Title	Nostosin G and Spiroidesin B from the Cyanobacterium Dolichospermum sp. NIES-1697
Author(s)	Phan, Chin-Soon; Mehjabin, Jakia Jerin; Anas, Andrea Roxanne J. et al.
Citation	Journal of Natural Products, 85(8), 2000-2005 https://doi.org/10.1021/acs.jnatprod.2c00382
Issue Date	2022-08-10
Doc URL	https://hdl.handle.net/2115/90284
Rights	This document is the Accepted Manuscript version of a Published Work that appeared in final form in Journal of Natural Products, copyright c American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see https://pubs.acs.org/doi/full/10.1021/acs.jnatprod.2c00382.
Type	journal article
File Information	Nat.Prod.J_v.85 n.8(2022).pdf

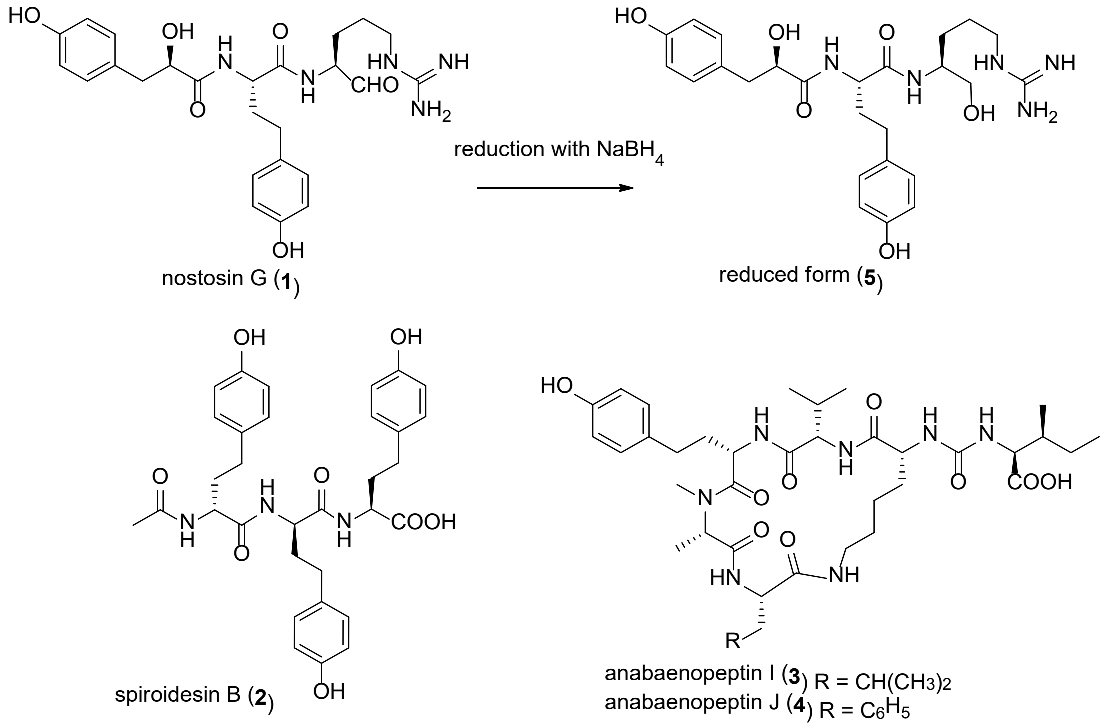
Nostosin G and Spiroidesin B from the Cyanobacterium *Dolichospermum* sp. NIES-1697

Chin-Soon Phan,[§#] Jakia Jerin Mehjabin,[§#] Andrea Roxanne J. Anas,[†] Masahiro Hayasaka,[†] Reiko Onoki,[†] Juting Wang,[†] Taiki Umezawa,[†,§] Kenji Washio,[†,§] Masaaki Morikawa,[†,§] Tatsufumi Okino[*,†,§]

[†]Graduate School of Environmental Science and [§]Faculty of Environmental Earth Science, Hokkaido University, Kita-ku, Sapporo 060-0810, Japan

ABSTRACT: Chemical investigation of the cyanobacterium *Dolichospermum* sp. NIES-1697 afforded nostosin G (**1**), a linear tripeptide, spiroidesin B (**2**), and two known compounds, anabaenopeptins I (**3**) and J (**4**). Planar structures and absolute configurations for **1** and **2** were determined by 2D NMR, HRMS, Marfey's methodology, chiral-phase HPLC, and enzymatic degradation. Nostosin G (**1**) is a unique example of a linear peptide containing three subunits, 4-hydroxyphenyllactic acid (Hpla), homotyrosine (Hty), and argininal, with potent trypsin inhibitory properties. The biosynthetic gene clusters for nostosin G (**1**) and spiroidesin B (**2**) were investigated based on the genome sequence of *Dolichospermum* sp. NIES-1697.

The continuing exploration of natural products from cyanobacteria is a point of discussion to date. Many recent developments have been reported regarding the exploration of natural products from cyanobacteria. A few recent articles about cyanobacterial natural products are the discoveries of the cyanobactin tolypamide from *Tolypothrix* sp.,[1] nocuolactylates and hapalosins from *Nodularia* sp. LEGE 06071 and *Fischerella* sp. PCC9431, respectively[2] cyclic lipopeptide puwainaphycins and minutissamides from *Nodularia harveyana* UHCC-0300,[3] pentabrominated biindole alkaloid aetokthonotoxin from *Aetokthonos hydrillicola*,[4] guanidine bis-prenylated peptide argicyclamide A from *Microcystis aeruginosa* NIES-88,[5] and chlorinated fatty acid amide columbamides F, G, and H from *Moorea bouillonii*.[6] Interestingly, these secondary metabolites often exhibited biological activities such as antitrypanosomal, biosurfactant, cytotoxic, and enzyme inhibitory properties.[3-8] In previous studies, we have isolated spumigin J from *Dolichospermum compacta* NIES-835 (formerly *Anabaena compacta*). Both spumigins A and J showed thrombin and cathepsin B inhibitory activities.[7] Hence, as part of our search for new bioactive natural products from cyanobacteria, we investigated an unexplored strain from the NIES collection, *Dolichospermum* sp. NIES-1697. The chemical investigation of this strain has led to the isolation of two new metabolites, nostosin G (**1**), spiroidesin B (**2**), and two known compounds, anabaenopeptins I (**3**) and J (**4**) (Figure 1). Herein, we report the isolation, structure elucidation, absolute configurations, trypsin inhibitory activities, and genomic analyses of biosynthetic gene clusters for nostosin G (**1**) and spiroidesin B (**2**).



nostosin G (**1**)

reduced form (**5**)

spiroidesin B (**2**)

anabaenopeptin I (**3**) R = $CH(CH_3)_2$

anabaenopeptin J (**4**) R = C_6H_5

Figure 1. Structures of **1**-**5** and chemical reduction of **1**.

RESULTS AND DISCUSSION

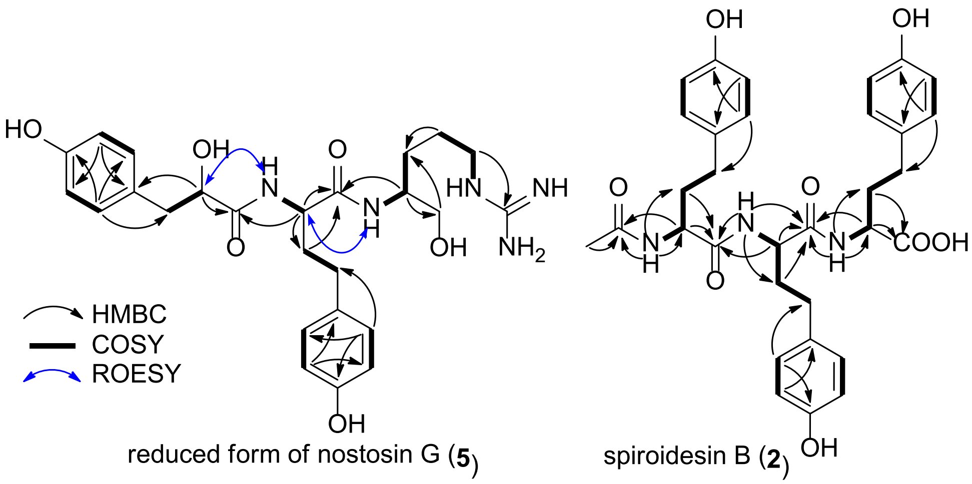
Nostosin G (**1**) was obtained as a mixture of interconverting cyclic and linear forms (Figure S1) from the hydrophilic fraction of the *Dolichospermum* sp. NIES-1697 culture grown in 50 L of BG-11 medium. Cyclic hemiaminal formation of the argininal (Argal) residue was observed in spumigins and aeruginosins, in which the sodium borohydride ($NaBH_4$) reduction was used to convert the mixture into a single peak containing argininol (Argol).[8-11] In addition to nostosin G (**1**), spiroidesin B (**2**), anabaenopeptins I (**3**), and J (**4**) were also isolated from the hydrophilic fraction.

Table 1. NMR Spectroscopic Data (DMSO-d_6) for **5** (Reduced Form of Nostosin G)

Unit	position	δ_C[a], type	δ_H (J in Hz)[b]	HMBC[c]
Hpla[d]	1	173.2, C		
	2	72.5, CH	4.05 dd (8.6, 3.8)	1, 3, 4
	3	39.6, CH_2	2.88 dt (13.8, 3.8)	1, 2, 4, 5/9
			2.55 ddd (13.8, 8.6, 4.1)	1, 2, 4, 5/9
	4	128.6, C		
	5/9	130.2, CH	7.01 d (7.6)	3, 5/9, 7
	6/8	114.7, CH	6.65 m	4, 6/8
	7	155.6, C		
Hty	1	171.0, C		
	2	52.1, CH	4.26 m	1, 3, 4 Hpla 1
	3	34.9, CH_2	1.85 m	1, 2, 4, 5
			1.76 m	1, 2, 4, 5
	4	30.3, CH_2	2.39 m	2, 3, 5, 6/10
	5	131.3, C		
	6/10	129.0, CH	6.91 d (7.9)	4, 6/10, 8
	7/9	115.1, CH	6.65 m	5, 7/9
	8	155.3, C		
	NH		7.84 d (8.3) / 7.79 d (8.3)	2, Hpla 1
Argol	1	63.1, CH_2	3.37 td (10.7, 4.8)	2, 3
			3.29 td (10.7, 5.8)	2, 3
	2	50.2, CH	3.72 m	1, Hty 1
	3	27.8, CH_2	1.58 m	2, 4, 5
			1.31 m	2, 4, 5
	4	25.1, CH_2	1.49 m	2, 3, 5
			1.43 m	2, 3, 5
	5	40.6, CH_2	3.07 m	4, 6
	6	156.7, C		
	NH		7.75 d (8.3)	2, Hty 1
	N^{δ}H		7.54 d (8.3)	4, 5, 6

[a]Measured at 150 MHz. [b]Measured at 600 MHz. [c]HMBC correlations are from the proton(s) stated to the indicated carbon. [d]Hpla = 4-hydroxyphenyllactic acid.

The structure (**5**) of a reduced form of **1** was determined by NMR, MS, and Marfey's methodology (Table 1, Figures S16-17). A molecular formula of $C_{25}H_{35}N_5O_6$ was determined based on the HRESIMS deprotonated molecule $[M - H]^-$ at *m/z* 500.2505. The ^1H and ^{13}C NMR spectra in DMSO-d_6 (Table 1) showed three sp^3 methines at δ_C 72.5, 52.1, and 50.2; δ_H 4.26, 4.05, and 3.72 suggested signals for three amino acid Cα methines, which later confirmed to be homotyrosine (Hty), argininol (Argol), and hydroxyphenyllactic acid (Hpla). Two disubstituted benzenes were deduced from overlapped 1D NMR signals at δ_C 130.2, 129.0, 115.1, and 114.7; δ_H 7.01, 6.91, and 6.65, COSY correlations of δ_H 7.01 to 6.65 and 6.91 to 6.65, and HMBC correlations of signals at δ_H 7.01, 6.91 and 6.65 to its corresponding HSQC cross-peaks. Additional correlations from both COSY and HMBC spectra confirmed the presence of Hpla and Hty (Figure 2, Table 1). The presence of Argol was determined by the remaining correlations of both COSY and HMBC data (Figure 2, Table 1). The key HMBC peaks of the Cα methines at δ_H 4.26, 4.05, and 3.72 to carbonyl carbons at δ_C 173.2 and 171.0, together with selected ROESY correlations between NH signals (δ_H 7.79/7.84 and 7.75) and Cα methines (δ_H 4.26 and 4.05) allowed the connection of a linear peptide of Hpla-Hty-Argol (Figure 2). The presence of isomers was confirmed when most signals of Hty-Argol found in ^{13}C NMR were observed as split peaks (Figure S3). The structures of these isomers were determined as *R*-Hpla-L-Hty-D-Argol and *R*-Hpla-L-Hty-L-Argol based on Marfey's analyses of the non-reduced form and authentic Hpla chromatographic comparison (Figures S16-17). However, we believe nostosin G (**1**) was originally synthesized as *R*-Hpla-L-Hty-L-Argal due to the absence of an epimerase domain in its biosynthetic gene cluster (*nst*), as shown in Figure 3. Most analogs of spumigins and aeruginosins have four residues,[7-10,12-15], and only a few examples contain three units.[11,16]



reduced form of nostosin G (**5**) spiroidesin B (**2**)

Figure 2. 2D NMR correlations of the reduced form of nostosin G (**5**) and spiroidesin B (**2**).

Compound **2** has the molecular formula of $C_{32}H_{37}N_3O_8$ with a HRESIMS deprotonated molecule $[M - H]^-$ at *m/z* 590.2511. The ^{1}H and ^{13}C NMR spectra in DMSO-d_6 (Table 2) showed three Cα methines of α-amino acids at δ_C 52.7, 52.3, and 51.1; δ_H 4.30, 4.26, and 4.10, three NH protons at δ_H 8.16, 8.13 and 7.91, four carbonyl carbons at δ_C 173.5, 171.7, 171.5 and 169.6, where one carbonyl is part of an acetyl group, and the rest indicated three amino acid residues. Three sets of two methylenes and disubstituted benzenes showed the existence of three homotyrosines (Hty). The COSY and HMBC correlations (Figure 2) confirmed a linear structure of *N*Ac-Hty-Hty-Hty. Marfey's analysis of the acid hydrolysate of **2** indicated the presence of both L- and D-Hty. The L- and D-Hty sequence was determined by the cleavage of C-terminal amino acid of **2** using carboxypeptidase Y.[17] The Hty released from the reaction was purified using HPLC along with dipeptide, *N*Ac-Hty-Hty (Figure S18). The molecular formula of the dipeptide was confirmed to be $C_{22}H_{26}N_2O_6$ from the observed $[M+H]^+$ ion at *m/z* 415.1876 (calculated for $[M+H]^+$ 415.1869). From Marfey's analysis of the hydrolysates of the dipeptide and the released Hty, the structure of spiroidesin B was confirmed as *N*Ac-D-Hty-D-Hty-L-Hty (Figures S19-20). A close example to **2** was spiroidesin, hexanoyl-D-Hty-L-Hty-D-Phe, isolated from the cyanobacterium *Anabaena* *spiroides*.[18] The known compounds, anabaenopeptins I (**3**) and J (**4**), were identified in comparison with literature data.[19]

Table 2. NMR Spectroscopic Data (DMSO-d_6) for Spiroidesin B (**2**)

Unit	position	δ_C[a], type	δ_H (*J* in Hz)[b]	HMBC[c]
*N*Ac	1	169.6, C		
	2	22.5, CH_3	1.89 s	1
Hty (1)	1	171.7, C		
	2	52.7, CH	4.26 td (8.6, 5.2)	1, *N*Ac 1
	3	33.9, CH_2	1.91 m	1, 2, 4, 5
			1.76 m	1, 2, 4, 5
	4	30.7, CH_2	2.44-2.51 m	2, 3, 5, 6/10
	5	131.5, C		
	6/10	129.1, CH	6.95 d (8.0)	4, 6/10, 8
	7/9	115.1, CH	6.65 d (8.0)	5, 7/9
	8	155.4, C		
	NH		8.16 d (8.0)	2, *N*Ac 1
Hty (2)	1	171.5, C		
	2	52.3, CH	4.30 td (8.6, 4.6)	1, Hty(1) 1
	3	34.5, CH_2	1.92 m	1, 2, 4, 5
			1.80 m	1, 2, 4, 5
	4	30.4, CH_2	2.40-2.50 m	2, 3, 5, 6/10
	5	131.4, C		
	6/10	129.1, CH	6.95 d (8.0)	4, 6/10, 8
	7/9	115.1, CH	6.65 d (8.0)	5, 7/9
	8	155.3, C		
	NH		7.91 d (8.0)	1, 3, Hty(1) 1
Hty (3)	1	173.5, C		
	2	51.1, CH	4.10 td (9.2, 4.6)	1, Hty(2) 1
	3	33.2, CH_2	1.90 m	1, 2, 4, 5
			1.83 m	1, 2, 4, 5
	4	30.4, CH_2	2.40-2.50 m	2, 3, 5, 6/10
	5	130.9, C		

6/10	129.2, CH	6.91 d (8.0)	4, 6/10, 8
7/9	115.1, CH	6.63 d (8.0)	5, 7/9
8	155.4, C		
NH		8.13 d (8.0)	2, 3, Hty(2) 1

[a]Measured at 150 MHz. [b]Measured at 600 MHz. [c]HMBC correlations are from the proton(s) stated to the indicated carbon.

To reveal the biosynthetic gene clusters of these peptides, we sequenced the genome of *Dolichospermum* sp. NIES-1697. AntiSMASH software (version 5.0),[20] allowed us to propose the assembly lines for nostosin G (**1**), spiroidesin B (**2**), and anabaenopeptins I (**3**) and J (**4**). The putative nostosin G biosynthetic gene cluster (*nst*) was found to consist of two genes, NstAB as nonribosomal peptide synthase (NRPS), where NstA is a ketoreducing NRPS (Figure 3). Ketoreducing NRPS module are composed of A-KR-PCP sequences with ketoreductase (KR) domain that catalyzed β-ketoreduction on α-keto acid to hydroxy acid residue.[21] The NRPSs were identified to composed of three NRPS modules. They were predicted to produce a peptide containing three units consistent with the structure of nostosin G, where NstA A_1 domain (VGVWIAASGK), NstB A_1 domain (DLGFTGCV), and NstB A_2 domain (DVETTGAV) activated Hpla, Hty, and Arg, respectively.[14] The architecture of the NRPSs was similar to a typical spumigin biosynthetic gene cluster containing four NRPS modules [9,10,14,22,23] in contrast to three NRPS modules for the biogenesis of nostosin G (Figure 3). As for the biosynthesis of Hty in nostosin G, three putative enzymes (Figure 3) found in the upstream region of *nst* are believed to be involved in homotyrosine synthesis, and share homology to HphA (74.74%; YP_001865968), HphB (79.72%; YP_001865961), and HphCD (82.24%; YP_001865962) from homophenylalanine biosynthetic genes in cyanobacterium *Nostoc punctiforme* PCC73102.[24] The region upstream of *nst* encodes four putative NRPSs (Figure 3) that share homology to AptA

(91.41%; WP_006196339), AptB (85.85%; WP_006196340), AptC (73.43%; WP_006196341), and AptD (77.43%; WP_006196342) from the biosynthesis of anabaenopeptin in cyanobacterium *Nodularia spumigena* CCY9414,[25] suggesting these four putative NRPSs are most likely associated with the biosynthesis of the isolated anabaenopeptins I (**3**) and J (**4**). This gene organization (Figure 3), where nostosin and anabaenopeptin assembly lines are co-located in the genome and are separated by a 13 kb region containing putative HphBCD, was also observed in the genome of the cyanobacterium *Sphaerospermopsis torques-reginae* ITEP-024, which is also known as a producer of Hty containing spumigins K-N.[8,22] A homology analysis that was carried out on the Integrated Microbial Genomes and Microbiomes (IMG/M) System using NstA as query, found five homolog genes (Figure S23).[26] The domain structures of these homologs SpuA and SpuB shared similarities to NstA and NstB except for their genetic organization (Figure S24).

A unique feature in the gene cluster was the presence of an unrelated NRPS positioned between NstA and NstB (Figure 3). This is different from the typical *spu* assembly line (from spumigin biosynthesis), where SpuA and SpuB are located side by side, without another NRPS in between.[9,10,14,22,23] An attempt to predict the activated substrates for all three A domains of SprA via *in silico* prediction showed identical specificity-conferring codes DLGFTGCV, suggesting that they activated the same amino acid. We believed that this is Hty based on the high similarity of the binding pocket sequence to the AptC A_2 domain (DLGFAGCV), in accordance with the structures of anabaenopeptins I (**3**) and J (**4**) found in this study strain. The binding pockets activated Hty residue from anabaenopeptin biosynthetic gene clusters have AptC A_2 domains (DLGFTGVC) in several *Planktothrix* strains and (DLGFTGCV) from *Sphaerospermopsis torquesreginae* ITEP-024,[22,27] and SpuB A_1 domain (DLGFTGCV) from spumigin biosynthesis

(Table S1).[14] Thus, we reason that peptide D-Hty-D-Hty-L-Hty of spiroidesin B (**2**) was a product of this NRPS (SprA), spiroidesin B biosynthetic gene cluster (*spr*). The existence of epimerases found in modules 1 and 2 within SprA matched the structure determined by NMR, cleavage of the C-terminal amino acid, and Marfey's analyses. The structure of **2** has an acetyl group at *N*-terminus, which is most likely catalyzed by an acetyltransferase. This can be exemplified by the MvdB acetyltransferase catalyzed Tyr *N*-acetylation in the biosynthesis of microviridin K.[28] However, no such acetyltransferase was encoded in the regions adjacent to this NRPS. The *N*-acetyltransferase involved in the production of **2** may be located at a different locus; this has been seen in the *N*-acetylation of erythrochelin, where *N*-acetyltransferase EryM/Mcd (SACE_1304) was located outside of erythrochelin biosynthetic gene cluster.[29] Two putative acetyltransferases (NG743_14015 and NG743_14050) that share homology with the NeuD *O*-acetyltransferase at 95.45% (WP_168358461.1) and 98.62% (WP_200888902.1), respectively, were found in a different locus that was separated by a distance of 1 Mbp from the *spr* gene cluster. The acetylation likely occurs after the tethered tripeptide is hydrolytically released by the TE domain. Further study is needed to elucidate this biosynthetic pathway.

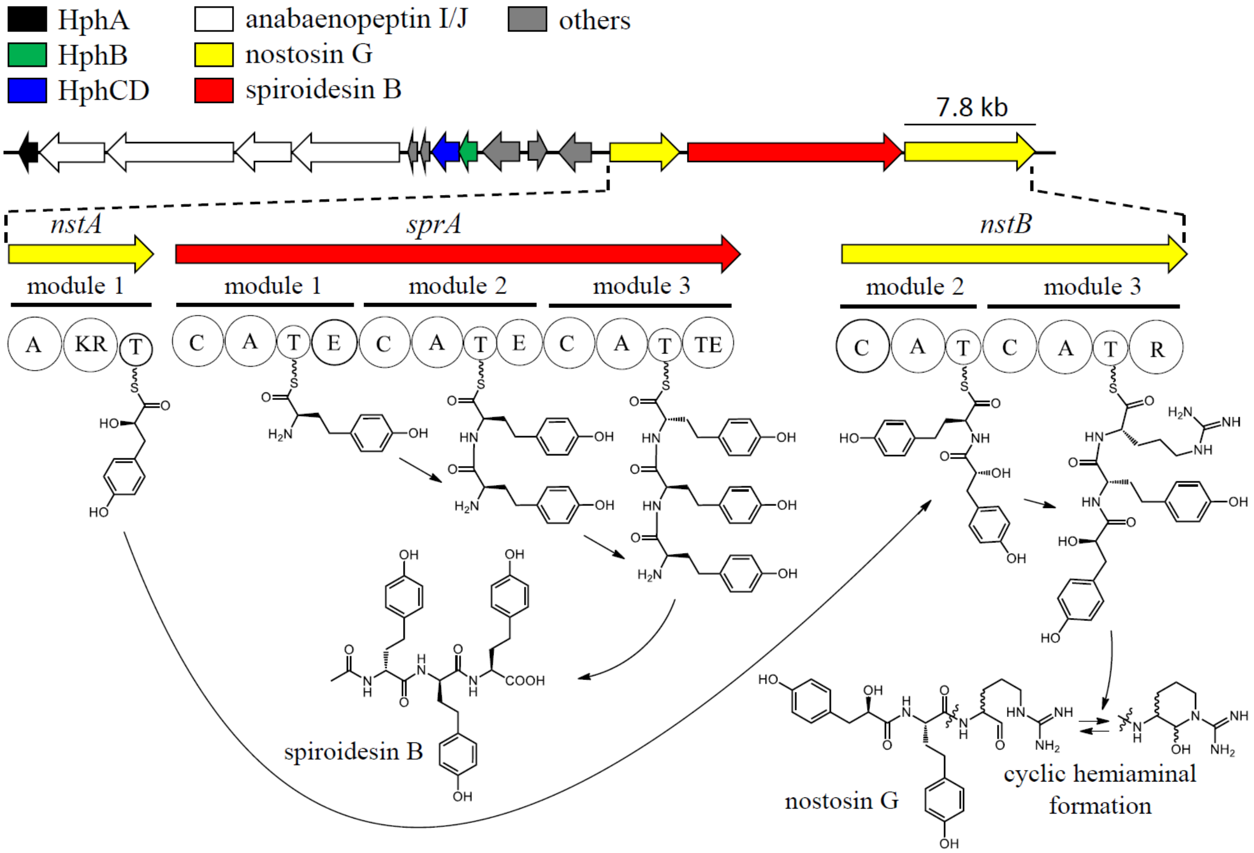


Figure 3. Proposed nostosin G and spiroidesin B biosynthetic gene clusters. Domains of A, adenylation; KR, ketoreductase; T, thiolation; C, condensation; E, epimerization; TE, thioesterase; R, reductase.

The linear peptide spumigins, aeruginosins, and nostosins are reported as serine and cysteine protease inhibitors and showed potent inhibitory activities against thrombin, trypsin, plasmin, and cathepsin B.[7,12,14,16] The mixture of L/D-Argal **1** (non-reduced) inhibited trypsin with IC_{50} value 0.1 μM. Similar trypsin inhibitory properties were found in the related compounds, spumigin E[14] and nostosin A,[16] containing an Argal unit. The reduced form of nostosin G (**5**) did not show any significant activities (IC_{50} value >200 μM) because the Argol unit could not covalently bind with trypsin active sites.[8,16] Spiroidesin B (**2**), anabenopeptin I (**3**), and anabenopeptin J (**4**) were also subjected to trypsin inhibitory properties, however, none of them showed significant activities. This study discovered two new natural products, nostosin G (**1**) and spiroidesin B (**2**) from *Dolichospermum* sp. NIES-1697, whose biosynthetic gene clusters were proposed based on the genome sequence, and nostosin G (**1**) exhibited trypsin inhibition with an IC_{50} value of 0.1 μM.

EXPERIMENTAL SECTION

General Experimental Procedures

Optical rotation and UV data were measured on a HORIBA SEPA-300 polarimeter and JASCO V-650 spectrophotometer, respectively, with MeOH as the solvent. ECD spectrum was taken using a JASCO J-720 spectropolarimeter. ^{1}H and ^{13}C NMR spectra were measured on a JEOL ECA 600 NMR spectrometer. Chemical shifts were denoted in δ (ppm) relative to residual solvent peaks as internal standard (DMSO-d_6, δ_H 2.50, δ_C 39.5). High-performance liquid chromatography (HPLC) experiments were performed with a JASCO HPLC system equipped with a PU-980 pump and UV-970 UV/vis detector. HR-ESI (-) MS experiments were performed with Agilent 1100 Series HPLC system coupled with a micrOTOF-HS mass spectrometer (Bruker Daltonics).

Culture Conditions

Dolichospermum sp. NIES-1697 was grown in BG-11 media,[30] with aeration (filtered air, 0.3 L/min) at 25 °C under illumination of 80 $\mu E/m^2s^1$ on a 12L:12D cycle. Cells were collected by continuous flow centrifugation at 10,000 rpm after 4-5 weeks of incubation.

DNA Extraction and Genome Sequencing

High molecular weight DNA was extracted by first crushing the frozen cells with a mortar and pestle, then dissolving the powder with CTAB buffer (3% CTAB, 1.4 M NaCl, 0.2% β-mercaptoethanol, 20 mM EDTA, 100 mM Tris pH 8, RNase A), and incubating at 50 °C for 30 min. Later, 750 μL $CHCl_3$ was added, then centrifuged with an inverted tube at 15,000 rpm for 5 min. The upper layer was gently transferred into a centrifuge tube containing 200 μL isopropanol. The appearance of a white string (precipitated DNA) became visible after gentle rotation of the tube; this precipitated DNA was collected for genomic sequencing using the PacBio RS II system. The 125x coverage data was obtained. The reads were subjected to *de novo* assembly using Hierarchical Genome Assembly Process (HGAP v.2), resulting in one circular chromosome (accession number: CP099464) with 5,833,376 Mbp of sequence data, which contained 38.52% G + C and 5,969 open reading frames. The genome sequence was analyzed by AntiSMASH software (version 5.0) for the detection of biosynthetic gene clusters.[20]

Extraction and Isolation

Freeze-dried cells (2.8 g from 50 L of culture in 4-5 weeks) were homogenized and extracted with MeOH (200 mL × 3). The concentrated extract was partitioned between H_2O and EtOAc. The H_2O (202.8 mg) extracts were applied to ODS (YMC-Gel, 150 μm) with an aqueous MeOH elution system to obtain 20 %, 40 %, 60 %, and 100 % MeOH fractions. The combined fractions of 20 % and 40 % MeOH fractions (100.0 mg) were purified by HPLC with C18 (Wakosil-II 5C18 AR, 20 x 250 mm, UV detection 215 nm, flow rate 4.0 mL/min) and isocratic condition (0-40 min, 33 % MeCN with 0.1 % TFA) to yield **2** (7.0 mg), **3** (1.0 mg) and **4** (7.0 mg). Reduction of a half portion of the 60 % MeOH fraction (20.0 mg from 40.0 mg) was done with $NaBH_4$ to improve peak resolution. Subsequently, this reduction product was purified by HPLC with C18 (Wakosil-II 5C18 AR, 20 x 250 mm, UV detection 210 nm, flow rate 4.0 mL/min) and isocratic condition (0-20 min, 20 % MeCN with 0.1 % TFA) to yield **5** (8.0 mg).

Reduced form of nostosin G (**5**): Colorless oil; $[\alpha]^{27}_D$ +9.4 (*c* 0.67, MeOH); ECD (0.001 M, MeOH) λ_{max} ($\Delta\varepsilon$) 226 (2.25) nm, (supporting information S22); UV (MeOH) λ_{max} (log ε) 278 (3.71) nm, 224 (4.27) nm; NMR (DMSO-d_6, 600 MHz) see Table 1; HRESIMS *m/z* 500.2505 $[M - H]^-$ (calcd for $C_{25}H_{34}N_5O_6$, 500.2515).

Spiroidesin B (**2**): Colorless oil; $[\alpha]^{18}_D$ -52 (*c* 0.4, MeOH); UV (MeOH) λ_{max} (log ε) 278 (3.44) nm 224 (4.26) nm; NMR (DMSO-d_6, 600 MHz) see Table 2; HRESIMS *m/z* 590.2511 $[M - H]^-$ (calcd for $C_{32}H_{36}N_3O_8$, 590.2508).

Oxidation, Hydrolysis and Advanced Marfey's Analyses for the Absolute Configurations of the Amino Acids

Nostosin G (**1**) was dissolved in 30 % H_2O_2 and one drop of conc. HCl was added. The solution was kept at room temperature for 24 hours. Then the solution was lyophilized and hydrolyzed. Hydrolysis of **1**, and **2** (0.2 mg) was executed by adding 200 μL 6 M HCl solutions. The solution was placed in a vacuum hydrolysis tube and heated at 110 °C for 16 h. The hydrolysates were dried and dissolved in 100 μL H_2O. The aqueous solutions of amino acid standards and hydrolysates of **1** and **2** were subjected to Marfey's analysis by adding 200 μL FDLA solution (in acetone) followed by 50 μL 1M $NaHCO_3$, and the resulting solution was heated to 40 °C for 1 h.[31] The reaction mixture was cooled to room temperature and neutralized with 50 μL 2 M HCl solution. The resulting residues were resuspended in 200 μL MeCN and directly analyzed by LC-MS. The LC-MS analysis of arginine derivatives from Marfey's analysis was performed on a cadenza CD-C18 column (2 x 150 mm) with a linear gradient elution of 20-80% MeCN containing 0.1% formic acid for 30 min and a flow rate of 0.2 mL/min. For the analysis of homotyrosine derivatives, a linear gradient of 25% MeCN containing 0.1% formic acid to 65% MeCN containing 0.1% formic acid using YMC-Pack Pro C18 column (2 x 50 mm) and 0.2 mL/min flow rate was used. The enzymatic reaction applying carboxypeptidase Y to cleave the COOH-terminal Hty of compound **2** was done by the previously reported method.[17] Compound **2** (1 mg) was dissolved in 200 μL of 0.1 M pyridine acetate buffer (pH 5.5), and a few μL of carboxypeptidase Y were added. The solution was incubated for 16 h and dried to remove the buffer. The remaining dipeptide and released Hty were purified using HPLC with Cosmosil MS-II (4.6 x 250 mm, UV detection 215 nm, flow rate 0.5 mL/min) and gradient condition (0-30 min, 5-30% MeCN with 0.1% TFA). The dipeptide containing two Hty moieties was hydrolyzed using 6 M HCl at 110 °C for 16 h, followed by Marfey's analysis to determine the sequence of Hty for **2**. The masses and retention time of the DLA derivatives of the L- and D-amino acid standards were as follows (*m/z*, L-, D-): Arg (469, 15 min, 14 min); Hty (490, 20 min, 23 min).

Determination of the Absolute Configuration of Hydroxyphenyllactic Acid (Hpla)

The acid hydrolysate of **1** was extracted with ethyl ether, which separated Hpla from the amino acid mixtures. The ether extract of the acid hydrolysates was then dried *in vacuo* and dissolved in MeOH. The MeOH solution was analyzed on an Astec Chirobiotic TAG column (250 x 2.1 mm), with a flow rate of 0.3 mL/min, UV detection at 277 nm, and a linear elution with 19:1 MeOH: 1% aq. triethylammonium acetate buffer (pH 4). The coelution experiments of the Hpla derivative from **1** with an authentic standard of *R*/*S*-Hpla confirmed the absolute configuration. The retention time of hydrolysate of **1** has t_R 3.6 min, *R*-Hpla has t_R 3.6 min and *S*-Hpla has t_R 3.2 min.

Trypsin Inhibition Assay

All of the enzymes and substrates used in this study were purchased from Sigma Chemical, Co. The trypsin was dissolved in 50 mM Tris-HCl (pH 7.6) to prepare 150 U/mL enzyme solutions. The *N*α-Benzoyl-D/L-arginine-4-nitroanilide hydrochloride (43.3 mg) was dissolved in 1 mL dimethyl sulfoxide (DMSO). The DMSO solution was diluted 100 times with 50 mM Tris-HCl (pH 7.6) and used as the substrate for the trypsin assay. In this study, each 200 µL assay mixture containing 30 µL of Tris-HCl buffer, 50 µL of enzyme solution, and 20 µL of test solution or inhibitor were added to each microtiter well using a 96-well plate.[32] The assay mixture and the substrate solution were preincubated at 37 °C for 5 min, and then 100 µL substrate solutions were added to start the reaction. The absorbance of the well was measured immediately and after 30 min incubation at 405 nm. Leupeptin from Peptide Institute, *Inc* was used as a positive control with an IC_{50} value of 0.6 µM.

ASSOCIATED CONTENT

Supporting Information

^{1}H NMR, ^{13}C NMR, HSQC, HMBC, COSY, ROESY spectra of compounds **1** and **2**, LC-MS chromatogram of compounds **1** and **2**, Marfey's derivatives of the hydrolysates of **1** and **2**.

AUTHOR INFORMATION

Corresponding Author

*Phone: +81 11 706 4519; Email: okino@ees.hokudai.ac.jp; Fax: +81 11 706 4867

Author Contributions

#C.-S.P. and J.J.M. contributed equally.

Notes: The authors declare no competing financial interest.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 19F19096. C.-S.P. is a recipient of the JSPS Postdoctoral Fellowship for Foreign Researchers (ID No. P19096).

REFERENCES

1. Purushothaman, M.; Sarkar, S.; Morita, M.; Gugger, M.; Schmidt, E. W.; Morinaka, B. I. *Angew. Chem. Int. Ed*. **2021**, *60*, 8460-8465.

2. Figueiredo, S. A. C.; Preto, M.; Moreira, G.; Martins, T.; Abt, K.; Melo, A.; Vasconcelos, V. M.; Leão, P. N. *Angew. Chem. Int. Ed*. **2021**, *60*, 10064-10072.

3. Saurav, K.; Caso, A.; Urajová, P.; Hrouzek, P.; Esposito, G.; Delawská, K.; Macho, M.; Hájek, J.; Cheel, J.; Saha, S.; Divoká, P.; Arsin, S.; Sivonen, K.; Fewer, D. P.; Costantino, V. *ACS Omega* **2022**, *7*, 11818-11828.

4. Breinlinger, S.; Phillips, T. J.; Haram, B. N.; Mareš, J.; Martínez Yerena, J. A.; Hrouzek, P.; Sobotka, R.; Henderson, W. M.; Schmieder, P.; Williams, S. M.; Lauderdale, J. D.; Wilde, H. D.; Gerrin, W.; Kust, A.; Washington, J. W.; Wagner, C.; Geier, B.; Liebeke, M.; Enke, H.; Niedermeyer, T. H. J.; Wilde, S. B. *Science* **2021**, *371*, eaax9050.

5. Phan, C.-S.; Matsuda, K.; Balloo, N.; Fujita, K.; Wakimoto, T.; Okino, T. *J. Am. Chem. Soc.* **2021**, *143*, 10083-10087.

6. Mehjabin, J. J.; Wei, L.; Petitbois, J. G.; Umezawa, T.; Matsuda, F.; Vairappan, C. S.; Morikawa, M.; Okino, T. *J. Nat. Prod.* **2020**, *83*, 1925-1930.

7. Anas, A. R. J.; Kisugi, T.; Umezawa, T.; Matsuda, F.; Campitelli, M. R.; Quinn, R. J.; Okino, T. *J. Nat. Prod.* **2012**, *75*, 1546-1552.

8. Sanz, M.; Salinas, R. K.; Pinto, E. *J. Nat. Prod.* **2017**, *80*, 2492-2501.

9. Fewer, D. P.; Jokela, J.; Paukku, E.; Österholm, J.; Wahlsten, M.; Permi, P.; Aitio, O.; Rouhiainen, L.; Gomez-Saez, G. V.; Sivonen, K. *PLoS ONE* **2013**, *8*, e73618.

10. Jokela, J.; Heinilä, L. M. P.; Shishido, T. K.; Wahlsten, M.; Fewer, D. P.; Fiore, M. F.; Wang, H.; Haapaniemi, E.; Permi, P.; Sivonen, K. *Front. Microbiol.* **2017**, *8*, 1963.

11. Nowruzi, B.; Wahlsten, M.; Jokela, J. *Iran. J. Biotech.* **2019**, *17*, 71-78.

12. Matsuda, H.; Okino, T.; Murakami, M.; Yamaguchi, K. *Tetrahedron* **1996**, *52*, 14501-14506.

13. Fujii, K.; Sivonen, K.; Adachi, K.; Noguchi, K.; Sano, H.; Hirayama, K.; Suzuki, M.; Harada, K. *Tetrahedron Lett.* **1997**, *38*, 5525-5528.

14. Fewer, D.; Jokela, J.; Rouhiainen, L.; Wahlsten, M.; Koskenniemi, K.; Stal, L. J.; Sivonen, K. *Mol. Microbiol.* **2009**, *73*, 924-937.

15. Mazur-Marzec, H.; Kaczkowska, M. J.; Blaszczyk, A.; Akcaalan, R.; Spoof, L.; Meriluoto, J. *Mar. Drugs* **2013**, *11*, 1-19.

16. Liu, L.; Jokela, J.; Wahlsten, M.; Nowruzi, B.; Permi, P.; Zhang, Y. Z.; Xhaard, H.; Fewer, D. P.; Sivonen, K. *J. Nat. Prod*. **2014**, *77*, 1784-1790.

17. Hayashi, R.; Moore, S.; Stein, W. H. *J. Biol. Chem.* **1973**, *248*, 2296-2302.

18. Kaya, K.; Mahakhant, A.: Keovara, L.; Sano, T.; Kubo, T.; Takagi, H. *J. Nat. Prod*. **2002**, *65*, 920-921.

19. Murakami, M.; Suzuki, S.; Itou, Y.; Kodami, S.; Ishida, K. *J. Nat. Prod*. **2000**, *63*, 1280-1282.

20. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Madema, M. H.; Weber, T. *Nucleic Acids Res*. **2019**, *47*, W81-W87.

21. Alonzo, D. A.; Chiche-Lapierre, C.; Tarry, M. J.; Wang, J.; Schmeing, T. M. *Nat. Chem. Biol.* **2020**, *16*, 493-496.

22. Lima, S. T.; Alvarenga, D. O.; Etchegaray, A.; Fewer, D. P.; Jokela, J.; Varani, A. M.; Sanz, M.; Dörr, F. A.; Pinto, E.; Sivonen, K.; Fiore, M. F. *ACS Chem. Biol*. **2017**, *12*, 769-778.

23. Voβ, B.; Bolhuis, H.; Fewer, D. P.; Kopf, M.; Möke, F.; Haas, F.; El-Shehawy, R.; Hayes, P.; Bergman, B.; Sivonen, K.; Dittmann, E.; Scanlan, D. J.; Hagemann, M.; Stal, L. J.; Hess, W. R. *PLoS ONE* **2013**, *8*, e60224.

24. Koketsu, K.; Mitsuhashi, S.; Tabata, K. *Appl. Environ. Microbiol*. **2013**, *79*, 2201-2208.

25. Rouhiainen, L.; Jokela, J.; Fewer, D. P.; Urmann, M.; Sivonen, K. *Chem. Biol*. **2010**, *17*, 265-273.

26. Chen, I.-M. A.; Chu, K.; Palaniappan, K.; Ratner, A.; Huang, J.; Huntemann, M.; Hajek, P.; Ritter, S.; Varghese, N,; Seshadri, R.; Roux, S.; Woyke, T.; Eloe-Fadrosh, E. A.; Ivanova, N. N.; Kyrpides, N. C. *Nucleic Acids Res*. **2021**, *49*, D751-D763.

27. Christiansen, G.; Philmus, B.; Hemscheidt, T.; Kurmayer, R. *J. Bacteriol*. **2011**, *193*, 3822-3831.

28. Philmus, B.; Christiansen, G.; Yoshida, W. Y.; Hemscheidt, T. K. *ChemBioChem*. **2008**, *9*, 3066-3073.

29. Lazos, O.; Tosin, M.; Slusarczyk, A. L.; Boakes, S.; Cortés, J.; Sidebottom, P. J.; Leadlay, P. F. *Chem. Biol*. **2010**, *17*, 160-173.

30. Waterbury, J. B.; Stanier, R. Y. *The Prokaryotes* **1981**, *7*, 221-223.

31. Marfey, P. *Carlsberg Res. Commun*. **1984**, *49*, 591–596.

32. Shin, H. J.; Murakami, M.; Matsuda, H.; Yamaguchi, K. *Tetrahedron* **1996**, *52*, 8159-8168.