Organic & Biomolecular Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Org. Biomol. Chem.*, 2022, **20**, 2681

Discovery of varlaxins, new aeruginosin-type inhibitors of human trypsins†

L. M. P. Heinilä, [a] J. Jokela,[a] M. N. Ahmed, [a,b] M. Wahlsten, [a] S. Kumar, [c] P. Hrouzek, [c] P. Permi,[d,e] H. Koistinen, [b] D. P. Fewer *[a] and K. Sivonen *[a]

Low-molecular weight natural products display vast structural diversity and have played a key role in the development of novel therapeutics. Here we report the discovery of novel members of the aeruginosin family of natural products, which we named varlaxins. The chemical structures of varlaxins 1046A and 1022A were determined using a combination of mass spectrometry, analysis of one- and two-dimensional NMR spectra, and HPLC analysis of Marfey's derivatives. These analyses revealed that varlaxins 1046A and 1022A are composed of the following moieties: 2-*O*-methylglyceric acid 3-*O*-sulfate, isoleucine, 2-carboxy-6-hydroxyoctahydroindole (Choi), and a terminal arginine derivative. Varlaxins 1046A and 1022A differ in the cyclization of this arginine moiety. Interestingly, an unusual α-D-glucopyranose moiety derivatized with two 4-hydroxyphenylacetic acid residues was bound to Choi, a structure not previously reported for other members of the aeruginosin family. We sequenced the complete genome of *Nostoc* sp. UHCC 0870 and identified the putative 36 kb varlaxin biosynthetic gene cluster. Bioinformatics analysis confirmed that varlaxins belong to the aeruginosin family of natural products. Varlaxins 1046A and 1022A strongly inhibited the three human trypsin isoenzymes with IC_{50} of 0.62–3.6 nM and 97–230 nM, respectively, including a prometastatic trypsin-3, which is a therapeutically relevant target in several types of cancer. These results substantially broaden the genetic and chemical diversity of the aeruginosin family and provide evidence that the aeruginosin family is a source of strong inhibitors of human serine proteases.

Received 21st December 2021,
Accepted 3rd March 2022

DOI: 10.1039/d1ob02454j

rsc.li/obc

Introduction

Advanced solid cancers are often difficult to treat because of the ability of tumour cells to escape from the primary tumour and form distant metastases.[1] Proteases play a critical role in tumour progression, especially in the process of metastatic dissemination of cancer cells and the promotion of tumour growth.[2,3] Trypsin is one of the best-characterized proteolytic enzymes.[4,5] The *PRSS1*, *PRSS2* and *PRSS3* genes encode trypsinogen-1, -2 and -3 trypsin precursors, respectively, which are especially abundant in pancreatic secretions.[5] These precursors can be proteolytically activated (*e.g.*, by enteropeptidase) to trypsin-1, -2, and -3 (also known as cationic and anionic trypsin, and mesotrypsin, respectively). Trypsin-3 is an enigmatic minor component of pancreatic secretions with a proposed role of modulating overall trypsin activity, as it efficiently degrades proteinaceous trypsin inhibitors.[6,7] Elevated levels of trypsin-3 are associated with several human diseases. Trypsin-3 has also been identified as a potential therapeutic target for the reduction of tumour growth and metastasis in several cancers, such as prostate, breast and pancreatic cancer.[8–12] However, the high sequence identity between trypsins extends to the active sites of the enzymes and makes selective targeting of trypsin-3 challenging.[7,13–16] Nevertheless, human trypsin isoenzymes have slight differences in substrate specificity, which suggests that the development of selective trypsin inhibitors is possible.[17]

Aeruginosins are a group of tetrapeptides produced by cyanobacteria.[18] They often possess potent inhibitory activity on trypsin-type serine proteases and, thus, represent potential lead molecules for pharmaceutical applications.[18] This class of non-ribosomal peptides was originally discovered in *Microcystis aeruginosa*.[18,19] Aeruginosins have also been identified in other cyanobacterial genera, such as *Nodularia*,

[a]*Department of Microbiology, Faculty of Agriculture and Forestry, University of Helsinki, Helsinki, Finland. E-mail: david.fewer@helsinki.fi, kaarina.sivonen@helsinki.fi*
[b]*Department of Clinical Chemistry, University of Helsinki and Helsinki University Hospital, Helsinki, Finland*
[c]*Laboratory of Algal Biotechnology-Centre Algatech, Institute of Microbiology of the Czech Academy of Sciences, Třeboň, Czech Republic*
[d]*Department of Chemistry, University of Jyväskylä, Jyväskylä, Finland*
[e]*Department of Biological and Environmental Science, Nanoscience Center, University of Jyväskylä, Jyväskylä, Finland*

†Electronic supplementary information (ESI) available. See DOI: 10.1039/d1ob02454j

Planktothrix, and *Nostoc.* Aeruginosin subgroups include pseudoaeruginosins,[20] nostosins,[21] suomilide,[22,23] and banyasides.[24] They have a complex chemical structure and typically contain the non-proteogenic 2-carboxy-6-hydroxyoctahydroindole (Choi) moiety and the C-terminal arginine derivatives argininal, argininol, agmatine, 1-amidino-2-ethoxy-3-aminopiperidine, and more rarely a 1-amino-2-(*N*-amidino-Δ^3-pyrrolinyl)-ethyl (Aaep) moiety (Fig. S1–S3†). Aeruginosins are sometimes decorated with sugars, which themselves may be further derivatized with fatty acids.[22,24,25] Aeruginosins are the products of a highly variable non-ribosomal peptide synthetase (NRPS) biosynthetic pathway.[23,26–31] The diversity of the aeruginosin family is quite broad and over 130 variants have been described to date (34, Table S1†). However, the chemical diversity of the aeruginosin family of natural products is underestimated given the known genetic diversity of this family.[28,33,34]

Members of the aeruginosin natural product family are often strong serine protease inhibitors.[18] It is noteworthy that all the previously reported aeruginosin groups contain well-known trypsin inhibitors.[18,20–24] The most potent microbial trypsin inhibitor reported from this family of natural products to date is chlorodysinosin A with an IC_{50} of 37 nM (Fig. S3 and Table S2†). Most reported biochemical assays employ nonhuman trypsin, such as porcine and bovine trypsin. However, the primary structures of porcine and bovine trypsins differ significantly from those of human trypsins.[35] Thus far, there has been only one study that examined aeruginosins (suomilide) as human trypsin inhibitors.[23] Notably, suomilide was also found to inhibit invasion of trypsin-3 expressing prostate cancer cells.[23] Here we report the discovery of two new members of the aeruginosin family of natural products, varlaxins 1022A and 1046A, that inhibit human trypsin isoenzymes in sub-nanomolar concentrations.

Experimental

Discovery of varlaxins

Varlaxins were discovered serendipitously from *Nostoc* sp. UHCC 0870 during routine screening of strains of the UHCC culture collection for the production of secondary metabolites. Varlaxins were subsequently identified from *Nostoc* spp. UHCC 0840, UHCC 0817, UHCC 0758 and UHCC 0757 (Table 1). *Nostoc* spp. UHCC 0840 and 0870 were collected 2014 from Varlaxudden, Porvoo from the brackish Gulf of Finland waterfront (N 60°12.225′, E 025°37.95′) and isolated into pure culture. *Nostoc* sp. UHCC 0817 was collected in 2012 from Kobben, Hanko, Finland from the rock surface of Gulf of Finland waterfront (N 59°49.92′, E 023°5.20′). *Nostoc* sp. UHCC 0758 was collected in 2003 from a mixture of gastropods collected from marina of Jurmo island, Finland (N 59°49.62′, E 021°35.05′). *Nostoc* sp. UHCC 0757 was collected from an undescribed location in Finland. Strains were grown for 3 weeks at ambient temperature (20–22 °C) under continuous photon irradiance of 15 μmol photons per m^2 per s in 40 mL of saline Z8 medium[36] lacking a source of combined nitrogen. Biomass was collected by centrifugation at 7000*g* for 7 min. The pellets were freeze-dried with a BETA 2–8 LSC plus freeze dryer with a LYO CUBE 4–8 chamber (Christ). Freeze-dried pellets were extracted with 1 mL of MeOH in 2 mL plastic tubes containing 0.5 mm glass beads by shaking with a FastPrep cell disrupter (Thermo Electron Corporation, Qbiogene, Inc.) for 30 s at a speed of 6.5 m s^{-1}. The extract was centrifuged at 10 000*g* for 5 min, and the supernatant was analysed with a high-resolution UPLC-QTOF (Acquity I-Class UPLC-Synapt G2-Si, Waters Corp., Milford, MA) system. A 1 μl aliquot was injected into a Kinetex C8 column (50 × 2 mm, 1.7 μm, Phenomenex Inc.), which was eluted at 40 °C with a flow rate of 0.3 mL min^{-1} from 5% of solvent B (acetonitrile or 1 : 1 acetonitrile/isopropanol with formic acid added to 0.1%) in 0.1% formic acid (solvent A) to 95% of B in 5 min, held for 2 min, then back to 5% of B in 0.5 min, and finally kept for 2.5 min before the next run. In MS/MS mode, the column was eluted from 5% of solvent B (1 : 1 acetonitrile/isopropanol with formic acid added to 0.1%) in 0.1% formic acid (solvent A) to 60% of B in 10 min, then in 0.01 min to 100% of B, held for 3.99 min, then back to 5% of B in 0.5 min, and finally kept for 5.5 min before the next run. QTOF was calibrated using sodium formate and Ultramark® 1621, which yielded a calibrated mass range from *m*/*z* 91 to 1921. Leucine Enkephalin was used at 10 s intervals as a lock mass reference compound. Data were collected in positive electrospray ionization Resolution Mode at scan range of *m*/*z* 50–2000. In MS/MS mode, parameters were same as in MS mode and Trap Collision Energy Ramp proceeded from 40.0 eV to 80.0 eV.

Varlaxin purification

Nostoc sp. UHCC 0870 was cultivated in a 100 L flat-panel bioreactor in Z8 medium bubbled with air enriched with 1.5% CO_2 at a constant temperature (28 °C) and illumination of 50 μmol photon per m^2 per s to obtain 14 g of lyophilized biomass. Biomass was extracted three times with 50% MeOH (in H_2O). The extract was then concentrated under vacuum using a rotary evaporator. The resulting extract (2.6 g) was chromatographed over normal-phase silica gel (100–200 mesh) flash column yielding 13 fractions eluted with step gradient hexane : diethyl ether (100 : 0, 25 : 75, 50 : 50 and 0 : 100) followed by chloroform : MeOH (100 : 0, 99 : 1, 98 : 2, 96 : 4, 92 : 8, 84 : 16, 68 : 32, 50 : 50 and 0 : 100). Fraction 10 (188 mg in total) eluted with 84 : 16 (chloroform : MeOH) was fractionated on preparative HPLC (Agilent 1260 Preparative Chromatograph equipped with quaternary pumps and MWD detector) on a Phenomenex Kinetex C18 column (250 × 21.2 mm, 5 μm) using H_2O (A)/acetonitrile (B) both containing 0.1% formic acid as a mobile phase at a flow rate of 10 mL min^{-1}. The linear gradient was as follows: A/B 85/15 (0 min), 85/15 (over 2 min), 0/100 (over 40 min), and 0/100 (over 45 min). Fraction 6 (eluted at 32–35 min, 25 mg) containing the desired compounds was further purified by semipreparative HPLC on a Phenomenex Kinetex EVO C18 column Phenomenex Kinetex EVO C18 column (250 × 10 mm, 5 μm, 100 Å) eluted with H_2O (A)/acetonitrile (B) both containing 0.1% formic acid at a flow rate of

Table 1 Varlaxin variants identified from five strains in the UHCC culture collection. Mgs = 2-*O*-methylglyceric acid 3-*O*-sulfate, Choi = 2-carboxy-6-octahydroindole, Agma = 4-amidinobutylamide, Aaep = 1-amidino-3-(2-aminoethyl)-3-pyrroline, Hpaa = 2-(4-hydroxyphenyl)acetic acid. Δ = difference in parts per million (ppm) between calculated and experimental $[M + Na]^+$. RI = percentage relative intensities of varlaxin variants in a strain. $[M - 80 + H]^+$ = desulfated (80 = SO_3) protonated ion

					Structural units						UHCC 0870		UHCC 0840		UHCC 0817		UHCC 0758		UHCC 0757	
			$[M - 80 + H]^+$	$[M + Na]^+$	Core				Side group		$[M + Na]^+$	$[M - 80 + H]^+$	$[M + Na]^+$	$[M - 80 + H]^+$	$[M + Na]^+$	$[M - 80 + H]^+$	$[M + Na]^+$	$[M - 80 + H]^+$	$[M + Na]^+$	$[M - 80 + H]^+$
No	Var[a]	t_R (min)	Calculated (m/z)[b]		1	2[c]	3	4	Sugar[d]	Hpaa	Δ (ppm)	RI (%)	Δ (ppm)	RI (%)	Δ (ppm)	RI (%)	Δ (ppm)	RI (%)	Δ (ppm)	RI (%)
1	740	1.00	661.3767	763.3154	Mgs	Val	Choi	Agma	D-Glc	—					−2.3[e]	0.1	−0.4	0.1	0.9	0.4
2	**764**	1.04	**685.3767**	787.3154	Mgs	Val	Choi	Aaep	D-Glc	—					−0.9[e]	1.5	**−0.7**	**1.8**	−0.5	0.5
3	**754**	1.24	**675.3923**	777.3311	Mgs	Ile	Choi	Agma	D-Glc	—	−3.5	0.05	7.7	0.04	−2.4	0.3	**−0.7**	**2.0**	1.6	5.6
4	**778**	1.30	**699.3923**	801.3311	Mgs	Ile	Choi	Aaep	D-Glc	—	−1.5	0.1	−3.3	0.2	0.1	10.5	**0.8**	**39**	2.2	7.9
5	874A	1.56	795.4135	897.3522	Mgs	Val	Choi	Agma	D-Glc	1	−1.9	0.02			−1.2	0.1	−2.5	0.02	−3.5	0.02
6	898A	1.61	819.4135	921.3522	Mgs	Val	Choi	Aaep	D-Glc	1	−0.7	0.05			−1.3	1.9	−0.7	0.4	−3.3	0.1
7	888A	1.64	809.4291	911.3679	Mgs	Ile	Choi	Agma	D-Glc	1	−3.0	0.3			−1.7	0.3	−1.8	0.4	−0.7	0.4
8	**912A**	1.67	**833.4291**	935.3679	Mgs	Ile	Choi	Aaep	D-Glc	1	−1.6	1.1	−6.3	0.2	−0.1	11.9	**0.4**	**12**	0.8	1.9
9	874B	1.69	795.4135	897.3522	Mgs	Val	Choi	Agma	D-Glc	1	−0.6	0.1			−1.6	0.1	−4.1	0.1	−0.6	0.1
10	898B	1.74	819.4135	921.3522	Mgs	Val	Choi	Aaep	D-Glc	1	−4.1	0.1	−1.3	0.1	−0.5	2.9	−0.6	1.1	0.4	0.5
11	**888B**	1.75	**809.4291**	911.3679	Mgs	Ile	Choi	Agma	D-Glc	1	−2.5	1.0	−3.1	0.6	−0.5	0.6	**−0.8**	**1.5**	0.1	2.3
12	912B	1.80	833.4291	935.3679	Mgs	Ile	Choi	Aaep	D-Glc	1	−1.3	2.2	−1.9	3.1	0.4	27.8	1.5	32	1.3	11
13	1008A	2.02	929.4502	1031.3890	Mgs	Val	Choi	Agma	D-Glc	2	−1.3	0.7	0.3	0.2	0.7	0.1			0.0	0.1
14	1032A	2.06	953.4502	1055.3890	Mgs	Val	Choi	Aaep	D-Glc	2	−1.9	2.6	−3.7	1.1	0.7	2.9	−1.4	0.3	0.5	0.4
15	**1022A**	2.07	**943.4659**	1045.4047	Mgs	Ile	Choi	Agma	D-Glc	2	−2.6	25	**0.1**	**12**	0.4	0.5	−0.8	0.2	1.7	7.1
16	**1046A**	2.10	**967.4659**	1069.4047	Mgs	Ile	Choi	Aaep	D-Glc	2	**−0.3**	**55**	−0.8	76	0.7	24.3	0.4	6.4	1.8	35
17	1008B	2.13	929.4502	1031.3890	Mgs	Val	Choi	Agma	D-Glc	2	−2.0	0.2	−2.3	0.3					−0.5	0.3
18	1032B	2.14	953.4502	1055.3890	Mgs	Val	Choi	Aaep	D-Glc	2	−3.7	0.5			0.9	1.6			1.3	1.7
19	1022B	2.20	943.4659	1045.4047	Mgs	Ile	Choi	Agma	D-Glc	2	−3.0	2.8	−0.4	4.2	−0.4	0.3	0.4	0.1	0.6	4.1
20	1046B	2.22	967.4659	1069.4047	Mgs	Ile	Choi	Aaep	D-Glc	2	−1.8	8.2			0.8	12.5	0.7	2.4	1.3	20
21	1142	—	1063.4870	1165.4258	Mgs	Val	Choi	Agma	D-Glc	3										
22	1166	—	1087.4870	1189.4258	Mgs	Val	Choi	Aaep	D-Glc	3										
23	1156	2.46	1077.5027	1179.4414	Mgs	Ile	Choi	Agma	D-Glc	3	−3.2	0.1	−5.1	0.1						
24	1180	2.45	1101.5027	1203.4414	Mgs	Ile	Choi	Aaep	D-Glc	3	−4.1	0.2	−3.6	0.4						

[a] Numbering base to exact mass of the neutral molecule. Bold font = product ion spectrum annotated. [b] M − 80 originates from neutral loss of SO_3 from M. [c] According to aa analysis of 1022A and 1046A aa is Ile. In other variants could also be isobaric Leu. [d] Analyzed from 1022A/1046A, Hpaa = number of Hpaa (4-hydroxyphenylacetic acid) units. [e] From $[M - 80 + H]^+$.

3 mL min^{-1} using following gradient: A/B 85/15 (0 min), 85/15 (in 1 min), 72/28 (in 8 min), and 72/28 (in 30 min). The compound elution was monitored on the MWD detector set to 220 nm, yielding 0.7 mg of purified varlaxin 1046A and 0.6 mg of purified varlaxin 1022A, eluting at 20 min and 19 min, respectively. The purified varlaxins 1046A and 1022A contained approximately 3–4% of other minor varlaxin variants.

Amino acid analysis

Solid purified varlaxins 1046A and 1022A and reference amino acids were transferred to 200 μL glass tubes inside a closed 4 mL glass vial containing 500 μL of 6 M HCl and hydrolysed at 110 °C for 21 h. Inner-tube HCl residues were vacuum evaporated and 100 μL of 1% Marfey's reagent (1-fluoro-2,4-dinitrophenyl-5-L-alanine or -L-leucine amide in acetone), 50 μL of H_2O and 20 μL of 1 M $NaHCO_3$ were added. The solution was incubated at 37 °C for 1 h and the reaction was stopped by adding 20 μL of 1 M HCl. Commercial reference compounds L-Ile, D-Ile, L-*allo*-Ile, and D-*allo*-Ile were obtained from Sigma-Aldrich (St. Louis, MO). The L-Choi reference amino acid was obtained from *Microcystis aeruginosa* NIES-298.[19] Cells were grown, freeze dried, and extracted as described above. Solvent was evaporated from the crude extract and dry material was used as L-Choi reference material.

Monosaccharide analysis

D- and L-Glucose were purchased from Sigma-Aldrich. Hydrolysis of varlaxin 1022A and sample derivatizations were performed as previously described[37] with the following exceptions: approximately 0.4 mg of varlaxin 1022A was hydrolysed, freeze dried instead of drying under N_2 gas stream, and the reagents L-cysteine methyl ester (Sigma-Aldrich) and *o*-tolyl isothiocyanate (Thermo Fisher GmbH, Germany) instead of phenyl isothiocyanate were added simultaneously, and the reaction mixture was incubated at 60 °C for 1 h and diluted with MeOH before UPLC-QTOF analysis. Samples of 1 μL were analysed with two LC methods. The sample was injected into a Kinetex C8 column (50 or 100 × 2 mm, 1.7 μm, Phenomenex Inc.), which was eluted at 0.3 mL min^{-1} at 40 °C with 95% H_2O (+0.1% formic acid, eluent A) and 5% ACN : isopropanol (1 : 1, +0.1% formic acid, eluent B). Eluent B was increased linearly to 100% in 5 minutes, kept at 2 minutes, then back to the initial condition in 0.5 minutes with 2.5 min post run before next injection. Additionally, the sample was eluted with Acquity UPLC® BEH C18 column (100 × 2.1 mm, 1.7 μm Waters Corp., MA) as previously described, but eluent B was increased linearly to 70% in 10 minutes, then to 100% in 0.10 minutes, kept at 3.99 minutes, and changed back to initial condition in 0.5 minutes with 5.5 min post run. QTOF was used in resolution mode with positive electrospray ionization. Leucine enkephalin was used as a lock mass, and sodium formate and Ultramark® 1621 for mass calibration. The retention time of monosaccharide thiocarbamoyl-thiazolidine derivative of D-Glc of varlaxin 1022A was 1.97 min (references D-Glc 1.97 min, L-Glc 1.88 min).

NMR and UV spectroscopy

Spectra were collected on a Bruker Avance III HD 800 MHz NMR spectrometer equipped with a cryogenically cooled TCI ^1H, ^{13}C, ^{15}N triple-resonance probe head. Data were collected at 30 °C in DMSO-d_6 (Table S3†). In addition to a ^1H experiment with the solvent pre-saturation and broadband decoupled ^{13}C experiment, homonuclear, two-dimensional (2D) TOCSY (TOtal Correlation SpectroscopY), NOESY (Nuclear Overhauser Enhancement SpectroscopY) and DQF-COSY (Double Quantum Filtered COrrelation SpectroscopY) experiments were performed. Heteronuclear 2D ^1H–^{13}C HSQC and ^1H–^{15}N HSQC (Heteronuclear Single Quantum Coherence) and 2D ^1H–^{13}C HMBC (Heteronuclear Multiple Bond Correlation) experiments (for the assignment of ^1H, ^{13}C and ^{15}N chemical shifts) were also performed. The 2D TOCSY experiment was acquired with an isotropic mixing time of 60 ms. The mixing time for 2D NOESY spectrum was 300 ms. The ^{13}C HMBC experiment (for observing long-range ^1H–^{13}C connectivity) was measured using $^nJ_{CH}$ transfer time optimized for 8 Hz couplings. For monosaccharide analysis, varlaxin 1046A was hydrolysed for 1 h at 100 °C in 0.5 mL of 2 M D_2SO_4 in D_2O. The proton spectrum was collected, and anomeric signals were compared with data from a previous publication.[38] UV spectra of varlaxins 1022A and 1046A, and a reference compound 4-hydroxyphenylacetic acid were measured in MeOH with a UV-1800 UV-Vis spectrophotometer (Shimadzu, Kyoto, Japan).

Trypsin inhibition

The inhibition of human trypsins was determined as described previously.[23] The recombinant trypsinogens-1, -2 and -3 were produced in *Escherichia coli* and activated to trypsin-1, -2, and -3 as described previously.[39] Enzyme inhibitory activity of varlaxins was determined in 96-well plates by adding 10 μL of varlaxins diluted in ultra-purified H_2O (containing up to 0.027% DMSO), 15 μL of 0.1% bovine serum albumin in 50 mM Tris-buffered saline (BSA/TBS) and 25 μL of each trypsin (1.8 nM). Ultra-purified H_2O (10 μL) was used as a control. DMSO at the concentrations used did not have any effect on trypsin activity (data not shown). The varlaxins were pre-incubated with trypsins at room temperature for 15 min before addition of 50 μL of 0.2 mM chromogenic substrate S-2222 (ChromogeniX) in ultra-purified H_2O. The tested varlaxin 1046A and 1022A concentrations, after the addition of substrate, ranged from 0.01 nM, 1076 nM to and from 0.73 nM to 7339 nM to 0.73 nM, respectively. The change of absorbance at 405 nm was followed for 15 min (VICTOR X4 Multiple plate reader, PerkinElmer). The experiment was performed three times, each with 2 replicates (n = 3). The absorbance change was determined at the phase of the substrate reaction during which the absorbance of the control reaction increased linearly over time. The absolute IC_{50} values were estimated using Quest Graph™ (https://www.aatbio.com/tools/ic50-calculator). The porcine trypsin assay was performed as previously described.[22]

Genome sequencing and gene cluster analysis

The cyanobacterial strain *Nostoc* sp. UHCC 0870 was grown for 8 days at 20–21 °C under photon irradiance of 15 μmol m^{-2} s^{-1} in 150 mL of Z8 culture medium in a 500 mL Erlenmeyer flask. Genomic DNA was isolated using phenol–chloroform extraction as previously described.[40] The quality and concentration of isolated DNA samples were assessed using an Agilent TapeStation (Agilent Technologies) and a NanoDrop 1000 spectrophotometer (Thermo Scientific), respectively. Whole-genome sequencing was performed using a combination of PacBio RSII and Illumina MiSeq. Nextera adapters were removed from Illumina data by cutadapt v1.14. PacBio data were assembled with HGAP3. Illumina data were assembled with SPAdes v3.11.1 with the "careful" option. Gap4 was used to combine assemblies and check sequence circularity. Bwa 0.7.12-r1039 was used to map Illumina reads against the combined assembly. Homopolymer errors were corrected using Pilon v1.16. The complete sequence was annotated using Prokka v1.13 and PANNZER2. All sequencing, assembly, and annotation were performed at the DNA Sequencing and Genomics laboratory, Institute of Biotechnology, University of Helsinki, Finland. The complete genome sequence of *Nostoc* sp. UHCC 0870 was deposited in the GenBank database (CP091913–CP091924). We analysed the substrate specificity of adenylation domains using NRPSpredictor2 analysis website (http://nrps.informatik.uni-tuebingen.de/). The condensation domain of VarB was analyzed with Natural Product Domain Seeker (NaPDoS) and compared with other known aeruginosin AerB.

Results

Analysis of methanol extracts of *Nostoc* sp. UHCC 0870 using UPLC-QTOF mass spectrometry led to the serendipitous discovery of two novel aeruginosins, which we named varlaxins 1046A and 1022A, after Varlaxudden on the Gulf of Finland from where the strain was isolated. Total ion current chromatogram from the methanol extract showed prominent peaks (retention time 2.07 and 2.10 min) at the peptide region (Fig. S4A†). The upper-range mass spectrum from these peaks contained two pairs of mass peaks, representing sodiated ions. The same compounds in desulfated protonated form showed the presence of a single sulfate group (Fig. S4B†). A broader-range mass spectrum showed neutral loss of methylglyceric acid (MgA)-Ile and ions matching to one of the common aeruginosin structures (Fig. S4C†). Low-mass nonspecific product ion spectrum (MS^E) from these peaks showed diagnostic ions for aeruginosin substructures as Choi, Aaep, and MgA-Ile (Fig. S4D†). Analysis of product ion spectra (MS^2) from the desulfated protonated ions *m*/*z* 967.46 and 943.46 gave preliminary structures (Fig. 1 and 2) without the exact locations of 4-hydroxyphenylacetic acid (Hpaa) residues or clarifying the nature of the hexose sugar (Fig. S5 and Table S4†).

Further analysis of strains from the UHCC culture collection led to identification of 22 varlaxin chemical variants from five strains (Table 1). Mass spectrometric analyses (Fig. S5 and Table S4†) and retention time data showed that the chemical variation originates from the following three positions: Ile or Val in position 2, Aaep or Agma in position 4, and the number and location of Hpaa residues in the structure. Two variants with identical masses, but with different retention times, were detected from varlaxin chemical variants with one and two Hpaa residues (Table 1). Two variants with three Hpaa residues were detected, but low abundance prevented further characterization of these variants. This variation was not observed when Hpaa was absent from the structure. These observations suggest that variants with identical masses differ with respect

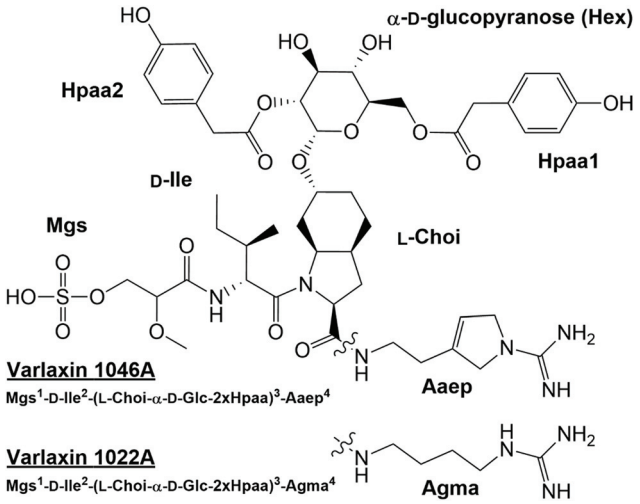

Fig. 1 Structures of varlaxins 1046A and 1022A from *Nostoc* sp. UHCC 0870. Mgs=2-*O*-methylglyceric acid 3-*O*-sulfate, Choi = 2-carboxy-6-octahydroindole, Agma = 4-amidinobutylamide, Aaep = 1-amidino-3-(2-aminoethyl)-3-pyrroline, Hpaa = 4-hydroxyphenylacetic acid.

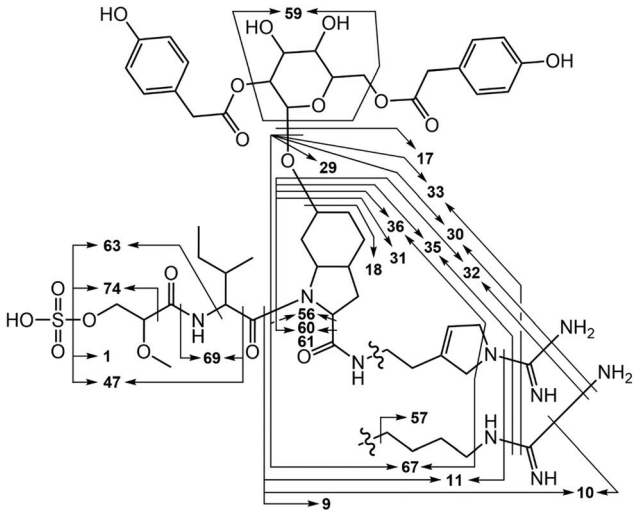

Fig. 2 Major varlaxin MS^2 fragments from desulfated variants 1046A, 1022A, 912A, 888B, 778, 764, and 754. Fragment numbers are given in Table S3.†

to the positions of the Hpaa residues in the hexose unit. Therefore, the subunit sequence for the 22 varlaxins can be summarized as Mgs[1]-Ile/Val[2]-(Choi-Glc-(0–3 Hpaa))[3]-Agma/Aaep[4] (Table 1). In all strains, Ile and Aaep dominated in positions 2 and 4, respectively, and in three strains variants with two Hpaa residues dominated. The Choi moiety was decorated with a hexose esterified with 0–3 Hpaa residues. UV spectra of varlaxin 1046A and 1022A near their maximum of 278 nm matched closely with spectrum of reference Hpaa, which is consistent with the existence of Hpaa substructure in varlaxins (Fig. S6†).

The structures of purified varlaxins 1046A and 1022A were completed and verified with 1D ^{1}H and ^{13}C spectra and 2D DQF-COSY, TOCSY, NOESY, ^{1}H–^{1}C HSQC, ^{1}H–^{15}N HSQC, and HMBC NMR spectra (Fig. S7, S8 and Tables S5, S6†). Two sets of $\delta_{C/H}$ values were recognized from both 1046A and 1022A data originating from *cis*/*trans* Ile-Choi rotamers in a 4 : 6 ratio (Table S5†). DQF-COSY correlations and HMBC/NOESY correlations were used for subunit connections (Fig. 3). HMBC correlations connected the two Hpaa residues to hexose positions 2 and 6 (Fig. 3). The hexose connected to Choi-OH was analysed from the hydrolysis mixture with NMR and LC-MS to be D-glucose (Fig. S8 and S9†). The configuration of the two amino acids in the varlaxin substructure were analysed with LC-MS of the Marfey's derivatives of purified varlaxin 1046A and 1022A acid hydrolysates. Comparison to commercial amino acid reference Ile isomers and Choi from *Microcystis aeruginosa* NIES-298 raw extract showed that Ile was in D-configuration and Choi was in the L-configuration in both varlaxins, which is the same configuration as Choi from *Microcystis aeruginosa* NIES 298 (Fig. S10†).

We sequenced the complete 7.17 Mb genome of *Nostoc* sp. UHCC 0870, which was organized in a 6.42 Mb chromosome, 10 plasmids, and a plasmid prophage. Bioinformatic analysis identified the putative 36 kb varlaxin (*var*) biosynthetic gene cluster, which contained 22 ORFs of which 10 shared similarity to gene products residing in other aeruginosin biosynthetic gene clusters (Table 2 and Fig. 4). VarB and VarG encode two bimodular NRPS enzymes, which form the varlaxin peptide backbone (Fig. 4). The first module of VarB consists of an *O*-methylglyceric acid transferase (OMT), a FkbH domain, peptidal carrier protein (PCP) and sulfotransferase (ST) (Fig. 4). The second module consists of condensation (C), adenylation (A), PCP and epimerization (E) domains (Fig. 4). The first module of this NRPS likely incorporate glycerate, which is further methylated and sulfated (Fig. 4), the second module loads allo-isoleucine. A loading module with an identical set of catalytic domains from the suomilide and dysinosin biosynthetic pathways is also proposed to synthesize Mgs.[23,30] The VarB second adenylation domain binding pocket Stachelhaus sequence is DALWMGGVFK, which predicts, with a moderate 80% match, Val in the second position (Table S7†). However, according to chemical analysis by Marfey's method, this amino acid in varlaxins 1022A and 1046A was D-Ile. Biosynthetic gene clusters for ulleungmycin and desotamide cyclic peptides contain identical adenylation domain binding pocket sequence DALWMGGVFK in genes $UlmA_4$ and $DsaH_4$ without following epimerase.[41,42] This resulted in incorporation and existence of L-*allo*-Ile in these peptides. Notably, the second domain of VarB contains an epimerase domain. When the substrate of this domain is L-*allo*-Ile, the C_2 epimerization results in D-Ile, which is found in varlaxins 1022A and 1046A. Based on these three cases, the Stachelhaus sequence DALWMGGVFK is more specific for L-*allo*-Ile than for Val. The binding pocket of second adenylation domain of AerB from aeruginosins 98-A, -B, and -C is DAFFLGVTFK, of which amino acid prediction is Ile with strong 100% match (Tables S7 and S28†). Thus, the action of these two VarB second domains resulted in the incorporporation of D-Ile or D-*allo*-Ile in varlaxins or aeruginosins 98-A, -B, and -C, as adenylation domains recognize different amino acids (L-*allo*-Ile or L-Ile, respectively), and C_2 epimerization in both cases generate the final Ile configurations (Fig. 4 and Table S7†). VarG, consists of C, A, PCP, C, and PCP domains, which load the Choi and arginine to the preceding moiety (Fig. 4). VarI glycosylates the Choi hydroxyl group. VarO is an adenylation (A) domain that could recognize the Hpaa substrate, which would be bound to acyl carrier protein (ACP) VarP. VarQ belongs to the MBOAT enzyme family, which could catalyse the addition of Hpaa to the glucose (Fig. 4).

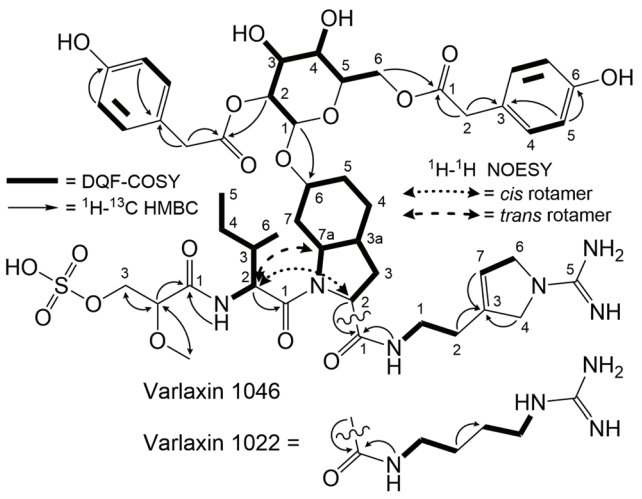

Fig. 3 DQF-COSY, HMBC, and NOESY correlations of varlaxins 1022A and 1046A.

The trypsin inhibitory activity of varlaxins 1046A and 1022A was tested using both porcine and human trypsins. Inhibition of porcine trypsin was very effective with both varlaxin variants, with IC_{50} values of 14 nM and 740 nM, respectively (Fig. 5). This makes the varlaxin 1046A the most potent porcine trypsin inhibitor among the reported aeruginosins (Table S2†). As varlaxins are strong porcine trypsin inhibitors, the three human trypsin isoenzymes were also tested for inhibition. Both varlaxin variants inhibited all human trypsin isoenzymes (Fig. 5). Varlaxin 1046A inhibited human trypsin-2 and -3 with an IC_{50} values of 0.62 nM and 0.96 nM, respectively (Fig. 5), while trypsin-1 was inhibited at higher concentrations (IC_{50} 3.6 nM).

Table 2 Predicted varlaxin (var) biosynthetic gene cluster gene products

Protein	Accession number	Predicted function	Length (aa)	BLASTp top hit			
				Description	Scientific name	Identity %	Accession
ORF1	UKO95746		156	Tetratricopeptide repeat protein	*Anabaena minutissima*	96	WP_190695341
ORF2	UKO95747		449	FAD-binding oxidoreductase	*Anabaena minutissima*	94	WP_190696554
VarB	UKO95748	NRPS	2853	Amino acid adenylation domain-containing protein	*Nostoc* sp. FACHB-145	85	WP_190754768
ORF3	UKO95749		274	Hypothetical protein	*Nostoc* sp. KVJ20	86	WP_069074352
ORF4	UKO95750		126	Nuclear transport factor 2 family protein	*Nostoc* sp.	84	WP_199337282
ORF5	UKO95751	Sulfo-transferase	192	Adenylyl-sulfate kinase	*Scytonema* sp. NIES-4073	90	WP_096564148
VarC	NA		720	Rieske 2Fe-2S domain-containing protein	*Nostoc* sp. B(2019)	86	WP_162397143
VarK	UKO95752		351	Type 2 isopentenyl-diphosphate Delta-isomerase	*Scytonema* sp. NIES-4073	92	WP_096564146
VarD	UKO95753	Choi formation	198	Bacilysin biosynthesis protein BacABacA	*Nostoc* sp. LEGE 12450	90	WP_193945672
VarE	UKO95754	Choi formation	235	Cupin domain-containing protein	*Scytonema* sp. NIES-4073	84	WP_096564145
VarF	UKO95755	Choi formation	264	SDR family oxidoreductase	*Nostoc* sp. B(2019)	90	WP_162397147
VarG	UKO95756	NRPS	1615	Non-ribosomal peptide synthetase	Scytonema sp. NIES-4073	88	WP_096564143
ORF6	UKO95757		267	SDR family oxidoreductase	*Nostoc* sp. '*Peltigera membranacea* cyanobiont' 213	93	WP_094328879
VarN1	UKO95758	ABC transporter	596	ABC transporter substrate-binding protein	*Anabaena* sp. 4-3	83	WP_066426904
ORF7	UKO95759		385	Hypothetical protein	*Anabaena* sp. CA = ATCC 33047	78	WP_066382146
ORF8	UKO95760		214	Hypothetical protein	*Anabaena*	84	WP_066382147
VarO	UKO95761	NRPS	541	Amino acid adenylation domain-containing protein	*Nostoc linckia*	82	WP_096542302
VarP	UKO95762	ACP	88	Acyl carrier protein	*Anabaena*	80	WP_066382151
VarQ	UKO95763	Hpaa transfer	504	MBOAT family protein	*Anabaena* sp. 4-3	89	WP_066426900
VarI	UKO95764	Glycosylation	423	Glycosyltransferase family 4 protein	Anabaena sp. 4-3	85	WP_066426897
ORF9	UKO95765		457	Major facilitator superfamily transporter	*Nostoc* sp. '*Peltigera membranacea* cyanobiont' N6	87	AVH65385
VarN2	UKO95766	ABC transporter	600	ATP-binding cassette domain-containing protein	*Anabaena* sp. CA = ATCC 33047	83	WP_066382156
ORF10	UKO95767		294	Hypothetical protein	*Anabaena minutissima*	89	WP_190700187
ORF11	UKO95768		273	3-Deoxy-7-phosphoheptulonate synthase	*Anabaena minutissima*	97	WP_190700184

Discussion

We report here the identification and structural characterization of novel members of the aeruginosin family of natural products, varlaxin 1046A and 1022A. All identified varlaxin variants contain Choi at the third amino acid, a moiety which is characteristic for most of the aeruginosins, except for subclasses such as pseudoaeruginosins, suomilide, banyasides and nostosins (Table 1 and Table S1†). Thus, we propose that varlaxins are members of the aeruginosin family of natural compounds. The Choi moiety of the varlaxins reported here is decorated with a sugar, a decoration previously identified in about one quarter of the aeruginosins that contain a Choi moiety (Table S1†). The varlaxin α-D-glucopyranose moiety is esterified by 0–3 carboxylic acids (Table 1 and Fig. 1, 4). Approximately half of the other aeruginosin sugar units are decorated with carboxylic or sulfate groups.[32] However, the Hpaa residue of varlaxins is part of the Choi-glucose subunit, making this substructure markedly different and unique among aeruginosins. This unique sugar decoration has been reported in plant sesquiterpene lactones, where glucose is decorated with three Hpaa residues.[43] The N-terminus moiety of varlaxin is Mgs, which is rare among aeruginosins (Table 1 and Table S1†); only suomilide, banyasides, and four dysinosins contain this moiety.[22,24,32] While the second moiety of varlaxins is primarily *allo*-Ile (identified from variants 1046A and 1022A, but could be isobaric Leu in other variants), Val was also found (Table 1). Both are nonpolar amino acids, which are the typical second amino acids in aeruginosins (Table S1†). The fourth substructure unit of varlaxins is either Aaep or Agma, of which Aaep is rarer than Agma in aeruginosins (Table S1†). In summary, the Mgs and Aaep structure and the Choi-Glu-Hpaa moiety make varlaxins unique amongst the aeruginosin family of natural products and justifies naming these compounds as a separate aeruginosin subclass.

The 36 kb varlaxin biosynthetic gene cluster is similar to previously reported aeruginosin biosynthetic gene clusters.[23,26–31] Varlaxin biosynthesis begins with the incorporation of Mgs, catalyzed by a FkbH domain, which tailors *O*-methylglyceric acid (OMT) and sulfotransferase (ST)

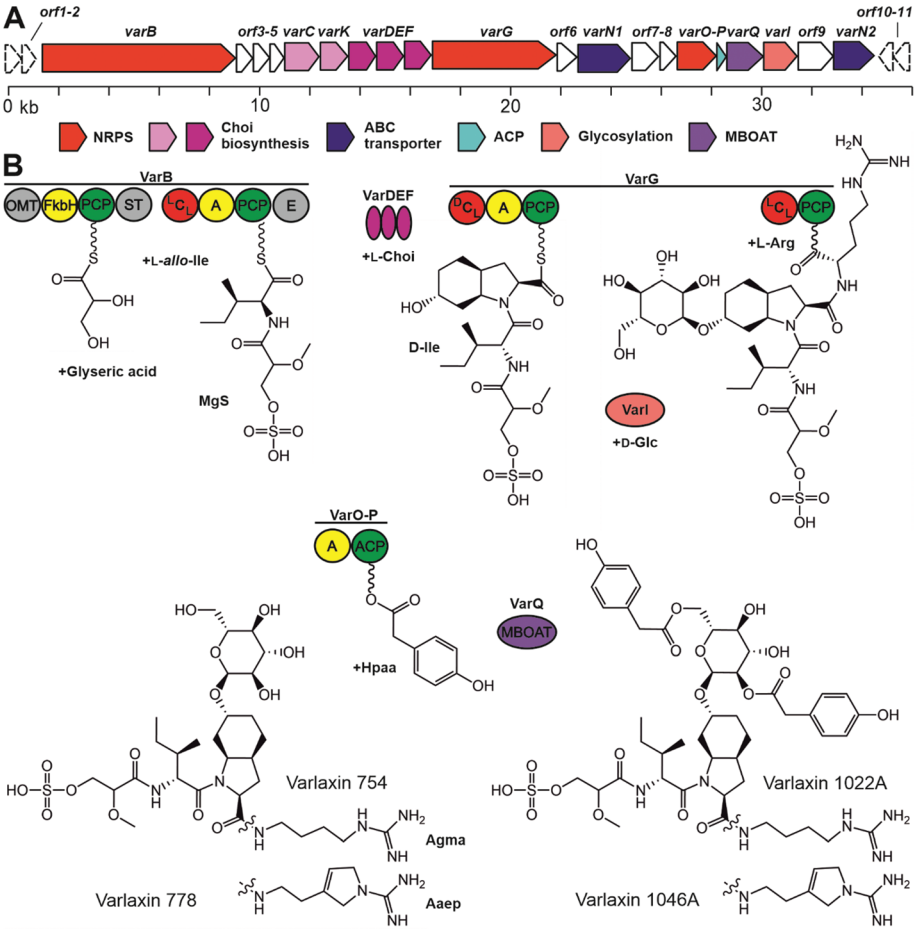

Fig. 4 The varlaxin (*var*) biosynthetic gene cluster and putative biosynthetic scheme in *Nostoc* sp. UHCC 0870. A – Organization of predicted biosynthetic genes. B – Proposed biosynthetic pathway of varlaxin 1046A. FkbH (domain of unknown function), OMT (*O*-methylglyceric acid transferase); PCP (peptidyl carrier protein); ST (sulfotransferase); C (condensation domain); A (adenylation domain); E (epimerization domain); and MBOAT (membrane bound *O*-acyl transferase).

domains (Fig. 5). A similar organization of Mgs-incorporating domains is seen in the biosynthesis of cyanopeptolin 1138,[48] dysinosin[30] and suomilide.[23] Phormidolide and oscillapeptin G both contain similar glyceric acid units that are incorporated by FkbH domains.[49,50] FkbH domains use phosphoglycerate as their substrate.[51] The predicted substrate for sulfotransferase is 3′-phosphoadenylyl sulfate, produced by ORF5 (adenylylsulfate kinase). The varlaxin biosynthetic gene cluster encodes the three Choi-forming enzymes, AerDEF (VarDEF in varlaxin BGC).[27] The varlaxin biosynthetic gene cluster also includes both AerC (VarC) and AerK (VarK), which are predicted to participate in Choi formation. However, Choi-containing aeruginosin NAL2 is produced without either of these enzymes.[29] No study to date has provided a reasonable model for formation of Aaep and this is likely to involve novel biochemistry. However, when comparing aeruginosin biosynthetic gene clusters and their products, AerC (VarC) can be found exclusively in biosynthetic gene clusters that produce aeruginosins with Aaep.[25] AerH has been predicted to act in Aaep formation.[27] However, no AerH homolog was found in the varlaxin biosynthetic gene cluster (Fig. 5). The varlaxin biosynthetic gene clusters encodes three unique proteins, VarO-Q, that are not found in other aeruginosin biosynthetic gene clusters and predicted here to attach Hpaa residues to the glucose (Fig. 5). MBOAT enzymes have also been shown to act on saccharides.[52] The timing of individual biosynthetic steps is unclear but glycosylation of Choi may occur prior to release of the final product because it appears in all of the varlaxins detected. The fact that there are mono-, di- and tri-Hpaa-ester derivatives suggest that this step occurs subsequent to release from VarG. Varlaxin 1046A is produced in two-fold higher abundance than varlaxin 1022A, indicating that varlaxin 1046A is the main product of the varlaxin biosynthetic gene cluster (Table 1). The other varlaxin chemical variants are likely to be side products from unspecific substrate specificities of the VarB adenylation domains and incomplete action of the possible tailoring reactions.

The aeruginosins have frequently been observed to inhibit serine proteases.[18] Varlaxin 1046A appears to be the most potent cyanobacterial trypsin inhibitor reported and is con-

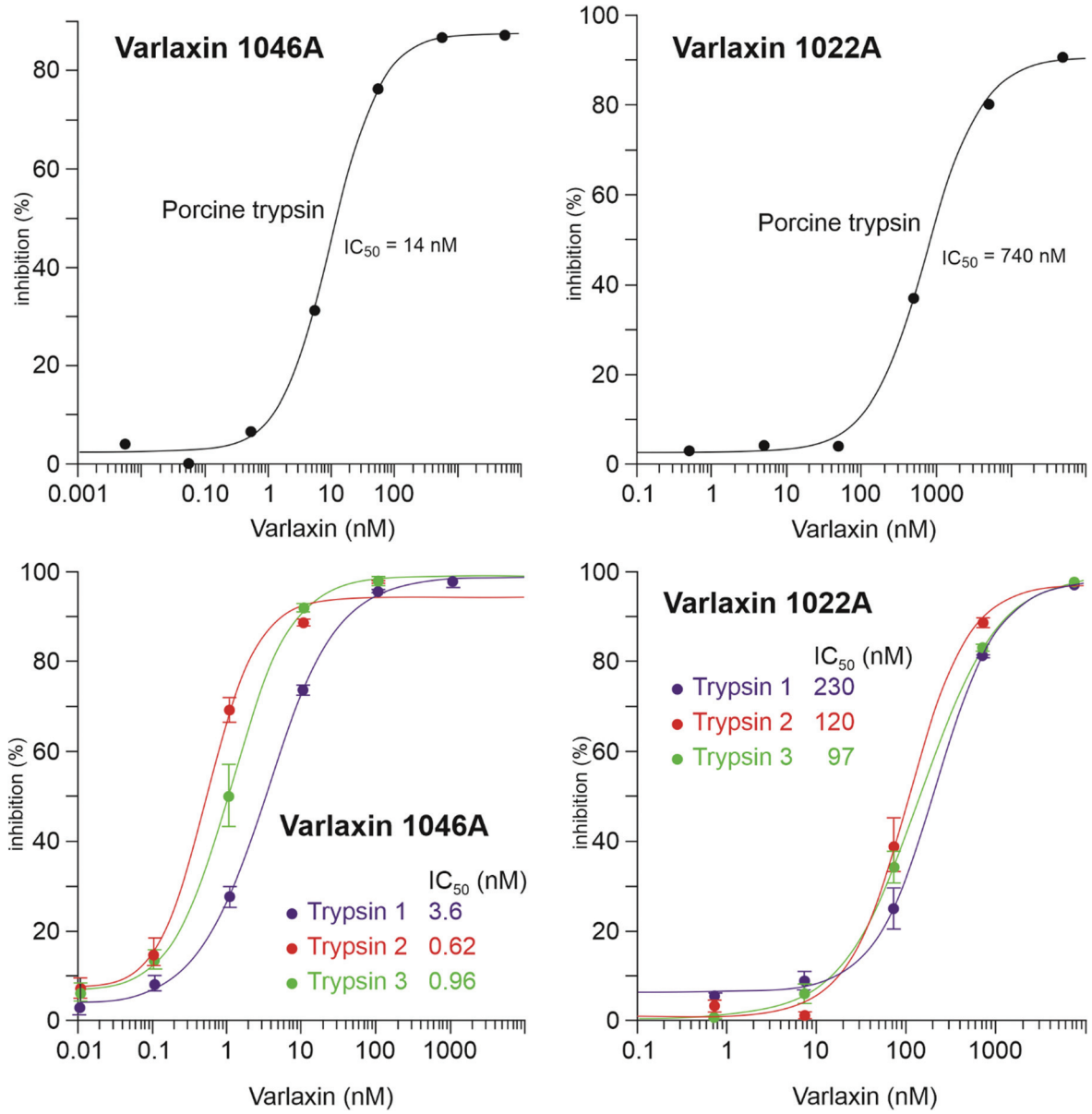

Fig. 5 Dose response curves and the IC_{50} values of varlaxins 1046A and 1022A on porcine trypsin and three human trypsin-isoenzymes.

siderably more potent than commercial peptidic inhibitors leupeptin and antipain (Fig. 5 and Table S2†). The high potency of varlaxin 1046A in inhibiting trypsin extends to human trypsin isoenzymes. Furthermore, varlaxins exhibit similar inhibition profile as suomilide, inhibiting trypsin-2 and -3 more potently than trypsin-1 (23, Fig. 5). Significantly, varlaxin 1046A was approximately 50 to 200 times more potent than varlaxin 1022A in inhibiting the trypsin isoenzymes (Fig. 5). The only difference between these two varlaxins is the fourth moiety, which is Aaep in varlaxin 1046A and Agma in varlaxin 1022A. In previous studies, the presence of Aaep was shown to impact the trypsin-inhibiting ability of aeruginosins, as studied using natural and synthetic aeruginosin analogues.[40,44] Furthermore, molecular-dynamics simulation has shown that the Aaep moiety of suomilide is important for interaction with trypsins.[23] Some other potent trypsin inhibitors (*i.e.*, aeruginosins 205A and 205B) have Agma in that position,[44] which is consistent with our results that Agma-containing varlaxin 1022A is also a potent human trypsin inhibitor, although not as potent as varlaxin 1046A with Aaep. Crystal structures of trypsin-aeruginosin 98-B and thrombin-aeruginosin 298-A complexes have been reported.[45,46] In these studies, the arginine derivative is recognized in an active site and a guanidinyl group derivative of Aaep appears to bind more strongly than other Arg derivatives, again suggesting the importance of Aaep for the inhibitory activity. However, the other three moieties (hydroxy acid, hydrophobic amino acid and Choi) in the aeruginosins may also interact with the cata-

lytic site-surrounded surface (Table S2†). It should be noted that the aeruginosins in the studies above were studied with undefined trypsin,[44–48] which is most likely of non-human origin as in most of such studies. Since our results indicate that there are significant differences in the potency of varlaxins inhibiting highly similar human trypsins, the results reported in the literature with different trypsins should be interpreted with caution. Furthermore, it would be important to describe the trypsin preparations used in more detail.

Conclusion

We report here the structural elucidation and activity of two novel linear peptides belonging to the aeruginosin family, varlaxins 1046A and 1022A. Their classification as members of the aeruginosin family is based on their linear nature and the presence of moieties such as Choi, Agma, Aaep, and Mgs, where Choi and Agma are shared by many aeruginosins. Bioinformatic analysis of the varlaxin biosynthetic gene cluster confirms that varlaxins are members of the aeruginosin family of natural products and share a common biosynthetic pathway. Varlaxin 1046A was found to be the most potent reported trypsin inhibitor among the aeruginosins. Importantly, varlaxins also inhibited human trypsins. Thus, to the best of our knowledge, varlaxin is the second cyanobacterial trypsin inhibitor tested with human trypsin isoenzymes after suomilide.[23] Inhibition of human trypsin-3 may be particularly significant, as trypsin-3 is a potential drug target for metastatic cancers.[8–10,53] As there are currently no specific small-molecule inhibitors for trypsin-3, varlaxin 1046A, as a sub-nanomolar trypsin inhibitor showing preference for trypsin-2 and -3 inhibition, is a potential lead molecule for drug development. While we call for testing of other cyanobacterial trypsin inhibitors against human trypsin isoenzymes, the varlaxin variants reported here may facilitate the development of more selective trypsin-3 inhibitors for clinical applications.

Author contributions

LH, KS, and DF designed the study. MW, MNA, PP, KS, PH, and LH performed the experiments. LH, HK, JJ, MNA, KS, PH, and DF analysed and interpreted the data. LH, MNA, JJ, HK, DF, and KS wrote the manuscript, which was corrected, revised, and approved by all authors.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We would like to thank Annikki Löfhjelm and Lyudmila Saari for their excellent technical assistance. We would also like to acknowledge support from the Sigrid Jusélius Foundation, the Magnus Ehrnrooth Foundation, the Jane and Aatos Erkko Foundation, the Nordforsk Nordic centre of Excellency NordAqua (project number #82845), and the University of Helsinki's Doctoral Programme in Microbiology and Biotechnology funding to L. P. H. and M. N. A. We thank the Language Centre of the University of Helsinki for their assistance in language revisions.

References

1 A. W. Lambert, D. R. Pattabiraman and R. A. Weinberg, Emerging Biological Principles of Metastasis, *Cell*, 2017, **168**(4), 670–691.
2 B. Turk, Targeting proteases: successes, failures and future prospects, *Nat. Rev. Drug Discovery*, 2006, **5**(9), 785–799.
3 L. Sevenich and J. A. Joyce, Pericellular proteolysis in cancer, *Genes Dev.*, 2014, **28**(21), 2331–2347.
4 P. Nyberg, M. Ylipalosaari, T. Sorsa and T. Salo, Trypsins and their role in carcinoma growth, *Exp. Cell Res.*, 2006, **312**(8), 1219–1228.
5 A. Paju and U. H. Stenman, Biochemistry and clinical role of trypsinogens and pancreatic secretory trypsin inhibitor, *Crit. Rev. Clin. Lab. Sci.*, 2006, **43**(2), 103–142.
6 C. N. M. Nyaruhucha, M. Kito and S. I. Fukuoka, Identification and expression of the cDNA-encoding human mesotrypsin(ogen), an isoform of trypsin with inhibitor resistance, *J. Biol. Chem.*, 1997, **272**(16), 10573–10578.
7 M. Sahin-Tóth, Human mesotrypsin defies natural trypsin inhibitors, From passive resistance to active destruction, *Protein Pept. Lett.*, 2005, **12**(5), 457.
8 A. Hockla, D. C. Radisky and E. S. Radisky, Mesotrypsin promotes malignant growth of breast cancer cells through shedding of CD109, *Breast Cancer Res. Treat.*, 2010, **124**(1), 27.
9 G. Jiang, F. Cao, G. Ren, D. Gao, V. Bhakta, Y. Zhang, *et al.*, PRSS3 promotes tumour growth and metastasis of human pancreatic cancer, *Gut*, 2010, **59**(11), 1535–1544.
10 A. Hockla, E. Miller, M. A. Salameh, J. A. Copland, D. C. Radisky and E. S. Radisky, PRSS3/mesotrypsin is a therapeutic target for metastatic prostate cancer, *Mol. Cancer Res.*, 2012, **10**(12), 1555–1566.
11 C. Rolland-Fourcade, A. Denadai-Souza, C. Cirillo, C. Lopez, J. O. Jaramillo, C. Desormeaux, *et al.*, Epithelial expression and function of trypsin-3 in irritable bowel syndrome, *Gut*, 2017, **66**(10), 1767–1778.
12 J. Oiva, O. Itkonen, R. Koistinen, K. Hotakainen, W. M. Zhang, E. Kemppainen, *et al.*, Specific Immunoassay Reveals Increased Serum Trypsinogen 3 in Acute Pancreatitis, *Clin. Chem.*, 2011, **57**(11), 1506–1513.
13 G. Katona, G. I. Berglund, J. Hajdu, L. Gráf and L. Szilágyi, Crystal structure reveals basis for the inhibitor resistance of human brain trypsin, *J. Mol. Biol.*, 2002, **315**(5), 1209–1218.

14 P. Wu, J. Weisell, M. Pakkala, M. Peräkyla, L. Zhu, R. Koistinen, *et al.*, Identification of novel peptide inhibitors for human trypsins, *Biol. Chem.*, 2010, **391**(2–3), 283–293.

15 M. A. Salameh, A. S. Soares, A. Hockla, D. C. Radisky and E. S. Radisky, The P(2)' residue is a key determinant of mesotrypsin specificity: engineering a high-affinity inhibitor with anticancer activity, *Biochem. J.*, 2011, **440**(1), 95–105.

16 O. Kayode, Z. Huang, A. S. Soares, T. R. Caulfield, Z. Dong, A. M. Bode, *et al.*, Small molecule inhibitors of mesotrypsin from a structure-based docking screen, *PLoS One*, 2017, **12**(5), e0176694.

17 O. Schilling, M. L. Biniossek, B. Mayer, B. Elsässer, H. Brandstetter, P. Goettig, *et al.*, Specificity profiling of human trypsin-isoenzymes, *Biol. Chem.*, 2018, **399**(9), 997–1007.

18 K. Ersmark, J. R. del Valle and S. Hanessian, Chemistry and biology of the aeruginosin family of serine protease inhibitors, *Angew. Chem., Int. Ed.*, 2008, **47**(7), 1202–1223.

19 M. Murakami, Y. Okita, H. Matsuda, T. Okino and K. Yamaguchi, Aeruginosin 298-A, a thrombin and trypsin inhibitor from the blue-green alga *Microcystis aeruginosa* (NIES-298), *Tetrahedron Lett.*, 1994, **35**(19), 3129–3132.

20 L. Liu, A. Budnjo, J. Jokela, B. E. Haug, D. P. Fewer, M. Wahlsten, *et al.*, Pseudoaeruginosins, nonribosomal peptides in *nodularia spumigena*, *ACS Chem. Biol.*, 2015, **10**(3), 725–733.

21 L. Liu, J. Jokela, M. Wahlsten, B. Nowruzi, P. Permi, Y. Z. Zhang, *et al.*, Nostosins, trypsin inhibitors isolated from the terrestrial cyanobacterium Nostoc sp. strain FSN, *J. Nat. Prod.*, 2014, **77**(8), 1784–1790.

22 K. Fujii, K. Sivonen, K. Adachi, K. Noguchi, Y. Shimizu, H. Sano, *et al.*, Comparative study of toxic and non-toxic cyanobacterial products: A novel glycoside, suomilide, from non-toxic *Nodularia spumigena* HKVV, *Tetrahedron Lett.*, 1997, **38**(31), 5529–5532.

23 M. N. Ahmed, M. Wahlsten, J. Jokela, M. Nees, U. H. Stenman, D. O. Alvarenga, *et al.*, Potent Inhibitor of Human Trypsins from the Aeruginosin Family of Natural Products, *ACS Chem. Biol.*, 2021, **16**(11), 2537–2546.

24 A. Pluotno and S. Carmeli, Banyasin A and banyasides A and B, three novel modified peptides from a water bloom of the cyanobacterium Nostoc sp., *Tetrahedron*, 2005, **61**(3), 575–583.

25 A. Kapuścik, P. Hrouzek, M. Kuzma, S. Bártová, P. Novák, J. Jokela, *et al.*, Novel Aeruginosin-865 from Nostoc sp. as a potent anti-inflammatory agent, *ChemBioChem*, 2013, **14**(17), 2329–2337.

26 L. Frangeul, P. Quillardet, A. M. Castets, J. F. Humbert, H. C. P. Matthijs, D. Cortez, *et al.*, Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium, *BMC Genomics*, 2008, **9**, 274.

27 K. Ishida, G. Christiansen, W. Y. Yoshida, R. Kurmayer, M. Welker, N. Valls, *et al.*, Biosynthesis and structure of aeruginoside 126A and 126B, cyanobacterial peptide glycosides bearing a 2-carboxy-6-hydroxyoctahydroindole moiety, *Chem. Biol.*, 2007, **14**(5), 565–576.

28 K. Ishida, M. Welker, G. Christiansen, S. Cadel-Six, C. Bouchier, E. Dittmann, *et al.*, Plasticity and evolution of aeruginosin biosynthesis in cyanobacteria, *Appl. Environ. Microbiol.*, 2009, **75**(7), 2017–2026.

29 D. P. Fewer, J. Jokela, E. Paukku, J. Österholm, M. Wahlsten, P. Permi, *et al.*, New structural variants of aeruginosin produced by the toxic bloom forming cyanobacterium *Nodularia spumigen*a, *PLoS One*, 2013, **8**(9), e73618.

30 M. A. Schorn, P. A. Jordan, S. Podell, J. M. Blanton, V. Agarwal, J. S. Biggs, *et al.*, Comparative Genomics of Cyanobacterial Symbionts Reveals Distinct, Specialized Metabolism in Tropical Dysideidae Sponges, *mBio*, 2019, **10**(3), e00821-19.

31 D. S. May, C. M. Crnkovic, A. Krunic, T. A. Wilson, J. R. Fuchs and J. E. Orjala, 15 N Stable Isotope Labeling and Comparative Metabolomics Facilitates Genome Mining in Cultured Cyanobacteria, *ACS Chem. Biol.*, 2020, **15**(3), 758–765.

32 M. R. Jones, E. Pinto, M. A. Torres, F. Dörr, H. Mazur-Marzec, K. Szubert, *et al.*, CyanoMetDB, a comprehensive public database of secondary metabolites from cyanobacteria, *Water Res.*, 2021, **196**, 117017.

33 A. Calteau, D. P. Fewer, A. Latifi, T. Coursin, T. Laurent, J. Jokela, *et al.*, Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria, *BMC Genomics*, 2014, **15**(1), 977.

34 E. Dittmann, M. Gugger, K. Sivonen and D. P. Fewer, Natural Product Biosynthetic Diversity and Comparative Genomics of the Cyanobacteria, *Trends Microbiol.*, 2015, **23**(10), 642–652.

35 J. C. Roach, K. Wang, L. Gan and L. Hood, The molecular evolution of the vertebrate trypsinogens, *J. Mol. Evol.*, 1997, **45**(6), 640–652.

36 J. Kotai, *Instructions for preparation of modified nutrient solution Z8 for algae, Blindern B–11/69*, Norwegian Institute for Water Research, Oslo, 1972.

37 Y. H. Wang, B. Avula, X. Fu, M. Wang and I. A. Khan, Simultaneous determination of the absolute configuration of twelve monosaccharide enantiomers from natural products in a single injection by a UPLC-UV/MS method, *Planta Med.*, 2012, **78**(8), 834–837.

38 J. L. Giner, J. Feng and D. J. Kiemle, NMR Tube Degradation Method for Sugar Analysis of Glycosides, *J. Nat. Prod.*, 2016, **79**(9), 2413–2417.

39 H. Koistinen, R. Koistinen, W. M. Zhang, L. Valmu and U. H. Stenman, Nexin-1 inhibits the activity of human brain trypsin, *Neuroscience*, 2009, **160**(1), 97–102.

40 D. P. Fewer, J. Österholm, L. Rouhiainen, J. Jokela, M. Wahlsten and K. Sivonen, Nostophycin biosynthesis is directed by a hybrid polyketide synthase-nonribosomal peptide synthetase in the toxic cyanobacterium Nostoc sp. strain 152, *Appl. Environ. Microbiol.*, 2011, **77**(22), 8034–8040.

41 S. Son, Y. S. Hong, M. Jang, K. T. Heo, B. Lee, J. P. Jang, *et al.*, Genomics-Driven Discovery of Chlorinated Cyclic Hexapeptides Ulleungmycins A and B from a Streptomyces Species, *J. Nat. Prod.*, 2017, **80**(11), 3025–3031.

42 Q. Li, Y. Song, X. Qin, X. Zhang, A. Sun and J. Ju, Identification of the Biosynthetic Gene Cluster for the Anti-infective Desotamides and Production of a New Analogue in a Heterologous Host, *J. Nat. Prod.*, 2015, **78**(4), 944–948.

43 R. A. Sessa, M. H. Bennett, M. J. Lewis, J. W. Mansfield and M. H. Beale, Metabolite profiling of sesquiterpene lactones from Lactuca species, Major latex components are novel oxalate and sulfate conjugates of lactucin and its derivatives, *J. Biol. Chem.*, 2000, **275**(35), 26877–26884.

44 H. J. Shin, H. Matsuda, M. Murakami and K. Yamaguchi, Aeruginosins 205A and -B, serine protease inhibitory glycopeptides from the cyanobacterium *Oscillatoria agardhii* (NIES-205), *J. Org. Chem.*, 1997, **62**(6), 1810–1813.

45 B. Sandler, M. Murakami and J. Clardy, Atomic structure of the trypsin – Aeruginosin 98-B complex, *J. Am. Chem. Soc.*, 1998, **120**(3), 595–596.

46 J. L. R. Steiner, M. Murakami and A. Tulinsky, Structure of thrombin inhibited by aeruginosin 298-A from a blue-green alga, *J. Am. Chem. Soc.*, 1998, **120**(3), 597–598.

47 G. Wang, N. Goyal and B. Hopkinson, Preparation of l-proline based aeruginosin 298-A analogs: Optimization of the P1-moiety, *Bioorg. Med. Chem. Lett.*, 2009, **19**(14), 3798–3803.

48 T. B. Rounge, T. Rohrlack, A. Tooming-Klunderud, T. Kristensen and K. S. Jakobsen, Comparison of cyanopeptolin genes in *Planktothrix*, *Microcystis*, and *Anabaena* strains: evidence for independent evolution within each genus, *Appl. Environ. Microbiol.*, 2007, **73**(22), 7322–7330.

49 T. B. Rounge, T. Rohrlack, A. J. Nederbragt, T. Kristensen and K. S. Jakobsen, A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix rubescens* strain, *BMC Genomics*, 2009, **10**, 396.

50 M. J. Bertin, A. Vulpanovici, E. A. Monroe, A. Korobeynikov, D. H. Sherman, L. Gerwick, *et al.*, The Phormidolide Biosynthetic Gene Cluster: A trans-AT PKS Pathway Encoding a Toxic Macrocyclic Polyketide, *ChemBioChem*, 2016, **17**(2), 164–173.

51 Y. Sun, H. Hong, F. Gillies, J. B. Spencer and P. F. Leadlay, Glyceryl-S-acyl carrier protein as an intermediate in the biosynthesis of tetronate antibiotics, *ChemBioChem*, 2008, **9**(1), 150–156.

52 D. Sychantha, D. J. Little, R. N. Chapman, G. J. Boons, H. Robinson, P. L. Howell, *et al.*, PatB1 is an O-acetyltransferase that decorates secondary cell wall polysaccharides, *Nat. Chem. Biol.*, 2018, **14**(1), 79–85.

53 H. Ma, A. Hockla, C. Mehner, M. Coban, N. Papo, D. C. Radisky, *et al.*, PRSS3/Mesotrypsin and kallikrein-related peptidase 5 are associated with poor prognosis and contribute to tumor cell invasion and growth in lung adenocarcinoma, *Sci. Rep.*, 2019, **9**(1), 1844.