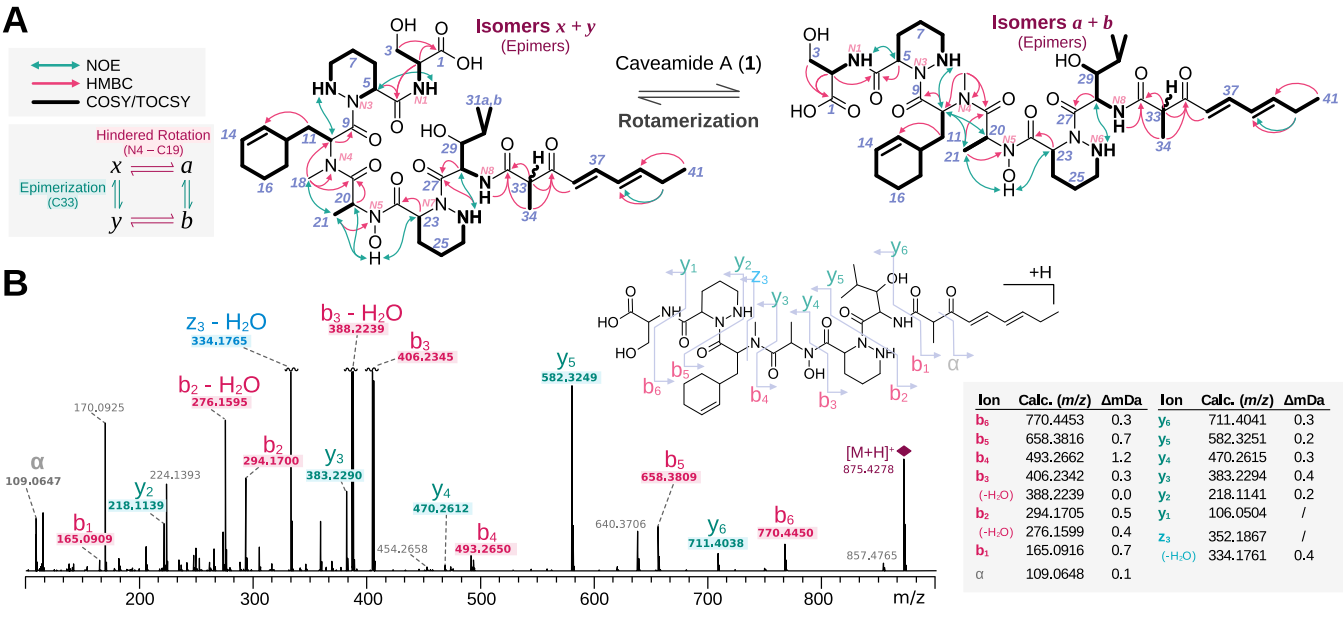


Figure 3: Determination of caveamide A (**1**) structure. (**a**) Key NMR correlations, with two pairs of rotamers distinguishable by NOE correlations in the *N*-OH-Ala → Piz-2 region. (**b**) Our proposed structure is corroborated by HRMS/MS fragmentation data acquired by collision-induced dissociation (CID) at 25 eV, with positive-mode electrospray ionization (ESI+). Peptide fragment ions are labeled according to standard nomenclature,[8] except *α* which denotes *α*-cleavage at the C-33/35 bond.

2A). The material in each of the peaks give identical NMR spectra, and in **1** they can be assigned to four roughly equally abundant isomeric species (designated as: *x, y, a, b*) (**Figure 2B**).

Structural assignments were made with additional information from gCOSY, tROESY, gHSQC, and gHMBC experiments. NMR data interpretation was challenging due to the isomer interconversion. Despite the spectral complexity, we were able to determine a full structure. The challenge is exemplified by a linear hexa-1,3-dienyl spin-system (C-36 to C-41) where each of the four alkene proton resonances, assigned H-36 through H-39, show two to four different chemical shifts with varying amounts of signal overlap. The spin-system can be disentangled based on HSQC and HMBC correlations (**Figure 3A**) and shown to extend into a saturated ethyl fragment (C-40 and C-41), also manifesting as multiple exchanging signals. An HMBC correlation between H-36 and a ketone carbon C-35, also producing four distinct chemical shifts (**Figure 2B**), reveals the spin-system to be an α,β,γ,δ-unsaturated ketone. ROESY correlations between CH-38 and CH_2-40 show the $\Delta^{38,39}$ alkene to be *E*, and comparative multiplet analysis of the CH-37 proton resonance at 600 MHz and 850 MHz establish the $\Delta^{36,37}$ alkene as *E* as well (**Figure S6**). A CH_3-CH fragment (C-33/C-34) shows HMBC correlations with both the ketone and four exchanging amide carbonyl resonances (C-32 δ_C 168.70–169.22), leading to the hypothesis that one of the isomerization processes is spontaneous epimerization at the acidic α-carbon (C-33). A comprehensive interpretation of NMR and MS data, described in **S37**, led to the identification of L-erythro-β-hydroxyleucine (βHle), D- and L-Piz residues, an *N*-OH-alanine residue, and D-serine. Identification of the *N*-Me-CHA residue was complicated by cyclic methylene signal overlap from exchanging isomers and Piz residues. We deduced its structure based on the molecular formula, a key HMBC correlation between CH_2-11/C-13, and later by comparison with synthesized standards. Differences in ROESY correlations between exchanging *N*-OH-Ala and *N*-Me-CHA systems lead us to propose the second isomerization process is hindered rotation about the *N*-4–C-19 tertiary amide bond. ROESY correlations between CH_3-18 and CH_3-21 are only seen in isomers *x*/*y*, and correlations between CH_3-21 and CH-10 are only seen in isomers a/b (**Figure 4A**).

Caveamide B (**2**) afforded NMR spectra with indications of other exchange processes likely unrelated to the N-4/C-19 amide. Likewise, incarnatapeptin B (**3**) was also reported as a mixture of conformers in the depsipeptide form,[5] suggesting that the shared polyketide–βHle–Piz-2 fragment induces a hindered rotation process when cyclized. Despite the complex NMR data, we identified key structural features, including an ester formed across the βHle residue (**S39**). Base-mediated hydrolysis of caveamide B (**2**) yields **1** (**S50**), leading to the conclusion that the latter is a *seco*- acid of the former. This is consistent with MS data, including CID fragmentation (**Figure S42**). Furthermore, reducing **2** with $NaBH_4$ results in D-serine no longer being detected by Marfey's method,[12] supporting our assignment (**S49**).

Most absolute configurations were straightforward to determine using Marfey's method (**S34**). However, acid-catalyzed isomerization and hydration of the *N*-Me-CHA residue during peptide hydrolysis led to scrambled stereochemistry (**Figure S30**). To circumvent this issue, we used stereoselective dibromination[13] to 'lock' the location of the alkene while preserving stereochemical information. After dibrominating **1**, the expected dibrominated CHA derivative was detected by Marfey's method (**Figure 4B**). All four possible isomers were then chemically synthesized and resolved (**S51** and **Figure 4B**). The configuration of caveamide-matching isomer was identified by NMR analysis of Marfey and Mosher[14] derivatives. The other three isomers were also assigned, enabling future researchers to determine *N*-Me-CHA configuration by preparing the dibrominated racemic standard and matching the elution order.

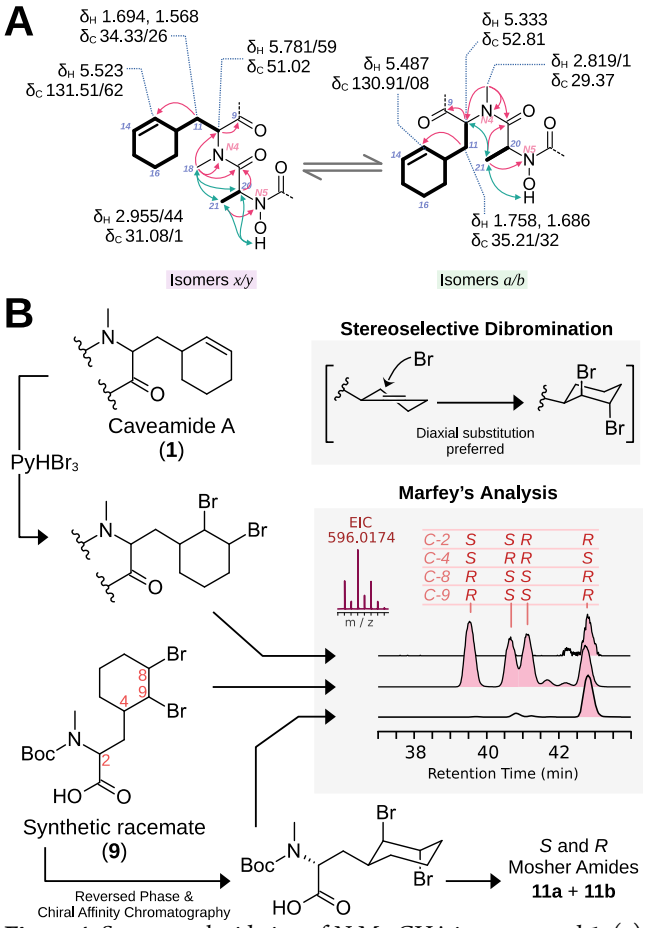


Figure 4. Structure elucidation of *N*-Me-CHA in compound **1**. (a) Key NMR correlations identifying *N*-Me-CHA. (b) Bromination using pyridinium tribromide leads to stereoselective diaxial substitution, converting *N*-Me-CHA into an acid-stable residue. Synthetic racemate was prepared, and isolation of the isomer matching that of compound **1** allowed configurational assignment by NMR analyses of its Marfey[12,15] and Mosher[14] derivatives (**S60** and **S65**).

Caveamides belong to the azinothricin family, 19-membered cyclic depsipeptides comprising the largest group of Piz-containing molecules[7] (examples given in **Figure S32**). Caveamides are distinct from the rest of the azinothricin family by having a CHA residue and an α,β,γ,δ-unsaturated ketone. We report the first instance of a free CHA residue in a peptide. The biosynthesis of CHA is not yet fully understood, though it is known to be an intermediate in the biosyntheses of salinosporamides[16] (**4**) and their closely related congeners, cinnabaramides.[17] CHA originates from a shunt pathway initiated by a prephenate decarboxylase (PDX), diverting prephenate from primary metabolism.[16] Typically, prephenate is decarboxylated and aromatized simultaneously *en route* to tyrosine or phenylalanine.[18] In contrast, PDXs effect decarboxylation but not aromatization, a challenging transformation yielding intermediates that have been implicated in the biosyntheses of bacilysin,[16] aeruginosins,[16] and 2,5-dihydrophenylalanine (2,5-H_2Phe, **5**),[19] in addition to salinosporamides[16] (**Figure 5**). None of these biosyntheses have yet been fully characterized, and no other natural products have been

experimentally verified to be PDX-derived. Therefore, caveamides likely represent an addition to a rare family of compounds, and provide another avenue for investigating the presently understudied biosynthesis of PDX-derived natural products.

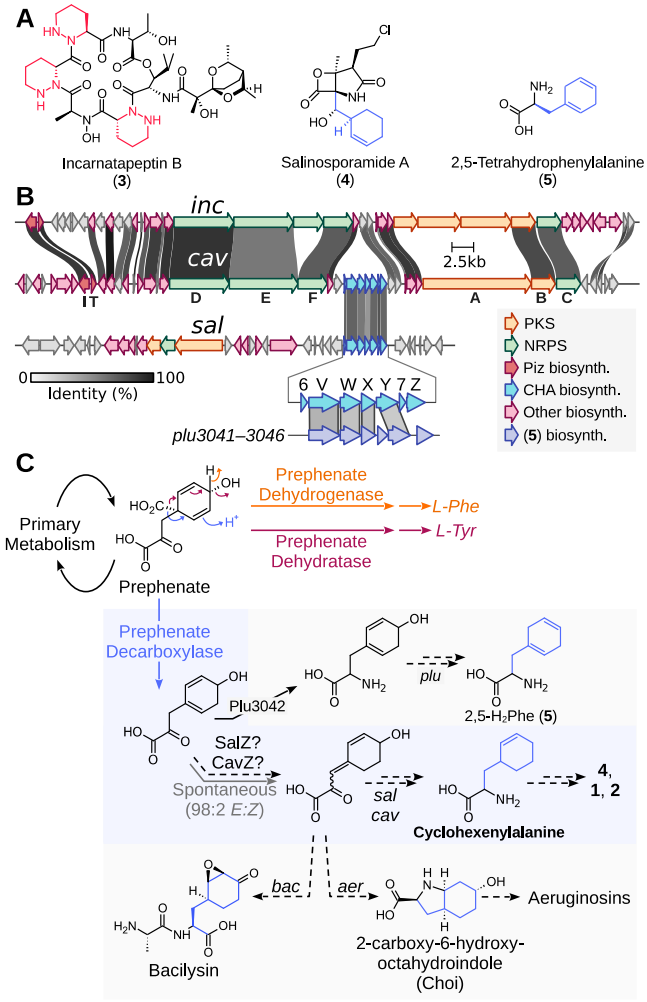


Figure 5: The putative caveamide BGC (*cav*) suggests a sub-cluster of *sal* genes is responsible for CHA production. (**a**) Caveamides share structural similarities with azinothricins such as incarnatapeptin B (**3**), and with salinosporamides (**4**). (**b**) Sequence similarities in the caveamide, incarnatapeptin, and salinosporamide BGCs (*cav, inc* and *sal*, respectively) reflect the structural similarities of their biosynthetic products. A putative CHA cassette is identified. Diagram generated using clinker,[20] highlighting protein sequence identities >30%. (**c**) Partially characterized biosynthetic pathways to PDX-derived moieties known thus far.

In addition to guiding further investigations of PDX-initiated biosyntheses, the isolation of caveamides and their putative BGC enabled us to identify several diverse groups of BGCs in public genomes bearing homologues to the putative CHA cassette, all co-localized with NRP synthetase genes (**Figure S66**). The biosynthetic products of these BGCs are not yet known and present further avenues for the investigation of CHA biosynthesis and/or isolation of structurally novel compounds.

Caveamides A (**1**) and B (**2**) were assayed for anti-microbial activity against a panel of pathogens, as well as cytotoxicity in HeLa, HepG2, LNCaP, and PC3 human cancer cell lines (**S68**). Where activity was detected, the depsipeptide **2** was markedly more potent than the *seco*- acid **1**, a common trend observed in macrocycles.[21] Caveamide B (**2**) showed nanomolar *in vitro* cytotoxicity against the human cancer cell lines, at 50–100 nM concentrations. It also showed *in vitro* anti-microbial activity against MRSA and Gram-negative *E. coli*. at 300 nM concentrations.

The potent bioactivity and unusual chemistry in caveamides underscore the value of targeting Piz-containing compounds for isolation. The discovery of **1** and **2** further demonstrates the utility of the ^{15}N-NMR/genome-mining approach for accessing a potentially enriched pool of natural products. We described a challenging structural elucidation, as both **1** and **2** exist as dynamically exchanging isomers in solution, and the hydrolytically unstable CHA residue necessitated dibromination and chemical synthesis to configurationally assign. Caveamides represent the first occurrence of CHA as a free peptide residue, and their discovery along with a putative BGC provides a critical 'second data point' identifying genes potentially involved in CHA production. Our findings will facilitate further investigation of CHA biosynthesis and the isolation of more natural products bearing this rare residue or analogues.

ASSOCIATED CONTENT

Data Availability Statement

The data underlying this study are available in the published article and its Supporting Information. The sequences reported in this paper have been deposited in the GenBank database with accession: JAQYWX000000000 (*Streptomyces* sp. strain BE230).

Supporting Information

Experimental procedures, compound characterizations, NMR spectra, MS data, chromatographic data, and genomic analysis data. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

* **Katherine S. Ryan** – *Department of Chemistry, The University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada*; orcid.org/0000-0003-1643-1940;
Email: ksryan@chem.ubc.ca

* **Raymond J. Andersen** – *Department of Chemistry and Department of Earth, Ocean, and Atmospheric Science, The University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada;* orcid.org/0000-0002-7607-8213;
Email: raymond.andersen@ubc.ca

Author Contributions

The manuscript was written through contributions of all authors.

ACKNOWLEDGMENT

The authors thank D. E. Williams (The University of British Columbia) for insightful discussions, J. Chen (The University of British Columbia) for MTT assays, and M. Okon (The University of British Columbia) for NMR spectrometer maintenance and advice. This work was funded in part by NSF-CAREER 1846995 to JAVB, and grants from NSERC (RJA, and RGPIN-2021-02626 to KSR).

REFERENCES

(1) Gaudêncio, S. P.; Pereira, F. Dereplication: Racing to Speed up the Natural Products Discovery Process. *Nat. Prod. Rep.* **2015**, *32* (6), 779–810. https://doi.org/10.1039/C4NP00134F.

(2) Fallarero, A.; Hanski, L.; Vuorela, P. How to Translate a Bioassay Into a Screening Assay for Natural Products: General Considerations and Implementation of Antimicrobial Screens. *Planta Med.*

2014, *80* (14), 1182–1199. https://doi.org/10.1055/s-0034-1383061.
(3) Bull, A. T.; Goodfellow, M. Dark, Rare and Inspirational Microbial Matter in the Extremobiosphere: 16 000 m of Bioprospecting Campaigns. *Microbiology* **2019**, *165* (12), 1252–1264. https://doi.org/10.1099/mic.0.000822.
(4) Bauman, K. D.; Butler, K. S.; Moore, B. S.; Chekan, J. R. Genome Mining Methods to Discover Bioactive Natural Products. *Nat. Prod. Rep.* **2021**, *38* (11), 2100–2129. https://doi.org/10.1039/D1NP00032B.
(5) Morgan, K. D.; Williams, D. E.; Patrick, B. O.; Remigy, M.; Banuelos, C. A.; Sadar, M. D.; Ryan, K. S.; Andersen, R. J. Incarnatapeptins A and B, Nonribosomal Peptides Discovered Using Genome Mining and 1H/15N HSQC-TOCSY. *Org. Lett.* **2020**, *22* (11), 4053–4057. https://doi.org/10.1021/acs.orglett.0c00818.
(6) Morgan, K. D.; Williams, D. E.; Ryan, K. S.; Andersen, R. J. Dentigerumycin F and G: Dynamic Structures Retrieved through a Genome-Mining/Nitrogen-NMR Methodology. *Tetrahedron Lett.* **2022**, *94*, 153688. https://doi.org/10.1016/j.tetlet.2022.153688.
(7) D. Morgan, K.; J. Andersen, R.; S. Ryan, K. Piperazic Acid-Containing Natural Products: Structures and Biosynthesis. *Nat. Prod. Rep.* **2019**, *36* (12), 1628–1653. https://doi.org/10.1039/C8NP00076J.
(8) Papayannopoulos, I. A. The Interpretation of Collision-induced Dissociation Tandem Mass Spectra of Peptides. *Mass Spectrom. Rev.* **1995**, *14* (1), 49–73. https://doi.org/10.1002/mas.1280140104.
(9) Harper, C. P.; Day, A.; Tsingos, M.; Ding, E.; Zeng, E.; Stumpf, S. D.; Qi, Y.; Robinson, A.; Greif, J.; Blodgett, J. A. V. Critical Analysis of Polycyclic Tetramate Macrolactam Biosynthetic Cluster Phylogeny and Functional Diversity. bioRxiv January 23, 2024, p 2024.01.22.576670. https://doi.org/10.1101/2024.01.22.576670.
(10) Hu, Y.; Qi, Y.; Stumpf, S. D.; D'Alessandro, J. M.; Blodgett, J. A. V. Bioinformatic and Functional Evaluation of Actinobacterial Piperazate Metabolism. *ACS Chem. Biol.* **2019**, *14* (4), 696–703. https://doi.org/10.1021/acschembio.8b01086.
(11) Li, C.; Hu, Y.; Wu, X.; Stumpf, S. D.; Qi, Y.; D'Alessandro, J. M.; Nepal, K. K.; Sarotti, A. M.; Cao, S.; Blodgett, J. A. V. Discovery of Unusual Dimeric Piperazyl Cyclopeptides Encoded by a *Lentzea Flaviverrucosa* DSM 44664 Biosynthetic Supercluster. *Proc. Natl. Acad. Sci.* **2022**, *119* (17), e2117941119. https://doi.org/10.1073/pnas.2117941119.
(12) Marfey, P. Determination of D-Amino Acids. II. Use of a Bifunctional Reagent, 1,5-Difluoro-2,4-Dinitrobenzene. *Carlsberg Res. Commun.* **1984**, *49* (6), 591. https://doi.org/10.1007/BF02908688.
(13) Bellucci, G.; Bianchini, R.; Vecchiani, S. Comparison of Molecular Bromine and Tribromide Ion as Brominating Reagents. 2. Kinetic and Product Investigation of the Bromination of 3-Substituted Cyclohexenes. *J. Org. Chem.* **1986**, *51* (22), 4224–4232. https://doi.org/10.1021/jo00372a023.
(14) Dale, J. A.; Mosher, H. S. Nuclear Magnetic Resonance Enantiomer Regents. Configurational Correlations via Nuclear Magnetic Resonance Chemical Shifts of Diastereomeric Mandelate, O-Methylmandelate, and .Alpha.-Methoxy-.Alpha.-Trifluoromethylphenylacetate (MTPA) Esters. *J. Am. Chem. Soc.* **1973**, *95* (2), 512–519. https://doi.org/10.1021/ja00783a034.
(15) Fujii, K.; Ikai, Y.; Oka, H.; Suzuki, M.; Harada, K. A Nonempirical Method Using LC/MS for Determination of the Absolute Configuration of Constituent Amino Acids in a Peptide: Combination of Marfey's Method with Mass Spectrometry and Its Practical Application. *Anal. Chem.* **1997**, *69* (24), 5146–5151. https://doi.org/10.1021/ac970289b.
(16) Mahlstedt, S.; Fielding, E. N.; Moore, B. S.; Walsh, C. T. Prephenate Decarboxylases: A New Prephenate-Utilizing Enzyme Family That Performs Nonaromatizing Decarboxylation En Route to Diverse Secondary Metabolites. *Biochemistry* **2010**, *49* (42), 9021–9023. https://doi.org/10.1021/bi101457h.
(17) Rachid, S.; Huo, L.; Herrmann, J.; Stadler, M.; Köpcke, B.; Bitzer, J.; Müller, R. Mining the Cinnabaramide Biosynthetic Pathway to Generate Novel Proteasome Inhibitors. *Chembiochem Eur. J. Chem. Biol.* **2011**, *12* (6), 922–931. https://doi.org/10.1002/cbic.201100024.
(18) Shende, V. V.; Bauman, K. D.; Moore, B. S. The Shikimate Pathway: Gateway to Metabolic Diversity. *Nat. Prod. Rep.* **2024**. https://doi.org/10.1039/D3NP00037K.
(19) Crawford, J. M.; Mahlstedt, S. A.; Malcolmson, S. J.; Clardy, J.; Walsh, C. T. Dihydrophenylalanine: A Prephenate-Derived Photorhabdus Luminescens Antibiotic and Intermediate in Dihydrostilbene Biosynthesis. *Chem. Biol.* **2011**, *18* (9), 1102–1112. https://doi.org/10.1016/j.chembiol.2011.07.009.
(20) Gilchrist, C. L. M.; Chooi, Y.-H. Clinker & Clustermap.Js: Automatic Generation of Gene Cluster Comparison Figures. *Bioinformatics* **2021**, *37* (16), 2473–2475. https://doi.org/10.1093/bioinformatics/btab007.
(21) Roxin, Á.; Zheng, G. Flexible or Fixed: A Comparative Review of Linear and Cyclic Cancer-Targeting Peptides. *Future Med. Chem.* **2012**, *4* (12), 1601–1618. https://doi.org/10.4155/fmc.12.75.