

Structure Determination and Biosynthesis of Dapalides A–C,¹ Glycosylated Kahalalide F Analogues from the Marine Cyanobacterium *Dapis* sp

Emma K. Ellis, Laura P. Ióca, Jie Liu, Manyun Chen, Steven D. Bruner, Yousong Ding, Valerie J. Paul,² Mohamed S. Donia,* and Hendrik Luesch*



Cite This: *J. Nat. Prod.* 2025, 88, 2138–2150 3



Read Online

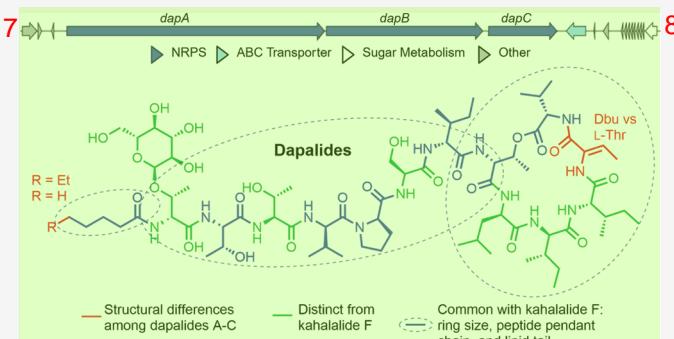
ACCESS | 4

Metrics & More | 5

Article Recommendations | 6

Supporting Information

ABSTRACT: Kahalalides were originally isolated from the marine mollusk *Elysia rufescens* and its green algal diet *Bryopsis* sp., but the true producer was revealed as the obligate bacterial symbiont *Candidatus Endobryopsis kahalalidefaciens*, residing within *Bryopsis* sp. The most notable is kahalalide F, a broad-spectrum antitumor depsipeptide that entered the clinic but failed from lack of efficacy. We have isolated three new glycosylated analogues of kahalalide F, termed dapalides A–C (**1–3**), from a marine cyanobacterium, *Dapis* sp., collected from Guam. The planar structures were determined by extensive NMR coupled with mass spectrometry. Acid hydrolysis of **1** using amino acid analysis revealed the absolute configuration of singlet and a mixture of duplicate amino acids. Metagenomic analysis unveiled a biosynthetic gene cluster (BGC) with a nonribosomal peptide synthetase (NRPS) system and downstream glycosylation enzymes, which assisted the configurational assignment through epimerization domain analysis. The discovered BGC, termed *dap*, was assigned to a high-quality metagenome-assembled genome of the *Dapis* sp. Dapalide A (**1**) was subjected to phenotypic bioassays and exhibited weak anticancer cytotoxicity. This discovery expands the chemical diversity of the kahalalide F family, suggests their broad ecological role across diverse organisms, and presents an intriguing case of natural product biosynthesis evolution.



Marine organisms produce specialized metabolites with structural diversity, complexity, and various biological activity. Decades of explorations of marine micro- and macroorganisms have culminated into the discovery of over 40,000 marine natural products (MNPs) and the development of 15 FDA approved drugs for various indications.^{1,2} Some of the most structurally remarkable and biologically active metabolites are those produced by marine cyanobacteria, which can create polyketides, polyketide-peptide hybrids, fatty acid amides, terpenes, alkaloids, and modified peptides from their biosynthetic gene clusters (BGCs), which have been reviewed extensively over the years.^{3–5} Genome mining and metabolomic analysis have become the forefront in determining the biosynthetic potential and chemical diversity produced from a given species.^{6,7} Notably, there is a rise in the number of previously isolated metabolites from collected sources that are biosynthesized in a cyanobacterial/bacterial metabolic source living in symbiosis within the host or in the surrounding area.^{6,8} Cyanobacteria are an ancient and rich source of diverse specialized metabolites that possess activity in different biological contexts, and these metabolites have been shown to move through food webs. Notable examples are the anticancer dolastatins such as dolastatin 10, originally isolated

from the marine sea hare *Dolabella auricularia*. Their true producer was found to be a food source of the sea hare, the cyanobacterium *Caldora penicillata*.^{9–12} Swinholidine A, a potent actin cytoskeleton disrupter, represents another example that was isolated from the marine sponge *Theonella swinhonis*; the true producer was later found to be an “Entotheonella” sponge symbiont.^{13,14} Notably, the same family of compounds were also isolated from marine cyanobacteria.^{15,16} Similar examples also exist beyond the cyanobacteria. The kahalalides were originally isolated from an herbivorous sacoglossan mollusk *Elysia rufescens*¹⁷ and later discovered within the diet of the mollusk, the green alga *Bryopsis* sp.¹⁸ However, further investigative work through metagenomic analysis, revealed the true producer of the kahalalides to be *Candidatus*

Received: June 19, 2025

Revised: August 12, 2025

Accepted: August 13, 2025

Published: August 26, 2025



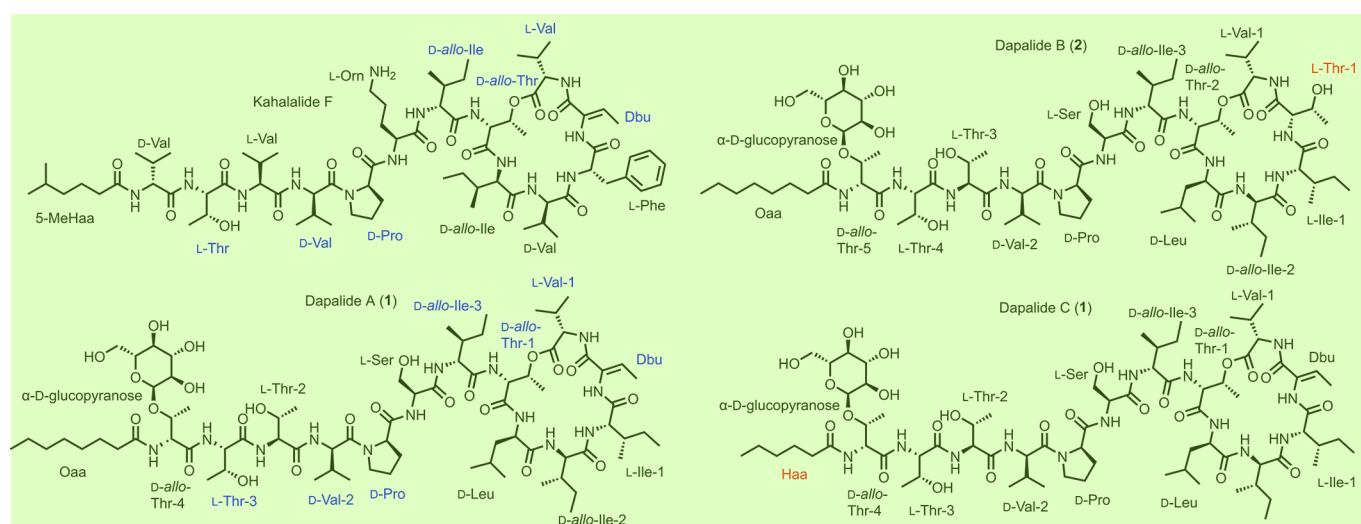


Figure 1. Structures of kahalalide F and dapalides A–C (**1–3**) with the amino acid similarities of dapalide A (**1**) and kahalalide F labeled blue and the structural differences of dapalides B (**2**) and C (**3**) compared to dapalide A (**1**) labeled red. All other units are labeled black.

Endobryopsis kahalalidefaciens, an intracellular bacterial symbiont of *Bryopsis* sp.¹⁹

The kahalalides are a distinct class of depsipeptides, containing both L-, D-, and modified amino acid building blocks, and fatty acid tails with varying substitutions at the N-terminus of the compounds. Successful synthetic efforts toward various analogues for structure–activity relationship (SAR) studies have been conducted using solid phase peptide synthesis (SPPS).^{20–24} The kahalalides have been extensively reviewed over the years^{25,26} and compounds are continuing to be reisolated and discovered from marine sources around the world.^{18,27} Nine of the 24 isolated kahalalides show significant biological activity, including anticancer, antifungal, antiviral, antileishmanial, and antimarial activity, in addition to immune suppressive and lymphatic activity.^{25,27,28} The most notably cytotoxic analogue is kahalalide F (KF), the endocyclic depsipeptide with the modified amino acid dehydrobutyryne (Dbu or Dhb), an ornithine residue (Orn), and at 5-methylhexanoic acid (5-MeHaa) fatty acid tail.¹⁷ KF showed significant activity against various human solid tumor cell lines and revealed diminished inhibitory activity against nontumor cell lines, indicating the possibility of an *in vivo* therapeutic window.^{17,29} This discovery led to KF entering clinical trials as a prime candidate for various indications,^{30–32} but the trials were eventually terminated due to lack of effects on tumor growth.^{25,33} A synthetic analogue of KF, elisidepsin (PM02734; Irvalec), also entered clinical trials as a potent cytotoxin against solid tumors.³⁴ Despite a successful phase 1 study to determine dose/dose-limiting toxicities, the phase II trial was terminated due to lack of efficacy, thus removing kahalalides or analogous from clinical trials entirely.^{35,36}

In our ongoing research efforts to aid in the discovery of new bioactive molecules from marine cyanobacteria, we report three new kahalalide F analogues, dapalides A–C (**1–3**), that have been isolated and characterized through traditional isolation chemistry, structural determination, and verification by NMR and LC/MS mass fragmentation from a marine filamentous cyanobacterial mat (Figure 1). Through assembling long-read metagenomic sequencing data from the sample, the biosynthetic gene cluster (BGC) responsible for the production of the dapalides was discovered in the genome of a *Dapis* sp., and the epimerization domains located within the

BGC were used to establish the absolute stereochemistry of the repeat amino acids. This discovery broadens the taxonomic origins of the kahalalide structural class as well as the chemical diversity between producers.

RESULTS AND DISCUSSION

The marine cyanobacterial sample VPG23-80 was collected from patch reefs inside Cocos Lagoon, Guam. The freeze-dried sample was extracted with 1:1 EtOAc:MeOH, then subjected to solvent–solvent partitioning, silica gel chromatography, and reversed-phase HPLC, yielding dapalide A (**1**) as the major compound along with two minor metabolites, dapalide B (**2**) and dapalide C (**3**). The structures were elucidated via a combination of NMR analysis using ¹H NMR, ¹H–¹H COSY, TOCSY, HSQC, HMBC, and ROESY (acquired in DMSO-*d*₆), and mass fragmentation analysis using LC-MS/MS.

Dapalide A (**1**) was isolated as a white powder with the molecular formula C₇₆H₁₃₁N₁₃O₂₄ as determined by HRESIMS data (*m/z* for [M + H]⁺ 1610.9490) coupled with NMR data. The degrees of unsaturation deduced from this formula were 18. Initial inspection of the ¹H NMR, HSQC, and HMBC data revealed the structure class was a peptide with indicative α -protons ($\delta_H \sim 3.8$ –4.8 ppm), amide carbonyls and exchangeable protons ($\delta_C \sim 168$ –173 ppm; $\delta_H \sim 6.75$ –8.14 ppm), decorated with a sugar unit ($\delta_{H/C}$ 4.95/99.04), and a fatty acid tail (δ_C 173.09; $\delta_{H/C}$ 2.19/34.90) (Table 1; acquired in DMSO-*d*₆). Briefly, ¹H–¹H COSY correlations paired with the TOCSY correlations established 12 spin systems of standard amino acids: four threonine (Thr) units, three isoleucine (Ile) units, two valine (Val) units, one leucine (Leu), one serine (Ser), and one proline (Pro) (Table 1, Figure 2a). Additionally, a separate spin system ascertained by the COSY/TOCSY correlations contained multiple exchangeable OH protons and a hemiacetal functional group (H-1'; $\delta_{H/C}$ 4.95/99.04), thus was determined to be the sugar group. However, two of the Thr units were missing a COSY correlation to an exchangeable OH, and the β -positions of the units were shifted downfield in the proton chemical shifts and upfield in the carbon for one and downfield for the other. Thus, Thr-1 was revealed to be the point of esterification due to the β -position having a downfield proton chemical shift of

Table 1. NMR Spectroscopic Data for Dapalide A (1) in $\text{DMSO}-d_6$ at 25 °C 1

Amino Acid	Position	δ_{H} (J in Hz) ^a	δ_{C} ^b	COSY	HMBC	TOCSY	ROESY
Val-1	1		169.2, C		2, 3, 5		
	2	3.85, dd (10.0, 9.0)	59.9, CH	6, 3	1, 3, 4, 5	6, 3, 4, 5	
	3	1.37, ov. ^c	29.3, CH	2, 4, 5	2	6, 2, 4, 5	
	4	0.64, d (6.5)	18.89, CH_3	3	2, 3, 5	6, 2, 3, 5	
	5	0.58, d (6.5)	18.2, CH_3	3	2, 3, 4	6, 2, 3, 4	
	6	6.75, d (9.0)	NH	2	2, 7	2, 3, 4, 5, 7	1, 5
Dbu	7		163.2, C		2, 6, 9, 10, 11		
	8		130.2, C		9, 10		
	9	6.46, dq (7.1, 1.1)	129.4, CH	11, 10	8, 10	11, 10	11
	10	1.67, d (7.1)	12.4, CH_3	9	9	11, 9, 8	11
	11	9.95, s	NH	2	8, 12	2, 3, 4, 5	11, 5
Ile-1	12		172.1, C		11, 13		
	13	3.91, dd (10.4, 5.8)	58.3 CH	14, 18	14, 15, 16	15, 16, 17, 18	18
	14	1.83, ov. ^c	32.8, CH	13, 15, 16	13, 15, 16	13, 15, 16, 17, 18	18
	15	1.54, ov. ^c	24.3, CH_2	14, 17	14, 16, 17	13, 14, 16, 17, 18	
		1.24, ov. ^c		14, 17	14, 16, 17	13, 14, 16, 17, 18	
	16	0.9, d (6.7)	14.7, CH_3	13	14, 15, 17	13, 14, 15, 17, 18	
	17	0.835, ov. ^c	9.3, CH_3	15	13, 14, 15	13, 14, 15, 16, 18	
	18	8.78, d (5.8)	NH	13	13, 19	13, 14, 15, 16, 17, 18	13, 14, 20, 21
	19		172.8, C		18, 20		
	20	4.57, dd (8.9, 4.5)	53.0, CH	21, 25	21, 22, 23, 26,	10, 21, 22, 24, 23	18
Ile-2	21	1.99, ov. ^c	39.4, CH	20, 22, 23	20	10, 20, 22, 24, 23	18
	22	1.25, ov. ^c	26.1, CH_2	21, 24	20, 24, 23	10, 20, 21, 22, 24, 23	
		1.01, ov. ^c		21, 24	20, 24, 23	10, 20, 21, 22, 24, 23	
	23	0.74, d (6.3)	13.7, CH_3	21	20, 21, 22	10, 20, 21, 22, 24	
	24	0.93, t (7.4)	11.8, CH_3	21	21, 22	10, 20, 21, 22, 23	
	25	7.48, d (8.9)	NH	20	20, 26	20, 21, 22, 24, 23	27, 32
	26		171.1, C		10, 24		
	27	4.38, ov. ^c	29.1, CH	28, 32	26, 28	28, 29, 30, 31, 32	25
	28	1.53, ov. ^c	42.4, CH_2	27, 29	27	27, 29, 30, 31, 32	25
		1.37, ov. ^c		27, 29	27, 30, 31	27, 29, 30, 31, 32	
Thr-1	29	1.28, ov. ^c	23.9, CH	28, 30, 31	26, 30, 31	27, 28, 30, 31, 32	
	30	0.825, d (6.2)	23.1, CH_3	29	28, 29	27, 29, 28, 31, 32	
	31	0.808, d (6.2)	20.9, CH_3	29	28, 29	27, 29, 28, 30, 32	
	32	8.89, d (10.0)	NH	27	27, 33	27, 28, 29, 30, 31	25, 34
	33		168.0, CH		32, 34		
	34	4.43, dd (10.2, 9.0)	56.0, CH	35, 37	33, 35, 36, 37, 38	35, 36, 37	32, 37, 44
	35	4.98, dd (6.4, 10.2)	69.4, CH	34, 36	34, 36, 88, 90	34, 36, 37	
	36	1.07, d (6.4)	16.5, CH_3	35	34, 35, 37	35, 34, 37	
	37	8.57, d, 9.0	NH	34	34, 38	34, 35, 36	34, 39, 44
Ile-3	38		169.3, C		37, 39		
	39	4.36, dd (8.5, 5.6)	55.6, CH	44	38, 40, 41,	40, 41, 42, 43, 44	37
	40	1.69, ov. ^c	37.1, CH	41	39, 41, 42, 43	39, 41, 42, 43, 44	
	41	1.32, ov. ^c	25.2, CH_2	40, 43	39, 40, 42, 43	39, 40, 42, 43, 44	
		1.05, ov. ^c		40, 43	39, 40, 42, 43	39, 40, 42, 43, 44	
	42	0.81, d (6.8)	11.4, CH_3	41	41	39, 40, 41, 42, 44	
	43	0.78, d (6.8)	14.0, CH_3	40	39, 40, 41	39, 40, 41, 42, 43	
	44	7.59, d (8.5)	NH	39	39, 45	39, 40, 41, 42, 43	34, 37, 46, 47, 49
Ser	45		170.9, C		44, 46		
	46	4.35, ov. ^c	54.6, C	47, 49	45, 47, 48	47, 48, 49	44
	47	3.53, ov. ^c	61.8, CH_2	46, 48	46, 50	47, 48, 49	44
		3.49, ov. ^c		46, 48	46, 50	47, 48, 49	44
	48	4.85, t (5.3)	OH	47	46, 47	46, 47, 49	
	49	7.83, d (8.0)	NH	46	46, 50	46, 47, 48	44, 28
	50		171.5, C		47, 49, 51		
Pro	51	4.44, ov. ^c	59.3, CH	52	52, 53	52, 53, 54	49
	52	2.01, ov. ^c	29.0, CH_2	53	50	51, 53, 54	
		1.81, ov. ^c		53	50	51, 53, 54	
	53	1.9, ov. ^c	24.0, CH_2	54, 52	52, 54	51, 52, 54	
		1.81, ov. ^c		54, 52	52, 54	51, 52, 54	
	54	3.7, w.p. ^d	46.8, CH_2	53	53	51, 52, 53	

Table 1. continued 1

Amino Acid	Position	δ_H (J in Hz) ^a	δ_C ^b	COSY	HMBC	TOCSY	ROESY	^c
Val-2		3.55, w.p. ^d		53	53	51, 52, 53		
	55		171.0, C		56			
	56	4.35, ov. ^c	55.3, CH	57, 60	34, 42, 45, 55			
	57	1.97, ov. ^c	29.9, CH	56	56, 58, 59	56, 58, 59, 60		
	58	0.89, d (6.8)	18.97, CH ₃	57	56, 57, 59	56, 57, 59		
	59	0.84, ov. ^c	18.0, CH	57	57, 58, 56	56, 57, 58		
Thr-2	60	7.80, d (8.0)	NH	56	56, 61	56, 57, 58, 59	62, 63, 71	
	61		169.3, C		60, 62			
	62	4.29, dd (8.2, 3.2)	57.6, CH	8, 29	61, 65, 66	61, 63, 65, 66	60	
	63	4.03, ov. ^c	66.5, CH	27, 62.5, 16	62, 64, 65	61, 62, 65, 66	60	
	64	1.02, d (6.2)	19.3, CH ₃	29	62, 63	62, 63, 65, 66		
	65	4.89, d (4.7)	OH	29	61, 62, 63	61, 62, 63, 66		
Thr-3	66	7.73, d (8.2)	NH	27	62, 67	61, 62, 63, 65	68, 71, 31	
	67		169.4, C		66, 68			
	68	4.37, dd (8.0, 4.3)	57.4, CH	69, 72	67, 70, 71, 72	69, 70, 71, 72	66	
	69	3.96, ov. ^c	66.2, CH	68, 70, 71		68, 69, 71, 72	66	
	70	1.02, d (6.2)	18.96, CH ₃	69	68, 69, 71	68, 69, 71, 72		
	71	5.09, d (4.8)	OH	69	68, 70	68, 69, 70, 72	60, 66	
Thr-4	72	8.14, d (8.0)	NH	68	68, 73	68, 69, 70, 71	66, 75, 77	
	73		169.1, C		72, 74			
	74	4.73, dd (8.5, 4.2)	55.7, CH	77, 75	73, 75, 76,	75, 76, 77	31	
	75	3.97, ov. ^c	75.6, CH	74, 76	1', 73, 76	74, 76, 77		
	76	1.11, d (6.5)	16.5, CH ₃	75	74, 75	74, 75, 77		
	77	8.07, d (8.5)	NH	74	74, 78	74, 75, 76	31, 1', 2"	
Glc	1'	4.95, d (3.8)	99.0, CH	2'	75, 2"	2', 2", 3', 3", 4", 6'	2'	
	2'	3.09, ddd (9.2, 9.2, 3.8)	72.0, CH	1', 2", 3'		1', 2", 3', 3", 4', 4", 5', 6"	1', 4'	
	2"	4.23, d (9.2)	OH	2'	1'	1', 2', 3", 4', 5', 6'	3'	
	3'	3.29, ddd (9.2, 9.2, 3.6)	73.5, CH	3", 4",		1', 2', 2", 3", 4", 4", 5', 6"	2", 4", 5'	
	3"	4.70, d (4.8)	OH	3'		1', 2', 2", 3", 4", 6",		
	4'	3.02, ddd (9.2, 9.2, 4.8)	69.9, CH	3', 5'	6'	1', 2', 2", 3", 3", 4", 5', 6"	2', 6', 6"	
	4"	4.81, d (4.8)	OH	4'	3', 4'	1', 2', 2", 3", 3", 6'	3'	
	5'	3.42, dd (9.2, 5.2)	31.7, CH	4', 6'	6"	1', 2', 2", 3", 3", 4', 4", 6"	3'	
	6'	3.59, dd (-9.0, 5.2)	60.5, CH ₂	6"		4', 4", 6"	4'	
	3.43, d (-9.0)			6"		4', 4", 6"	4', 76	
	6"	4.43, br s	OH	6'	6'	3', 3", 4", 5', 6'	4'	
	78		173.09, C		77, 79			
Oaa	79	2.19, m	34.9, CH ₂	80	78	80, 81		
	80	1.48, ov. ^c	25.0, CH ₂	79, 81		79, 81		
	81	1.25, ov. ^c	28.29, CH ₂		82	79, 80		
	82	1.25, ov. ^c	28.29, CH ₂		81, 83			
	83	1.23, ov. ^c	31, CH ₂		82, 84			
	84	1.24, ov. ^c	21.8, CH ₂		83, 85			
	85	0.85, ov. ^c	13.7, CH ₃	84	83, 84			

^a600 MHz (¹H). ^bDeduced from HSQC and HMBC. ^cNMR signal overlaps with other protons. ^dSignal is underneath the water peak. 3

~1 ppm (δ_H 4.98) and the upfield carbon chemical shift of 10.4 (δ_C 56.0). The other Thr unit had an HMBC correlation of the β -carbon ($\delta_{H/C}$ 3.97/75.6) to the hemiacetal methine ($\delta_{H/C}$ 4.95/99.04), suggesting the glycosylation unit was on Thr-4.

To selectively isolate the coupling constants of the sugar moiety to establish the relative configuration, a 1D selective TOCSY NMR experiment was conducted, irradiating the H-2' proton (δ_H 3.09). (Figure S6) Due to the large coupling constants of protons H-2'/H-3' (9.2 Hz), H-3'/H'4' (9.2 Hz), and H-4'/H-5' (9.2 Hz), the protons were assigned in the axial position (Figure S6). The relatively small coupling constant between H-1'/H-2' (3.8 Hz) revealed H-1' was in the equatorial position, thus the sugar was assigned as a α -glucopyranose (Figure 2b). Furthermore, the lack of ROE correlations between the H-1'/H-3' and the small coupling

constant between H-1'/H-2' confirms the equatorial orientation of the anomeric proton H-1' and indicates the presence of an α -sugar. Experimentally, anomeric chemical shifts of α -sugars have upfield carbon chemical shifts (δ_C 94–99) and a small coupling constant between 2 and 3.8 Hz.^{37–40} As mentioned above, the chemical shift values coupled with the coupling constants and ROESY correlations unequivocally assigned this unit as an α -glucopyranose (Figure 2b). The last spin system was revealed to be a dehydrated threonine unit (Dbu unit), as assigned via a singlet at the aliphatic β position ($\delta_{H/C}$ 6.46/129.43) corresponding to the amide proton signal at δ_H 9.91 and the HMBC correlation from the β proton ($\delta_{H/C}$ 6.46/129.43) to the nonprotonated sp² α -carbon signal (δ_C 130.22). The chemical shift of the olefinic methine (δ_C 129.43) and a ROE between NH and CH₃ protons in the Dbu unit

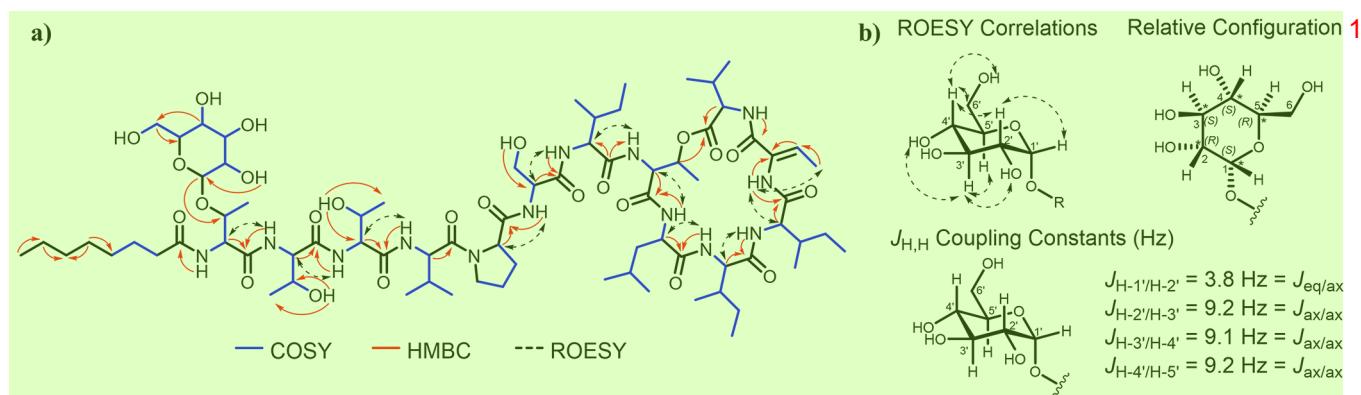


Figure 2. a) Key COSY (blue), HMBC (red), and ROESY (dashed) 2D NMR correlations for dapalide A (1) to assemble the planar structure and b) relative configuration of the α -glucopyranose including coupling constants deduced by 1D TOCSY.

indicated cis (Z) geometry of the double bond.^{41–43} Lastly, the presence of a fatty acid chain with eight carbon units (octanoic acid: Oaa) fit the molecular formula requirements and the degrees of unsaturation were fulfilled by an ester cyclization between the hydroxy group of Thr-1 and the carbonyl of Val-1 using HMBC correlations. The planar peptide sequence of dapalide A (1) was established using HMBC and ROESY correlations as Oaa-Thr-Thr-Thr-Val-Pro-Ser-Ile-Thr-Leu-Ile-Ile-Dbu-Val, with the underlined residues as the cyclization point.

To verify the structure of 1, MS/MS fragmentation analysis was conducted. Mass fragmentation provided the sequences of seven of the 13 amino acids and the sugar cleavage. The sequence for dapalide A (1) revealed the sequential loss of the sugar unit ($[-\text{Glc}+\text{H}]^+$ 1488.8 [calc 1488.9]), the GlcThr unit with the C8 (Oaa) fatty acid tail ($[\text{b}_2+\text{H}]^+$ 1221.8 [calc 1221.7]), the dipeptide Thr-GlcThr-Oaa ($[\text{b}_3+\text{H}]^+$ 1120.8 [calc 1120.7]), the tripeptide Thr-Thr-GlcThr-Oaa ($[\text{b}_4+\text{H}]^+$ 1019.7 [calc 1019.7]), the tetrapeptide Val-Thr-Thr-GlcThr-Oaa ($[\text{b}_5+\text{H}]^+$ 920.6 [calc 920.6]), and the rest of the linear heptapeptide Ile-Ser-Pro-Val-Thr-Thr-GlcThr-Oaa ($[\text{b}_8+\text{H}]^+$ 623.4 [calc 623.4]). The depsipeptide ring was not fragmented under the applied conditions (Figure 3).

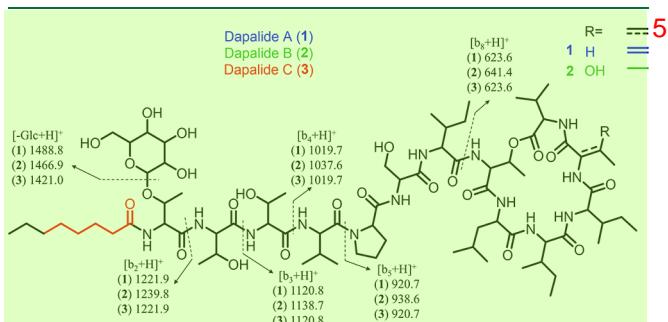


Figure 3. Structural verification of dapalides A–C (1–3) through MS/MS fragmentation, which revealed the structural differences in blue, green, and red.

The same extract yielded dapalide B (2) from a mixture of compounds in which the structure was solved using a combination of mass fragmentation and in-depth analysis of the unclean NMR spectra. The molecular formula of dapalide B (2) was determined to be $C_{76}H_{133}N_{13}O_{25}$ through HRESIMS data analysis ($m/z [M + H]^+$ 1627.9767) coupled with the NMR data (Table 2). The structure was fully

elucidated despite the presence of multiple compounds in the sample by using both the NMR and mass fragmentation patterns from the LC-MS. MS data revealed the addition of 18 amu, suggesting the formal addition of H_2O and presence of a hydroxy group. Further analysis of the 1H NMR data revealed the lack of the characteristic amide singlet proton at δ_H 9.91 and the doublet of quartet methine proton at δ_H 6.46, indicating the presence of the threonine group in the place of the Dbu unit. Mass fragmentation further indicated the additional groups are on the depsipeptide ring as the mass fragments reflected the same fragmentation as 1 (Figure 3). The sequential loss began with the sugar unit ($[-\text{Glc}+\text{H}]^+$ 1466.9 [calc 1466.9]), then the GlcThr with the Oaa fatty acid ($[\text{b}_2+\text{H}]^+$ 1239.8 [calc 1221.7] and the $[\text{a}_2 + \text{H}]^+$ 1266.9 [calc 1266.7]), then the tetrapeptide Val-The-The-GlcThr-Oaa ($[\text{b}_5+\text{H}]^+$ 938.6 [calc 938.6]), and last the linear heptapeptide Ile-Ser-Pro-Val-Thr-Thr-GlcThr-Oaa ($[\text{b}_8+\text{H}]^+$ 641.4 [calc 641.4]). The depsipeptide ring was not fragmented, but the NMR clearly revealed the Thr unit in place of the Dbu unit.

Dapalide C (3) was isolated as a mixture with dapalide A (1) (7:3 ratio based on HRESIMS) that was a colorless solid with the molecular formula of $C_{74}H_{127}N_{13}O_{24}$ as determined by HRESIMS data ($m/z [M + H]^+$ 1582.9185), coupled with NMR data to determine the structure. The structure was fully elucidated despite the presence of dapalide A (1) by using both the NMR and mass fragmentation patterns from the LC-MS. The 1H NMR spectrum was almost identical to that for 1, suggesting that the peptide and glycosylated unit were identical. The difference of 28 amu compared with 1 strongly suggested that the C8 chain (Oaa) in 1 is replaced by a C6 chain (hexanoic acid: Haa). The mass fragmentation sequence verified this tail reduction via the $[-\text{Glc}+\text{H}]^+$, indicating the loss of 28 amu. The fragmentation sequence for 3 revealed the loss of the glycosylated group from the Thr unit first ($[-\text{Glc}+\text{H}]^+$ 1421.0 [calc 1420.9]), then the GlcThr with the Haa tail ($[\text{b}_2+\text{H}]^+$ 1221.9 [calc 1221.8]). The subsequent fragmentation aligns with the fragmentation of 1 (Figure 3), validating the structural assignment for 3.

The analysis of the absolute configuration of dapalide A (1) was conducted first chemically by comparing amino acid standards to the acid hydrolyzate. The amino acids were liberated using acid hydrolysis, then subjected to enantioselective LC-MS analysis and comparison with amino acid standards. L-Ser, D-Pro, and D-Leu were unambiguously assigned as the only singly represented amino acids, and a

Table 2. NMR Spectroscopic Data for Dapalides B (2) and C (3) in $\text{DMSO}-d_6$ at 25 °C 1

Dapalide B (2)				Dapalide C (3)			
Amino Acid	Position	δ_{H} (J in Hz) ^a	$\delta_{\text{C}}^{\text{b}}$	Amino Acid	Position	δ_{H} (J in Hz) ^a	$\delta_{\text{C}}^{\text{b}}$
Val-1	1			Val-1	1		169.2, C
	2	3.85, t (10.0)	66.8, CH		2	3.85, dd (10.0, 9.0)	59.9, CH
	3	1.37, ov. ^c	27.7, CH		3	1.37, ov. ^c	29.3, CH
	4	0.64, d (6.5)	16.7, CH_3		4	0.64, d (6.5)	18.89, CH_3
	5	0.58, d (6.5)	18.2, CH_3		5	0.58, d (6.5)	18.2, CH_3
	6	6.75, d (9.0)	NH		6	6.75, d (9.0)	NH
Thr-1				Dbu	7		163.2, C
	8	4.24, d (7.8)	57.8, CH		8		130.2, C
	9	3.90, ov. ^c	66.6, CH		9	6.46, dq (7.1, 1.1)	129.4, CH
	10	1.07, ov. ^c	19.7, CH_3		10	1.67, d (7.1)	12.4, CH_3
	11	7.88, d (7.8)	NH		11	9.95, s	NH
Ile-1	12			Ile-1	12		172.1, C
	13	4.47 dd (9.0, 4.5)	53.9, CH		13	3.91, dd (10.4, 5.8)	58.3 CH
	14	1.83, ov. ^c	36.7, CH		14	1.83, ov. ^c	32.8, CH
	15	1.49, ov. ^c	24.8, CH_2		15	1.54, ov. ^c	24.3, CH_2
		1.35, ov. ^c				1.24, ov. ^c	
	16	0.92, ov. ^c	14.9, CH_3		16	0.9, d (6.7)	14.7, CH_3
	17	0.81, ov. ^c	10.0, CH_3		17	0.835, ov. ^c	9.3, CH_3
	18	7.63, d (9.0)	NH		18	8.78, d (5.8)	NH
Ile-2	19		172.8, C	Ile-2	19		172.8, C
	20	4.57, dd (8.9, 4.5)	53.0, CH		20	4.57, dd (8.9, 4.5)	53.0, CH
	21	1.99, ov. ^c	39.4, CH		21	1.99, ov. ^c	39.4, CH
	22	1.25, ov. ^c	26.1, CH_2		22	1.25, ov. ^c	26.1, CH_2
		1.01, ov. ^c				1.01, ov. ^c	
	23	0.74, d (6.3)	13.7, CH_3		23	0.74, d (6.3)	13.7, CH_3
	24	0.93, t (7.4)	11.8, CH_3		24	0.93, t (7.4)	11.8, CH_3
	25	7.48, d (8.9)	NH		25	7.48, d (8.9)	NH
Leu	26		171.1, C	Leu	26		171.1, C
	27	4.38, ov. ^c	29.1, CH		27	4.38, ov. ^c	29.1, CH
	28	1.53, ov. ^c	42.4, CH_2		28	1.53, ov. ^c	42.4, CH_2
		1.37, ov. ^c				1.37, ov. ^c	
	29	1.28, ov. ^c	23.9, CH		29	1.28, ov. ^c	23.9, CH
	30	0.825, d (6.2)	23.1, CH_3		30	0.825, d (6.2)	23.1, CH_3
	31	0.808, d (6.2)	20.9, CH_3		31	0.808, d (6.2)	20.9, CH_3
	32	8.89, d (10.0)	NH		32	8.89, d (10.0)	NH
Thr-2	33		168.0, CH	Thr-1	33		168.0, CH
	34	4.43, dd (10.2, 9.0)	56.0, CH		34	4.43, dd (10.2, 9.0)	56.0, CH
	35	4.98, dd (6.4, 10.2)	69.4, CH		35	4.98, dd (6.4, 10.2)	69.4, CH
	36	1.07, d (6.4)	16.5, CH_3		36	1.07, d (6.4)	16.5, CH_3
	37	8.57, d, 9.0	NH		37	8.57, d, 9.0	NH
Ile-3	38		169.3, C	Ile-3	38		169.3, C
	39	4.36, dd (8.5, 5.6)	55.6, CH		39	4.36, dd (8.5, 5.6)	55.6, CH
	40	1.69, ov. ^c	37.1, CH		40	1.69, ov. ^c	37.1, CH
	41	1.32, ov. ^c	25.2, CH_2		41	1.32, ov. ^c	25.2, CH_2
		1.05, ov. ^c				1.05, ov. ^c	
	42	0.81, d (6.8)	11.4, CH_3		42	0.81, d (6.8)	11.4, CH_3
	43	0.78, d (6.8)	14.0, CH_3		43	0.78, d (6.8)	14.0, CH_3
	44	7.59, d (8.5)	NH		44	7.59, d (8.5)	NH
Ser	45		170.9, C	Ser	45		170.9, C
	46	4.35, ov. ^c	54.6, C		46	4.35, ov. ^c	54.6, C
	47	3.53, ov. ^c	61.8, CH_2		47	3.53, ov. ^c	61.8, CH_2
		3.49, ov. ^c				3.49, ov. ^c	
	48	4.85, t (5.3)	OH		48	4.85, t (5.3)	OH
Pro	49	7.83, d (8.0)	NH	Pro	49	7.83, d (8.0)	NH
	50		171.5, C		50		171.5, C
	51	4.4, ov. ^c	59.3, CH		51	4.4, ov. ^c	59.3, CH
	52	2.01, ov. ^c	29.0, CH_2		52	2.01, ov. ^c	29.0, CH_2
		1.81, ov. ^c				1.81, ov. ^c	
	53	1.9, ov. ^c	24.0, CH_2		53	1.9, ov. ^c	24.0, CH_2

Table 2. continued 1

Dapalide B (2)				Dapalide C (3)			
Amino Acid	Position	δ_H (J in Hz) ^a	δ_C ^b	Amino Acid	Position	δ_H (J in Hz) ^a	δ_C ^b
Val-2	54	1.81, ov. ^c 3.7, w.p. ^d 3.55, w.p. ^d	46.8, CH ₂	Val-2	54	1.81, ov. ^c 3.7, w.p. ^d 3.55, w.p. ^d	46.8, CH ₂
	55		171.0, C		55		171.0, C
	56	4.35, ov. ^c	55.3, CH		56	4.35, ov. ^c	55.3, CH
	57	1.97, ov. ^c	29.9, CH		57	1.97, ov. ^c	29.9, CH
	58	0.89, d (6.8)	18.97, CH ₃		58	0.89, d (6.8)	18.97, CH ₃
	59	0.84, ov. ^c	18.0, CH		59	0.84, ov. ^c	18.0, CH
	60	7.80, d (8.0)	NH		60	7.80, d (8.0)	NH
	61		169.3, C		61		169.3, C
	62	4.29, dd (8.2, 3.2)	57.6, CH		62	4.29, dd (8.2, 3.2)	57.6, CH
	63	4.03, ov. ^c	66.5, CH		63	4.03, ov. ^c	66.5, CH
Thr-3	64	1.02, d (6.2)	19.3, CH ₃	Thr-2	64	1.02, d (6.2)	19.3, CH ₃
	65	4.89, d (4.7)	OH		65	4.89, d (4.7)	OH
	66	7.73, d (8.2)	NH		66	7.73, d (8.2)	NH
	67		169.4, C		67		169.4, C
	68	4.37, dd (8.0, 4.3)	57.4, CH		68	4.37, dd (8.0, 4.3)	57.4, CH
	69	3.96, ov. ^c	66.2, CH		69	3.96, ov. ^c	66.2, CH
Thr-4	70	1.02, d (6.2)	18.96, CH ₃	Thr-3	70	1.02, d (6.2)	18.96, CH ₃
	71	5.09, d (4.8)	OH		71	5.09, d (4.8)	OH
	72	8.14, d (8.0)	NH		72	8.14, d (8.0)	NH
	73		169.1, C		73		169.1, C
	74	4.73, dd (8.5, 4.2)	55.7, CH		74	4.73, dd (8.5, 4.2)	55.7, CH
Thr-5	75	3.97, ov. ^c	75.6, CH	Thr-4	75	3.97, ov. ^c	75.6, CH
	76	1.11, d (6.5)	16.5, CH ₃		76	1.11, d (6.5)	16.5, CH ₃
	77	8.07, d (8.5)	NH		77	8.07, d (8.5)	NH
	78		169.1, C		78		169.1, C
	79	4.95, d (3.8)	99.0, CH		79	4.95, d (3.8)	99.0, CH
Glc	2'	3.09, ddd (9.2, 9.2, 3.8)	72.0, CH	Glc	2'	3.09, ddd (9.2, 9.2, 3.8)	72.0, CH
	2''	4.23, d (9.2)	OH		2''	4.23, d (9.2)	OH
	3'	3.29, ddd (9.2, 9.2, 3.6)	73.5, CH		3'	3.29, ddd (9.2, 9.2, 3.6)	73.5, CH
	3''	4.70, d (4.8)	OH		3''	4.70, d (4.8)	OH
	4'	3.02, ddd (9.2, 9.2, 4.8)	69.9, CH		4'	3.02, ddd (9.2, 9.2, 4.8)	69.9, CH
	4''	4.81, d (4.8)	OH		4''	4.81, d (4.8)	OH
	5'	3.42, dd (9.2, 5.2)	31.7, CH		5'	3.42, dd (9.2, 5.2)	31.7, CH
	6'	3.59, dd (-9.0, 5.2)	60.5, CH ₂		6'	3.59, dd (-9.0, 5.2)	60.5, CH ₂
	6''	3.43, d (-9.0)			6''	3.43, d (-9.0)	
	6'''	4.43, br s	OH		6'''	4.43, br s	OH
Oaa	78			Haa	78		
	79	2.19, m			79	2.19, m	35.4, CH ₂
	80	1.48, ov. ^c			80	1.48, ov. ^c	24.6, CH ₂
	81	1.25, ov. ^c			81	1.28, ov. ^c	28.7, CH ₂
	82	1.25, ov. ^c			82	1.24, ov. ^c	21.8, CH ₂
	83	1.23, ov. ^c			83	0.85, d (6.2)	13.7, CH ₃
	84	1.24, ov. ^c			84		
	85	0.85, ov. ^c			85		

^a600 MHz (¹H). ^bDeduced from HSQC and HMBC. ^cNMR signal overlaps with other protons. ^dSignal is underneath the water peak. 3

mixture of L/D-Val with a ratio of 2:1 was observed (Figure 4a). As for Thr and Ile units, the assignments were verified as either L-Thr or D-*allo*-Thr and L-Ile or D-*allo*-Ile, respectively (Figure 4b). The location of epimerization domains of the BGC (see below) ultimately revealed the modules that transform the Thr and Ile units to the D-configuration, and, more specifically, the D-*allo*-configuration. This powerful combination method successfully established the absolute configuration for dapalide A (1) without further sacrificing precious material for partial hydrolysis studies. Given the trace quantities (<100 µg or less) of dapalides B (2) and C (3), chemical analyses of the absolute configurations were not

conducted. However, all the amino acid units for dapalide C (5) are the same, supported by NMR comparison, and the BGC indicated that the Thr configuration at the Dbu position for dapalide B (2) was L because of the lack of an epimerization domain (see below). Lastly, absolute configuration of the sugar moiety could be established based on ROE correlations between D-*allo*-Thr and the attached α-glucopyranose. Specifically, the two ROE correlations revealed the sugar is in the D configuration based on the presence of the ROE correlations from the anomeric proton to the Thr β-proton and the 6' methylene protons to the Thr γ methyl. With the α-configuration previously determined using coupling

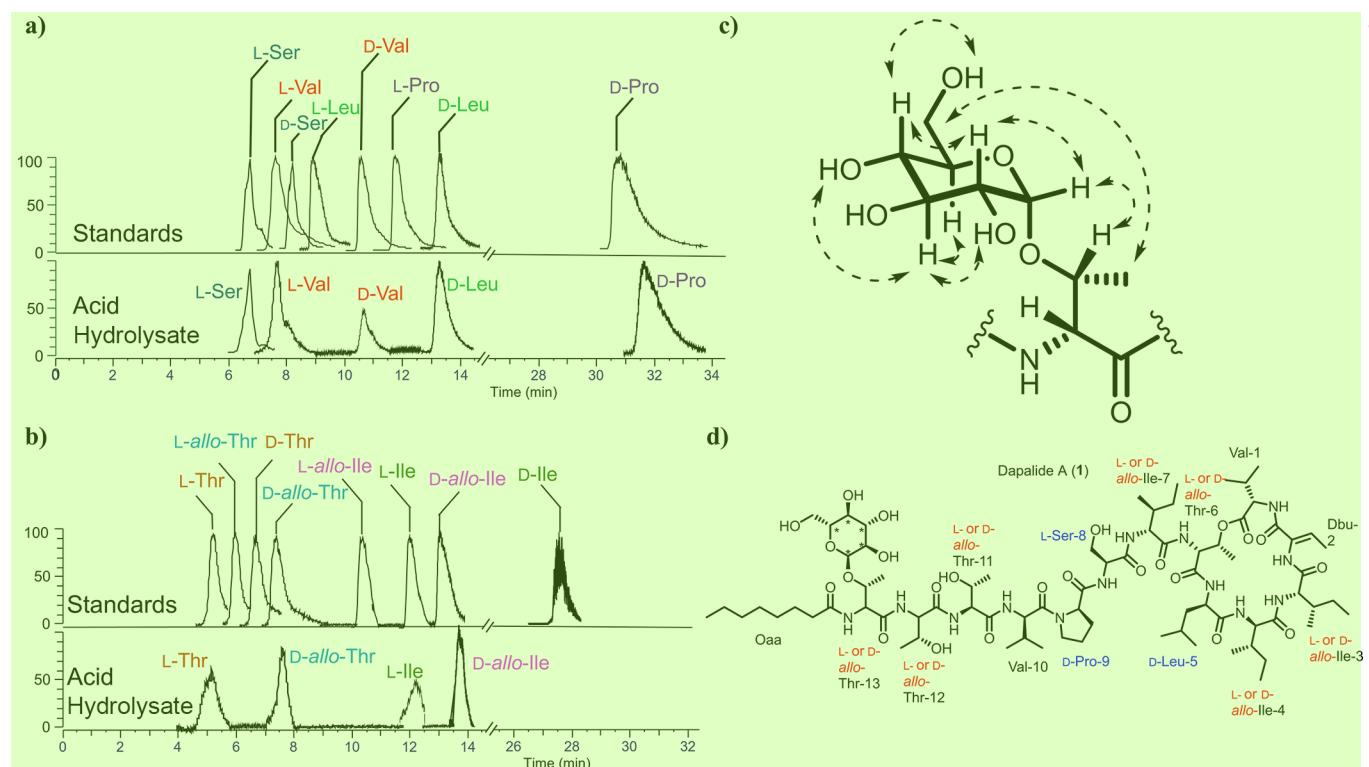


Figure 4. Chemical analysis of the absolute configuration of dapalide A (1). Chromatogram of the acid hydrolysate of dapalide A (1) compared to a) Ser, Val, Leu, and Pro, and b) Thr and Ile with each amino acid color coded. c) The absolute configuration of the sugar unit was established using ROESY correlations to the established D-allo-Thr. Two key ROE correlations included the anomeric proton (δ_H 4.95) to the one of the diastereotopic H-6' proton signal (δ_H 3.43) to the methyl of the Thr unit (δ_H 1.11). d) Structure of the dapalide A (1) configuration established by chemical methods unambiguously (blue) and partially (red).

Table 3. Amino Acid Placement of Dapalides A–C (1–3) Compared to Kahalalide F 3

	FA	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13
Kahalalide F	SMeHaa	V _D	T _L	V _L	V _D	P _D	O _L	I _D	T _D	I _D	V _D	F _L	DB _Z	V _L
Dapalide A	Oaa	T _D	T	T	V	P _D	S _L	I	T	L	I	I	T	V
Dapalide B	Oaa	T _D	T	T	V	P _D	S _L	I	T	L	I	I	DB _Z	V
Dapalide C	Haa	T _D	T	T	V	P _D	S _L	I	T	L	I	I	DB _Z	V

constants and the lack of a 1,3 diaxial interaction, the sugar unit was fully characterized as α -D-glucopyranose (Figure 4c). Thus, the absolute configuration was established using chemical methods for the leucine, proline, serine, and sugar units, and the partial absolute configuration was established for the isoleucine and threonine units to consist of L- or D-allo-amino groups (Figure 4d). The two valine units were revealed as a mixture of both L- and D-amino acids.

Dapalides A-C (1-3) share several structural similarities to kahalalide F. Both are lipotridecapeptides, cyclized through an ester bond formed between the carboxylic acid of the last residue (Val-1) and a hydroxy group of Thr-1; seven out of the 13 amino acids are identical (positions 2, 4, 5, 7, 8, 12, and 13; Figure 1, Table 3), while other two belong to the same amino acid class (positions 9 and 10; Figure 1, Table 3). Due to their structural similarity to kahalalide F, we hypothesized that dapalides A-C are likely produced by a nonribosomal peptide synthetase (NRPS) system similar to the *Ca. E. kahalalidifaciens* NRPS-8 pathway.¹⁹

To identify the BGC for dapalides A-C (1-3), we subjected VPG23-80 metagenomic DNA to long-read sequencing using the Pacific Biosciences platform (3,161,748 CCS reads, average read length 5,947 bps). After metagenomic assembly and

binning, we successfully recovered a ~ 11.2 Mbps metagenome-assembled genome (MAG, 93.4% completeness, 4.9% contamination), that encodes a 53.1-kbp NRPS pathway composed of 13 modules (Figure 5a,b). Satisfyingly, the predicted amino acid specificity for the adenylation domains of this pathway was in perfect agreement with the structure of dapalide A (1) (prediction was performed using NRPSpredictor2 in antiSMASH 5,⁴⁴ Figure 5c). As discussed above, the absolute configuration of the amino acids selected by modules 5, 6, and 9 were experimentally verified as D-Pro, L-Ser, and D-Leu in accordance with the presence/absence of an epimerization domain within the respective modules, further supporting that this BGC produces dapalide A (1) (Figure 5d). The 16S rRNA gene and 16S-23S internal transcribed spacer (ITS) region of this MAG share 100% identity to partial *Dapis pnigousa* NAB11-15 16S rRNA gene (MF061807) and 97.42% identity to *Dapis pnigousa* NAC11-53 ITS region (MF167708), respectively, indicating that this cyanobacterium belongs to the genus *Dapis*⁴⁵ (Figure 6). Additionally, the dapalides' producer MAG (VPG23-80 MAG-1) encodes 20 BGCs (identified by antiSMASH): 4 NRPSs, 2 NRPS-PKS, 1 PKS type-1, 1 PKS type-3, 8 RiPPPs, 3 terpenes, and 1 homoserinelactone. The only products isolated produced by

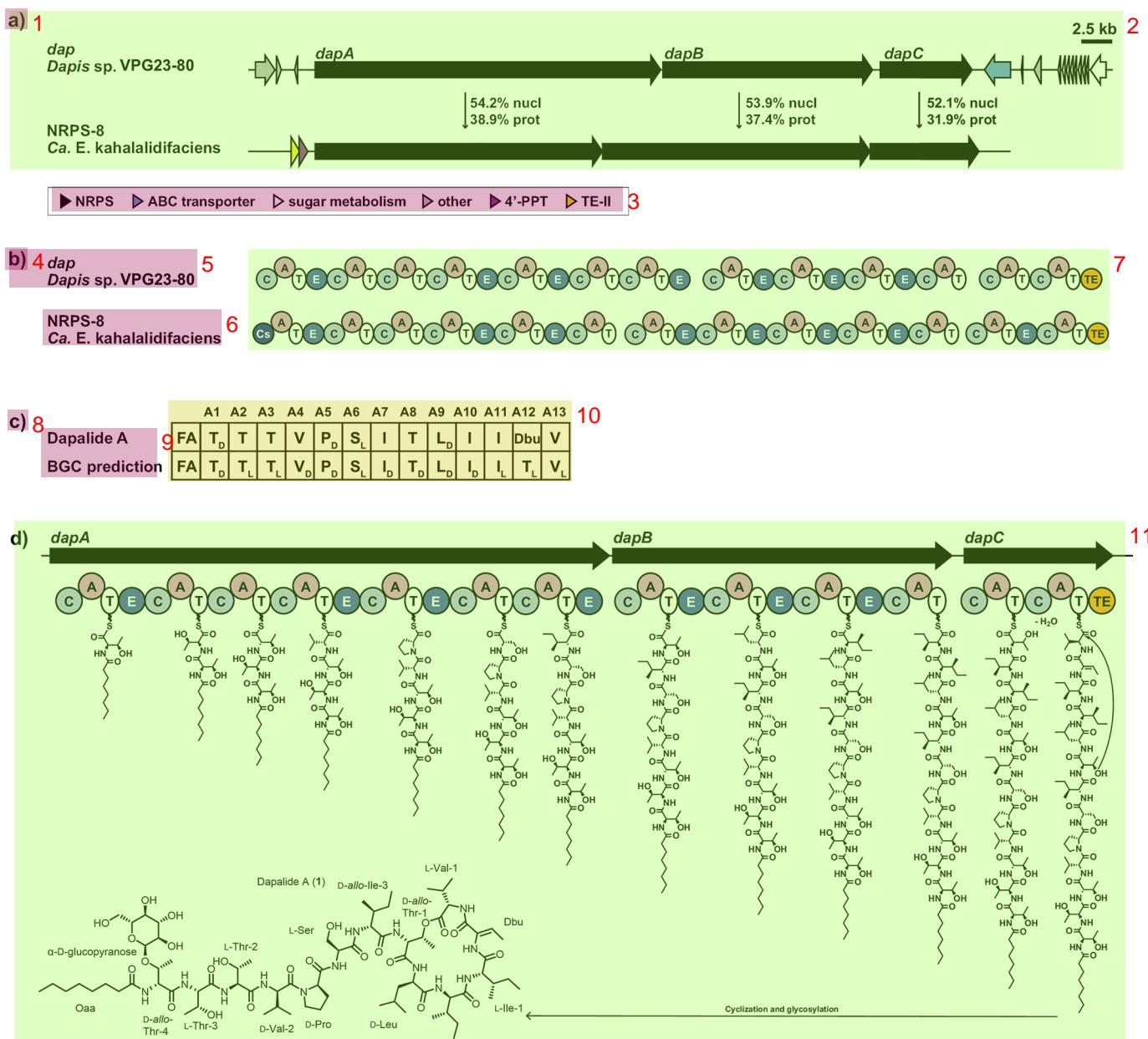


Figure 5. Genetic and domain architecture of the dapalide biosynthetic gene cluster. a) Comparison of the gene architectures between the dapalide BGC (*dap*) with the reported NRPS for kahalalide F (NRPS-8). Pairwise nucleotide and protein sequence identities between the NRPS genes of the two BGCs are shown. NRPS: nonribosomal peptide synthetase; 4'-PPT: 4'-phosphopantetheinyl transferase; TE-II: thioesterase II. b) Comparison of the NRPS domain architectures between the two BGCs shown in a. C: condensation domain; Cs: starter condensation domain; A: adenylation domain; T: thiolation domain; E: epimerase domain; TE: thioesterase domain. (c) Perfect agreement between the amino acid sequence determined experimentally for dapalide A (top) and the one predicted bioinformatically from the *dap* BGC, based on adenylation domain substrate specificity and the presence/absence of epimerization domains in each module. (d) Proposed dapalide A biosynthetic pathway.

VPG23-80 MAG-1 cyanobacterium were the dapalides A–C (1–3).

Overall, the *Ca. E. kahalalidifaciens* NRPS-8 and the *dap* BGC share 53.7% identity at the nucleotide level and 37.2% at the protein level. Interestingly, several differences can be pointed out between the two BGCs: (i) *dap* lacks the thioesterase type-II and 4'-phosphopantetheinyl transferase genes found in the kahalalide F BGC; (ii) the predicted amino acid stereochemistry for *dap* is identical to that of kahalalide F, except for module 12 that lacks an epimerization domain, and should select for L-Thr (establishing the absolute configuration for 2);

and (iii) downstream of the NRPS genes, the *dap* BGC harbor an ABC transporter gene, together with a gene involved in sugar metabolism.

Dapalide A (1), as the more abundant metabolite isolated, was screened for antifungal and antiproliferative activity. The compound had no activity against yeast *Saccharomyces cerevisiae* at the highest concentration tested (200 µg/mL) but showed weak cytotoxic activity against HCT 116 colorectal and Hep G2 liver cancer cells (Figure 7). To note, due to the limited amount of material, the highest concentration tested for the HepG2 cells was 10 µM. However, 1 exhibited potent growth inhibition at 100 µM in the HCT 116 cells and

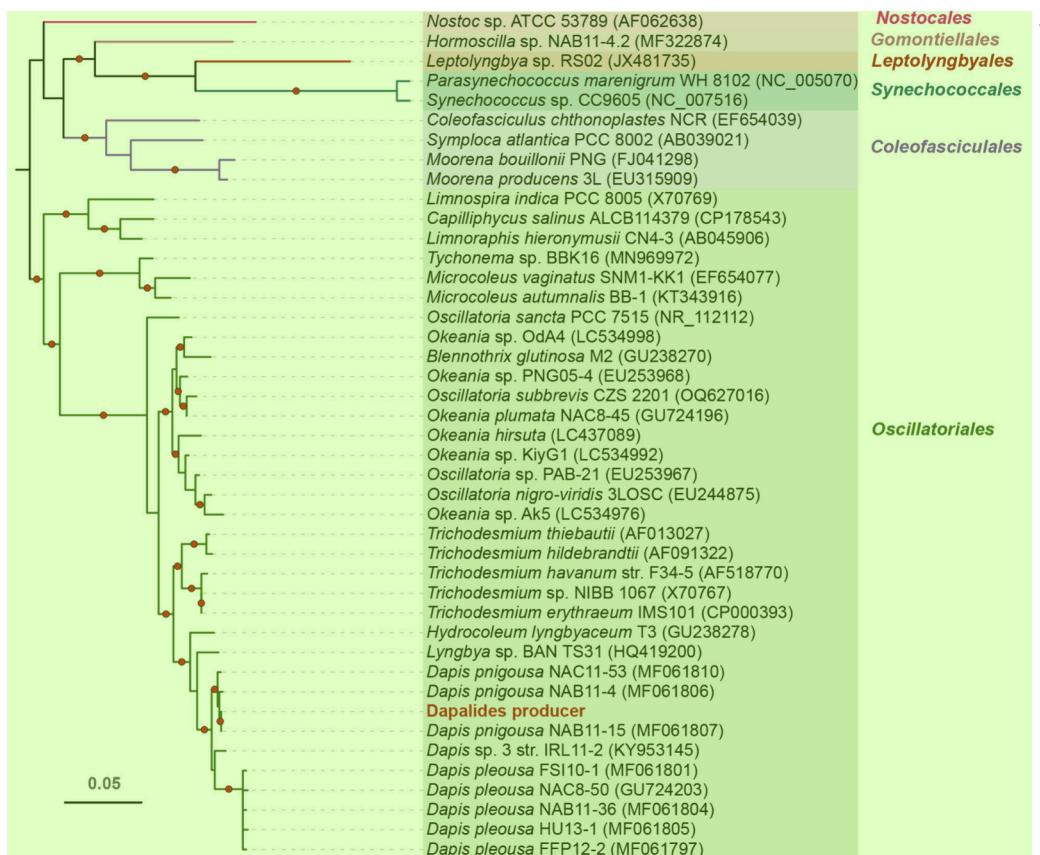


Figure 6. Phylogenetic tree of the dapalide producing *Dapis* sp. Phylogenetic tree was constructed based on the 16S rRNA gene sequence in comparison to selected reference cyanobacteria, using RAxML 8.2.11 and 500 bootstraps. Red dots on tree branches indicate a bootstrap value of >60%, and branch colors indicate the cyanobacterial taxonomic order.

moderate growth inhibition at 10 μ M in Hep G2 cells at 48 h (Figure 7). Kahalalide F was tested side by side for bioactivity comparison against both cancer cell lines as kahalalide F has μ M potency against both cell lines in both cell viability and growth inhibition assays.²⁸ Kahalalide F has 100-fold more potent antiproliferative activity against the HCT 116 and Hep G2 cells.

Previous extensive SAR evaluation of KF showed that the positions of the D- amino acids, the rigidity of the structural backbone due to the proline residue, and the aromatic phenylalanine and positively charged ornithine residues are the key features responsible for bioactivity of this compound class. Additionally, the key amino residues of the Phe on the depsipeptide ring and the positively charged ornithine that were removed during SAR studies depleted activity exponentially.²¹ As the dapalides possess only two of the four (replacing Phe with Ile and Orn with Ser), a lack of potent bioactivity in the cancer cell lines is to be expected.²¹ However, it is possible that the sugar moiety is used as a protecting group for unknown activity, as seen in other molecules that are prodrugs and become active once the glycosylated unit is cleaved. Such prodrugs include daunorubicin, cyclophosphamide, camptothecin, etc.⁴⁶

Cyanobacteria are a diverse group of ancient organisms that have produced several bioactive metabolites. The importance of this work lies in the discovery of a new source of kahalalide-like molecules - namely free-living marine cyanobacteria - and the structural differences between kahalalide F and the cyanobacterial metabolites.

During the development of this manuscript, the Gerwick laboratory released a preprint outlining the discovery of other kahalalide F-like analogues from a different genus of cyanobacterium from Panama, *Limnophysis* sp., through AI-assisted NMR analysis, supporting the widespread production of this class of molecules by diverse marine cyanobacteria.⁴⁷

Another dimension to the importance of this discovery lies within the relatively new genus producing new types of compounds not characterized in the literature. *Dapis* spp. are known to produce only six compounds, including two fatty acids, lyngbic acid and malyngolide,⁴⁸ carriebowlolin,^{45,48} and the linear modified peptides iheyamide A-C.^{49,50} Previously referred to as *Lyngbya*, this genus was only described in 2018,⁴⁵ thus the probability of discovering even more diverse metabolites or assignment of previously identified structures to *Dapis* is highly likely. We also contributed to the widely investigated SAR of kahalalide F analogues by introducing new glycosylated kahalalide congeners from a cyanobacterial source.

Additionally, a common dilemma that isolation chemists face is the limited amount of material that can be isolated from small collections of marine organisms, oftentimes precluding unambiguous structure determination, especially of the absolute configuration. In the case of complex peptides such as kahalalides or dapalides that are comprised of L- and D-amino acids, classic methods would have required partial hydrolysis and significant amounts of material that would have been irreversibly consumed.^{47,51} Bioinformatic analysis of the BGC can substantially aid in structure determination and

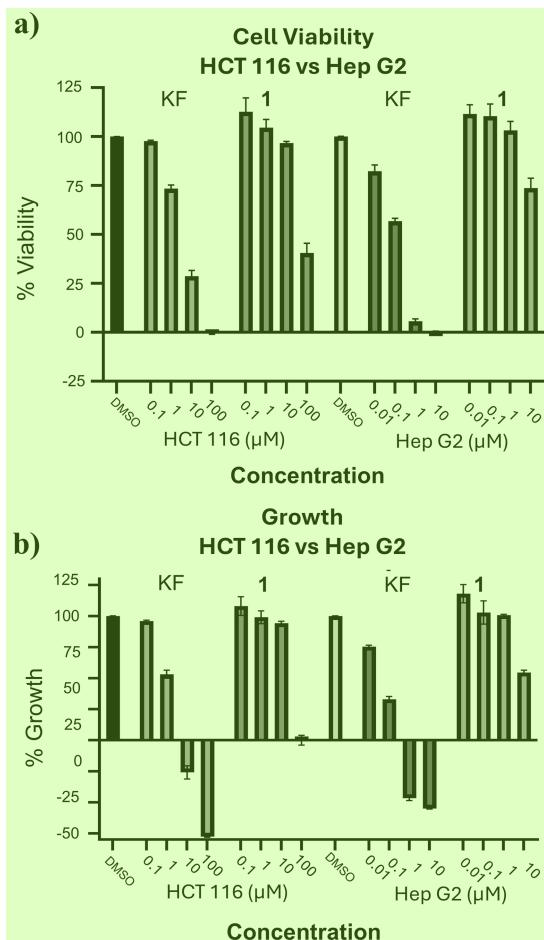


Figure 7. Antiproliferative activity of dapalide A (1) in HCT 116 and Hep G2 cells through a) cell viability and b) growth measurements.

configurational assignment for NRPS systems,^{52,53} exemplified by our work, without the need for extensive chemical degradation, saving precious material for the exploration of the biological potential of new compounds. Here we have characterized three new kahalalide F-like analogues, isolated in quantities of 0.1 to 1.0 mg, using this combinatorial approach to fully elucidate the structures through traditional methods of meticulous NMR analysis, without AI assistance, and completely characterize the absolute configuration. AI-assisted HSQC analysis has also proven useful for kahalalides as demonstrated for kahalalide Z5.⁴⁷

EXPERIMENTAL SECTION

General Experimental Procedures. Optical rotation was measured on a PerkinElmer 341 polarimeter. NMR data were collected on a Bruker Avance Neo 600 MHz spectrometer and high resolution Bruker Avance Neo 600 MHz 1.7 mm cryoprobe spectrometer, operating at 600 MHz for ¹H using residual solvent signal (δ_H 2.50; DMSO- d_6) as internal standard. HRESIMS data were obtained using a Q Exactive Focus with electrospray ionization (ESI). The LC-MS data were acquired in the negative mode using the TSQ Altis plus triple quadrupole mass spectrometer.

Cell Culture. HCT 116 human colorectal carcinoma cells and Hep G2 human hepatocellular carcinoma cells were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA) and cultured in Dulbecco's Modified Eagle Medium (DMEM) (DMEM, Invitrogen, Carlsbad, CA, USA) and Eagle's Minimum Essential Medium (EMEM, Invitrogen, Carlsbad, CA, USA), respectively, supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich, St.

Louis, MO, USA) and 1% antibiotic-antimycotic (penicillin, streptomycin, and Gibco Amphotericin B, Invitrogen) and maintained under a humidified environment with 5% CO₂ at 37 °C.

Biological Material. The marine cyanobacterium VPG23–80 was collected from patch reefs in Cocos Lagoon, Guam on April 4, 2023, as an orange-brown filamentous mat.

Extraction and Purification. The cyanobacterium (VPG23–80) was freeze-dried and extracted at the Smithsonian Marine Station using a nonpolar extraction with 1:1 EtOAc:MeOH (135.2 mg) and then a polar extraction with 30% aq. EtOH (88.44 mg). The nonpolar extract was subjected to solvent partitioning using EtOAc:H₂O, then the H₂O layer was partitioned again using BuOH:H₂O. The EtOAc partition (17.9 mg) was then subjected to silica gel chromatography using a Hexane:EtOAc MeOH gradient, with the fraction eluting with EtOAc:MeOH (1:1) (1.6 mg) being subjected to a C₁₈ Hypersep SPE cartridge using 1:1 H₂O:MeOH. The resulting fraction was then purified using semipreparative reverse phase HPLC (Synergy 4 μM Hydro-RP 80 Å, 250 × 10 mm, 3.0 mL/min; UV detection at 219 and 254 nm) using a binary gradient of H₂O:MeCN (20–100% MeCN for 20 min and then a 100% hold of MeCN for 5 min) to elute 1 at t_R 13.5 min (1.0 mg). The BuOH partition (5.1 mg) was subjected to C18 chromatography (100% MeOH) to clean up the sample then directly loaded onto the HPLC. The partition was purified using analytical reversed-phase HPLC (Synergy 4 μM Hydro-RP 80 Å, 250 × 4.6 mm, 1.5 mL/min; UV detection at 219 and 254 nm) using a binary gradient of H₂O: MeCN (20–100% MeCN for 22 min and then a 100% hold of MeCN for 10 min) to elute impure compound 2 at t_R 8.7 min (0.1 mg) and 3 (containing 1) at t_R 9.1 min (0.1 mg).

Dopalide A (1). White, amorphous solid; $[\alpha]^{20}_D$ −36 (c 0.002, CHCl₃); ¹H NMR, COSY, HSQC, HMBC, TOCSY, ROESY data (see Table 1); HRESIMS *m/z* 1610.9490 [M + H]⁺ (calc for C₇₆H₁₃₂N₁₃O₂₄).

Dopalide B (2). White, amorphous solid (30% mixture containing 2); ¹H NMR and HSQC NMR data (see Table 2); HRESIMS *m/z* 1627.9767 [M + H]⁺ (calc for C₇₆H₁₃₄N₁₃O₂₅).

Dopalide C (3). Colorless, amorphous solid (7:3 mixture of 3 with 1); $[\alpha]^{20}_D$ −29 (c 0.001, CHCl₃); ¹H NMR and HSQC NMR data (see Table 2); HRESIMS *m/z* 1582.9185 [M + H]⁺ (calc for C₇₄H₁₂₈N₁₃O₂₄).

Enantioselective LC-MS Analysis of Amino Acids. Dopalide A (1) (100 μg) was acid-hydrolyzed overnight (200 μL of 6 N HCl, 90 °C, 14 h), and the hydrolysate was dried under nitrogen and reconstituted in 100 μL of H₂O. The absolute configurations of the amino acids Ser, Leu, and Pro were determined by enantioselective HPLC-MS [column, Chirobiotic TAG (250 × 4.6 mm), Supelco; solvent, MeOH–10 mM NH₄OAc (40:60, pH 5.30 for Thr, Val, Ser, Leu, and Pro; 90:10, pH 5.3 for Ile); flow rate, 0.5 mL/min]. The acid hydrolysates of dapalide A showed retention times at 7.8, 13.6, and 28.5 min corresponding to L-Ser, D-Pro, and D-Leu respectively, and peaks at 10.3 and 9.2 min for L-Ile and D-*allo*-Ile, and 6.2 and 7.9 min for L-Thr and D-*allo*-Thr, respectively. The remaining α stereocenter chirality amino acid retention times revealed a mixture of L- and D-amino acids. The retention times (t_R , min; MS ion pair) of the authentic amino acids were as follows: L-Thr (6.4; 120), L-*allo*-Thr (6.7), D-*allo*-Thr (7.9), D-Thr (7.0), L-Val (7.0; 118.12), D-Val (9.3), L-Ser (6.5; 106), D-Ser (7.2), L-Ile (10.3; 132.1), L-*allo*-Ile (8.2), D-*allo*-Ile (9.2), D-Ile (26.6), L-Pro (10.5; 116), D-Pro (25.8) and L-Leu (9.9; 132.1), D-Leu (11.5). The MS parameters used were as follows: DP – 60.0, EP – 7.0, CE – 28.0, CXP – 7.4, CUR 40, CAD High, IS – 4500, TEM 750, GS1 40, GS2 40.

Metagenomic DNA Extraction, Sequencing, and Analysis. The cyanobacteria sample was stored frozen in RNAlater solution (Thermo). The metagenomic DNA of VPG23–80 was isolated using the QIAamp DNA Micro Kit (Qiagen) and further purified using the DNeasy PowerClean Pro Cleanup kit (Qiagen) following manufacturer's protocol. For PacBio sequencing, metagenomic DNA was first PCR amplified following the Ampli-Fi library preparation method protocol. Purified and size-selected PCR products (>5kb) were used for library construction using SMRTbell prep kit 3.0 following the manufacturer's protocol.⁵⁴ The SMRTbell library was sequenced on

PacBio Revio Sequencer using the SPRQ chemistry. Raw reads were processed using SMRT Link version 25.1.0.257715 to generate CCS reads. CCS reads were assembled using Flye (version 2.9.2-b1786) in -meta mode.⁵⁵ BGC prediction was performed on metagenomic assembly using antiSMASH 5.1.1, and the adenylation domain substrate specificity was predicted using NRPSpredictor2, also incorporated within antiSMASH tool.⁴⁴ Binning was performed on metagenomic assembly using HiFi-MAG-Pipeline for high quality MAG.⁵⁶

Cell Viability Assays. HCT 116 and Hep G2 cells were seeded (10,000 cells/well) in 96-well plates and incubated for 24 h. Cells were then treated in triplicates with varying concentrations of dapalide A (1), kahalalide F (KF), and vehicle control (DMSO) and incubated for 48 h. The MTT assay was used as a proxy for cell viability. Specifically, MTT reagent was added to measure mitochondrial function and then stop solution after 2 h. After incubating overnight, the absorbance was measured at 540 nM using SpectraMax MS plate reader. % cell viability and % growth calculations were done by GraphPad Prism 5.03 based on the triplicate experiments. Growth calculations were conducted using t_0 and t_{48} time points by calculating the growth at time of MTT addition at t_{48} and dividing it by the cell population at the time of compound addition at t_0 with the vehicle calculated into both factors.

Saccharomyces cerevisiae Growth Inhibition Assay. An overnight culture of *S. cerevisiae* was grown in YPD medium. Cells were diluted to 5×10^4 cells/mL in media, then seeded at 100 μ L per well in a 96 well plate. The cells were treated with 1 μ L of dapalide A (1) in DMSO at various concentrations (100 μ M, 10 μ M, 1 μ M) and the plate was incubated for 24 h at 30 °C before quantifying OD600 on the SpectraMax MS plate reader with 1% DMSO serving as the negative control.

ASSOCIATED CONTENT⁴

Data Availability Statement⁵

The NMR data for compounds 1–3 have been deposited in the Natural Products Magnetic Resonance Database (NP-MRD; www.np-mrd.org) and can be found at the NP ID numbers NP0351406 (https://np-mrd.org/natural_products/NP0351406), NP0351407 (https://np-mrd.org/natural_products/NP0351407), and NP0351408 (https://np-mrd.org/natural_products/NP0351408). The sequences of the *dap* BGC and the 16S rRNA gene of *Dapis* sp. VPG23-80 were deposited at NCBI under Project NCBI accession PX021390 (*Dapis* sp. VPG23-80 MAG-1 16S rRNA) and PX048350 (dapalide BGC).

Supporting Information⁷

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jnatprod.5c00757>.

1D and 2D NMR spectra for all new compounds (PDF)⁹

AUTHOR INFORMATION¹⁰

Corresponding Authors

Mohamed S. Donia – Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, United States; Email: donia@princeton.edu

Hendrik Luesch – Department of Medicinal Chemistry and Center for Natural Products, Drug Discovery, and Development, University of Florida, Gainesville, Florida 32610, United States; Program in Cancer and Stem Cell Biology, Duke-NUS Medical School, Singapore 169857, Singapore; orcid.org/0000-0002-4091-7492; Email: luesch@cop.ufl.edu

Authors¹²

- Emma K. Ellis – Department of Medicinal Chemistry and Center for Natural Products, Drug Discovery, and Development, University of Florida, Gainesville, Florida 32610, United States
Laura P. Ióca – Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, United States; orcid.org/0000-0002-9288-6405
Jie Liu – Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, United States
Manyun Chen – Department of Medicinal Chemistry, University of Florida, Gainesville, Florida 32610, United States; orcid.org/0000-0001-6803-044X
Steven D. Bruner – Center for Natural Products, Drug Discovery, and Development, University of Florida, Gainesville, Florida 32610, United States; Department of Chemistry, University of Florida, Gainesville, Florida 32611, United States; orcid.org/0000-0002-0522-480X
Yousong Ding – Department of Medicinal Chemistry and Center for Natural Products, Drug Discovery, and Development, University of Florida, Gainesville, Florida 32610, United States; orcid.org/0000-0001-8610-0659
Valerie J. Paul – Smithsonian Marine Station, Fort Pierce, Florida 34949, United States; orcid.org/0000-0002-4691-1569

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jnatprod.5c00757>¹⁴

Author Contributions¹⁵

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. E.K.E. and L.P.I. contributed equally.

Notes¹⁷

The authors declare no competing financial interest.¹⁸

ACKNOWLEDGMENTS¹⁹

This research is supported by the NIH grant RM1GM145426 (H.L., Y.D., S.D.B., M.S.D., V.J.P.), the NIH grant R35GM128742 (Y.D.), and the Debbie and Sylvia DeSantis Chair professorship (H.L.). E.K.E. is supported by NIGMS grant T32GM136583: Chemistry–Biology Interface Training Program at the University of Florida. We thank the Guam Department of Agriculture Division of Aquatic and Wildlife Resources for the research permit (SCR-MPA-23-007) and the University of Guam Marine Laboratory (UOGML) for hosting our visit. We thank UOGML staff and Jay Houk (Smithsonian Marine Station) for assistance with collections and Jay Houk for extracting this cyanobacterium. A portion of this work was performed in the McKnight Brain Institute at the National High Magnetic Field Laboratory’s Advanced Magnetic Resonance Imaging and Spectroscopy (AMRIS) Facility, which is supported by the National Science Foundation Cooperative Agreement DMR-2128556 and the State of Florida.

REFERENCES²¹

- (1) MarinLit - A database of the marine natural products literature. <https://marinlit.rsc.org/> (accessed 2025–06–12).
- (2) Marine Pharmacology. Mysite. <https://www.marinepharmacology.org> (accessed 2024–12–15).
- (3) Carroll, A. R.; Copp, B. R.; Davis, R. A.; Keyzers, R. A.; Prinsep, M. R. *Nat. Prod. Rep.* **2022**, 39 (6), 1122–1171.

- (4) Salvador-Reyes, L. A.; Luesch, H. *Nat. Prod. Rep.* **2015**, *32* (3), 478–503.
- (5) Luesch, H.; Ellis, E. K.; Chen, Q.-Y.; Ratnayake, R. *Nat. Prod. Rep.* **2025**, *42* (2), 208–256.
- (6) Zhou, Q.; Hotta, K.; Deng, Y.; Yuan, R.; Quan, S.; Chen, X. *Microorganisms* **2021**, *9* (12), 2551.
- (7) Xu, Y.; Du, X.; Yu, X.; Jiang, Q.; Zheng, K.; Xu, J.; Wang, P. *Mar. Drugs* **2022**, *20* (6), 341.
- (8) Piel, J. Bacterial Symbionts: Prospects for the Sustainable Production of Invertebrate-Derived Pharmaceuticals. *Curr. Med. Chem.* **2006**, *13* (1), 39–50.
- (9) Pettit, G. R.; Kamano, Y.; Herald, C. L.; Tuinman, A. A.; Boettner, F. E.; Kizu, H.; Schmidt, J. M.; Baczyński, L.; Tomer, K. B.; Bontems, R. *J. Am. Chem. Soc.* **1987**, *109* (22), 6883–6885.
- (10) Luesch, H.; Moore, R. E.; Paul, V. J.; Mooberry, S. L.; Corbett, T. H. *J. Nat. Prod.* **2001**, *64* (7), 907–910.
- (11) Kallifidas, D.; Dhakal, D.; Chen, M.; Chen, Q.-Y.; Kokkaliari, S.; Colon Rosa, N. A.; Ratnayake, R.; Bruner, S. D.; Paul, V. J.; Ding, Y.; Luesch, H. *Org. Lett.* **2024**, *26* (7), 1321–1325.
- (12) Engene, N.; Tronholm, A.; Salvador-Reyes, L. A.; Luesch, H.; Paul, V. J. *J. Phycol.* **2015**, *51* (4), 670–681.
- (13) Kobayashi, M.; Tanaka, J.; Katori, T.; Matsuura, M.; Kitagawa, I. *Tetrahedron Lett.* **1989**, *30* (22), 2963–2966.
- (14) Bubb, M. R.; Spector, I.; Bershadsky, A. D.; Korn, E. D. *J. Biol. Chem.* **1995**, *270* (8), 3463–3466.
- (15) Andrianasolo, E. H.; Gross, H.; Goeger, D.; Musafija-Girt, M.; McPhail, K.; Leal, R. M.; Mooberry, S. L.; Gerwick, W. H. *Org. Lett.* **2005**, *7* (7), 1375–1378.
- (16) Ueoka, R.; Uria, A. R.; Reiter, S.; Mori, T.; Karbaum, P.; Peters, E. E.; Helfrich, E. J. N.; Morinaka, B. I.; Gugger, M.; Takeyama, H.; Matsunaga, S.; Piel, J. *Nat. Chem. Biol.* **2015**, *11* (9), 705–712.
- (17) Hamann, M. T.; Scheuer, P. *J. Am. Chem. Soc.* **1993**, *115* (13), 5825–5826.
- (18) Hamann, M. T.; Otto, C. S.; Scheuer, P. J.; Dunbar, D. C. *J. Org. Chem.* **1996**, *61* (19), 6594–6600.
- (19) Zan, J.; Li, Z.; Tianero, Ma. D.; Davis, J.; Hill, R. T.; Donia, M. S. *Science* **2019**, *364* (6445), No. eaaw6732.
- (20) Bonnard, I.; Manzanares, I.; Rinehart, K. L. *J. Nat. Prod.* **2003**, *66* (11), 1466–1470.
- (21) Jiménez, J. C.; López-Macià, A.; Gracia, C.; Varón, S.; Carrascal, M.; Caba, J. M.; Royo, M.; Francesch, A. M.; Cuevas, C.; Giralt, E.; Albericio, F. *J. Med. Chem.* **2008**, *51* (16), 4920–4931.
- (22) Shilabin, A. G.; Kasanah, N.; Wedge, D. E.; Hamann, M. T. *J. Med. Chem.* **2007**, *50* (18), 4340–4350.
- (23) López-Macià, A.; Jiménez, J. C.; Royo, M.; Giralt, E.; Albericio, F. *J. Am. Chem. Soc.* **2001**, *123* (46), 11398–11401.
- (24) Goetz, G.; Yoshida, W. Y.; Scheuer, P. *J. Tetrahedron* **1999**, *55* (25), 7739–7746.
- (25) Gao, J.; Hamann, M. T. *Chem. Rev.* **2011**, *111* (5), 3208–3235.
- (26) Wyer, S.; Townsend, D. M.; Ye, Z.; Kourtidis, A.; Choo, Y.-M.; de Barros, A. L. B.; Donia, M. S.; Hamann, M. T. *Biomed. Pharmacother.* **2022**, *148*, No. 112676.
- (27) Ciavatta, M. L.; Devi, P.; Carbone, M.; Mathieu, V.; Kiss, R.; Casapullo, A.; Gavagnin, M. *Tetrahedron* **2016**, *72* (5), 625–631.
- (28) Shilabin, A. G.; Hamann, M. T. *Bioorg. Med. Chem.* **2011**, *19* (22), 6628–6632.
- (29) Suárez, Y.; González, L.; Cuadrado, A.; Berciano, M.; Lafarga, M.; Muñoz, A.; Kahalalide, F. *Mol. Cancer Ther.* **2003**, *2* (9), 863–872.
- (30) Rademaker-Lakhai, J. M.; Horenblas, S.; Meinhhardt, W.; Stokvis, E.; de Reijke, T. M.; Jimeno, J. M.; Lopez-Lazaro, L.; Lopez Martin, J. A.; Beijnen, J. H.; Schellens, J. H. M. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **2005**, *11* (5), 1854–1862.
- (31) Pardo, B.; Paz-Ares, L.; Tabernero, J.; Ciruelos, E.; García, M.; Salazar, R.; López, A.; Blanco, M.; Nieto, A.; Jimeno, J.; Izquierdo, M. Á.; Trigo, J. M. *Clin. Cancer Res.* **2008**, *14* (4), 1116–1123.
- (32) Hamann, M. T. *Curr. Opin. Mol. Ther.* **2004**, *6* (6), 657–665.
- (33) Martín-Algarra, S.; Espinosa, E.; Rubió, J.; López, J. J. L.; Manzano, J. L.; Carrión, L. A.; Plazaola, A.; Tanovic, A.; Paz-Ares, L. *Eur. J. Cancer* **2009**, *45* (5), 732–735.
- (34) Molina-Guijarro, J. M.; Macías, Á.; García, C.; Muñoz, E.; García-Fernández, L. F.; David, M.; Núñez, L.; Martínez-Leal, J. F.; Moneo, V.; Cuevas, C.; Lillo, M. P.; Villalobos Jorge, C.; Valenzuela, C.; Galmarini, C. M. *PLoS One* **2011**, *6* (4), No. e19042.
- (35) Ratain, M. J.; Geary, D.; Undevaia, S. D.; Coronado, C.; Alfaro, V.; Iglesias, J. L.; Schilsky, R. L.; Miguel-Lillo, B. *Invest. New Drugs* **2015**, *33* (4), 901–910.
- (36) Petty, R.; Anthoney, A.; Metges, J.-P.; Alsina, M.; Gonçalves, A.; Brown, J.; Montagut, C.; Gunzer, K.; Laus, G.; Iglesias Dios, J. L.; Miguel-Lillo, B.; Bohan, P.; Salazar, R. *Cancer Chemother. Pharmacol.* **2016**, *77* (4), 819–827.
- (37) Kokkaliari, S.; Luo, D.; Paul, V. J.; Luesch, H. *Mar. Drugs* **2023**, *21* (7), 378.
- (38) Al-Awadhi, F. H.; Kokkaliari, S.; Ratnayake, R.; Paul, V. J.; Luesch, H. *J. Nat. Prod.* **2024**, *87* (10), 2355–2365.
- (39) Chen, X.-Q.; Luo, Z.; Liu, K.; Tang, Z.-Z.; Peng, L.; Gao, C.; Xia, C.; Liu, Y.-H. *Rec. Nat. Prod.* **2024**, No. 1, 136–142.
- (40) Li, M.; Li, S.; Hu, J.; Gao, X.; Wang, Y.; Liu, Z.; Zhang, W. *Mar. Drugs* **2022**, *20* (1), 33.
- (41) Cai, W.; Matthew, S.; Chen, Q.-Y.; Paul, V. J.; Luesch, H. *Bioorg. Med. Chem.* **2018**, *26* (9), 2310–2319.
- (42) Taori, K.; Matthew, S.; Rocca, J. R.; Paul, V. J.; Luesch, H. *J. Nat. Prod.* **2007**, *70* (10), 1593–1600.
- (43) Keller, L.; Canuto, K. M.; Liu, C.; Suzuki, B. M.; Almaliti, J.; Sikandar, A.; Namani, C. B.; Glukhov, E.; Luo, D.; Duggan, B. M.; Luesch, H.; Koehnke, J.; O'Donoghue, A. J.; Gerwick, W. H. *ACS Chem. Biol.* **2020**, *15* (3), 751–757.
- (44) 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.
- (45) Engene, N.; Tronholm, A.; Paul, V. J. *J. Phycol.* **2018**, *54* (4), 435–446.
- (46) Martin, H.; Lazaro, L. R.; Gunnlaugsson, T.; Scanlan, E. M. *Chem. Soc. Rev.* **2022**, *51* (23), 9694–9716.
- (47) Ryu, B.; Avalon, N.; Cuau, M.; Almaliti, J.; Din, M. O.; Brennan, C.; Glukhov, E.; Knight, R.; Gerwick, W. Cyanobacteria Join the Kahalalide Conversation: Genome and Metabolite Evidence for Structurally Related Peptides. *ChemRxiv* June 6, 2025, DOI: 10.26434/chemrxiv-2025-dmqr1.
- (48) Soares, A. R.; Engene, N.; Gunasekera, S. P.; Snead, J. M.; Paul, V. J. *J. Nat. Prod.* **2015**, *78* (3), 534–538.
- (49) Youssef, D. T. A.; Mufti, S. J.; Badiab, A. A.; Shaala, L. A. *Mar. Drugs* **2022**, *20* (12), 768.
- (50) Kurisawa, N.; Iwasaki, A.; Jeelani, G.; Nozaki, T.; Suenaga, K. *J. Nat. Prod.* **2020**, *83* (5), 1684–1690.
- (51) Yamazaki, Y.; Yamabe, S.; Komori, S.; Yoshitake, K.; Fukuoka, M.; Sato, S.; Takada, K. *J. Nat. Prod.* **2023**, *86* (11), 2539–2545.
- (52) Chen, J.; Jia, Y.; Sun, Y.; Liu, K.; Zhou, C.; Liu, C.; Li, D.; Liu, G.; Zhang, C.; Yang, T.; Huang, L.; Zhuang, Y.; Wang, D.; Xu, D.; Zhong, Q.; Guo, Y.; Li, A.; Seim, I.; Jiang, L.; Wang, L.; Lee, S. M. Y.; Liu, Y.; Wang, D.; Zhang, G.; Liu, S.; Wei, X.; Yue, Z.; Zheng, S.; Shen, X.; Wang, S.; Qi, C.; Chen, J.; Ye, C.; Zhao, F.; Wang, J.; Fan, J.; Li, B.; Sun, J.; Jia, X.; Xia, Z.; Zhang, H.; Liu, J.; Zheng, Y.; Liu, X.; Wang, J.; Yang, H.; Kristiansen, K.; Xu, X.; Mock, T.; Li, S.; Zhang, W.; Fan, G. *Nature* **2024**, *633* (8029), 371–379.
- (53) Trindade, M.; van Zyl, L. J.; Navarro-Fernández, J.; Abd Elrazak, A. *Front. Microbiol.* **2015**, *6*, 890.
- (54) Documentation. PacBio. <https://www.pacb.com/support/documentation/> (accessed 2025-06-12).
- (55) Kolmogorov, M.; Bickhart, D. M.; Behsaz, B.; Gurevich, A.; Rayko, M.; Shin, S. B.; Kuhn, K.; Yuan, J.; Polevikov, E.; Smith, T. P. L.; Pevzner, P. A. *Nat. Methods* **2020**, *17* (11), 1103–1110.
- (56) PacificBiosciences/Pb-Metagenomics-Tools, 2025. <https://github.com/PacificBiosciences/pb-metagenomics-tools> (accessed 2025-06-12).