

## Nostosins, Trypsin Inhibitors Isolated from the Terrestrial Cyanobacterium *Nostoc* sp. Strain FSN

Liwei Liu,<sup>†</sup> Jouni Jokela,<sup>†</sup> Matti Wahlsten,<sup>†</sup> Bahareh Nowruzi,<sup>‡</sup> Perttu Permi,<sup>§</sup> Yue Zhou Zhang,<sup>⊥</sup> Henri Xhaard,<sup>⊥</sup> David P. Fewer,<sup>†</sup> and Kaarina Sivonen<sup>\*,†</sup>

<sup>†</sup>Division of Microbiology and Biotechnology, Department of Food and Environmental Sciences, University of Helsinki, P.O. Box 56, 00014, Helsinki, Finland

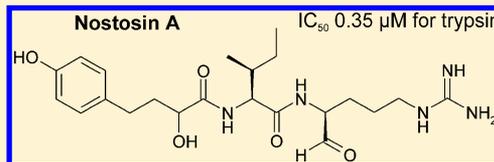
<sup>‡</sup>Department of Biology, Faculty of Science, Tarbiat Moallem University, 49 Dr. Mofatteh Avenue, P.O. Box 158153587, 15614, Tehran, Iran

<sup>§</sup>Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, P.O. Box 65, 00014, Helsinki, Finland

<sup>⊥</sup>Division of Pharmaceutical Chemistry and Technology, Faculty of Pharmacy, University of Helsinki, P.O. Box 56, 00014, Helsinki, Finland

### Supporting Information

**ABSTRACT:** Two new trypsin inhibitors, nostosin A (1) and B (2), were isolated from a hydrophilic extract of *Nostoc* sp. strain FSN, which was collected from a paddy field in the Golestan Province of Iran. Nostosins A (1) and B (2) are composed of three subunits, 2-hydroxy-4-(4-hydroxyphenyl)-butanoic acid (Hhpba), L-Ile, and L-argininal (1) or argininol (2). Nostosins A (1) and B (2) exhibited IC<sub>50</sub> values of 0.35 and 55 μM against porcine trypsin, respectively, suggesting that the argininal aldehyde group plays a crucial role in the efficient inhibition of trypsin. Molecular docking of nostosin A (1) (449 Da), leupeptin (426 Da, IC<sub>50</sub> 0.5 μM), and spumigin E (610 Da, IC<sub>50</sub> < 0.1 μM) with trypsin suggested prominent binding similarity between nostosin A (1) and leupeptin but only partial binding similarity with spumigin E. The number of hydrogen bonds between ligands and trypsin increased according to the length and size of the ligand molecule, and the docking affinity values followed the measured IC<sub>50</sub> values. Nostosin A (1) is the first highly potent three-subunit trypsin inhibitor with potency comparable to the known commercial trypsin inhibitor leupeptin. These findings expand the known diversity of short-chain linear peptide protease inhibitors produced by cyanobacteria.



Cyanobacteria are a prolific source of natural products and produce a range of linear and cyclic peptide inhibitors of serine proteases including aeruginosins,<sup>1,2</sup> spumigins,<sup>3</sup> banyasins,<sup>4</sup> cyanopeptolins,<sup>5</sup> micropeptins,<sup>6</sup> anabaenopeptins,<sup>6</sup> kempeptins,<sup>7</sup> microginins,<sup>8</sup> and microviridins.<sup>9</sup> Aeruginosins are a chemically diverse family of serine protease inhibitors composed of four subunits, derivatives of 4-hydroxyphenyllactic acid, a hydrophobic D-amino acid, 2-carboxy-6-hydroxyoctahydroindole (Choi), and a modified arginine.<sup>1,2,10</sup> Spumigins are structurally similar to aeruginosins, except Choi is replaced by (2S,4S)-4-methylproline or L-proline and subunit variability of spumigins is markedly less than in aeruginosins.<sup>3,11,12</sup> Many of these protease inhibitors are assembled on large nonribosomal peptide synthetase enzyme complexes.<sup>2,3</sup> Proteases play an important role in a multitude of biological processes from simple proteolysis to the degradation of important regulators of the major cellular pathways. Novel protease inhibitors from nature are important drug leads. Aeruginosin 98-B was used as a lead structure to design new thrombin inhibitors, which can be used in the treatment of thrombosis.<sup>13</sup> *Nostoc*, a genus of nitrogen-fixing cyanobacteria found in a number of habitats around the world, is known to produce a variety of bioactive peptides, including compounds such as banyasins,<sup>3</sup> micro-

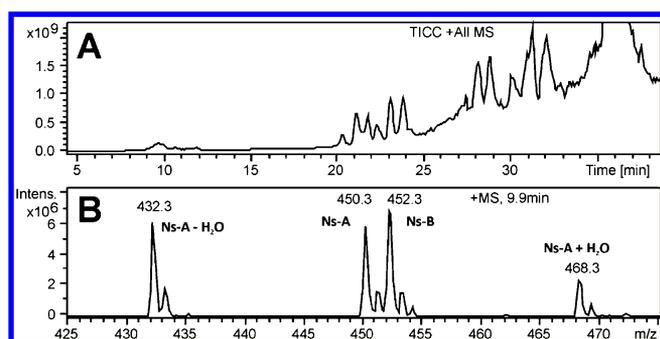
cystins,<sup>14</sup> microviridins,<sup>15</sup> nostopeptins,<sup>16</sup> and nostoginins.<sup>16</sup> Here, we describe the isolation, structure, and bioactivity of new trypsin inhibitors nostosin A (1) and nostosin B (2) isolated from a *Nostoc* strain.

### RESULTS AND DISCUSSION

*Nostoc* sp. strain FSN was isolated from a sample collected in a paddy field in the Golestan Province of Iran in 2003.<sup>17</sup> We analyzed a methanol extract of freeze-dried biomass with LC-MS in order to gain knowledge of the secondary metabolites produced by this strain. The total and extracted ion chromatograms and mass spectrum showed a pair of compounds, which were named nostosins A (Ns-A) and B (Ns-B) with masses *m/z* 450.3 (1) and *m/z* 452.3 (2) (Figure 1, Figure S1). Four additional compounds (*m/z* 434 and 436, *m/z* 436 and 438) with minor intensities were also identified from LC-MS analysis (Figure S1). The MS<sup>2</sup> fragmentation patterns indicated that the smaller ion (*m/z* 450) could contain argininal (neutral fragment 158 Da; Figure S2A) and the two mass units larger ion argininol (160 Da; Figure S2B). The MS<sup>2</sup>

Received: February 4, 2014

Published: July 28, 2014



**Figure 1.** Total positive ion-current chromatogram (TICC) and mass spectrum (+MS) from *Nostoc* sp. FSN. (A) TICC from *Nostoc* sp. FSN. (B) MS of positive ions from TICC at 9.9 min. Nostosins A (1) and B (2) protonated molecule ions, dehydrated 1, and water hydrate of 1 with their ion masses are marked in the spectrum.

spectrum from ion  $m/z$  450.3 (1) did not yield any further structural information. The spectrum of the  $m/z$  452.3 (2) suggested the presence of leucine or isoleucine (Figure S2B,D).

A culture of *Nostoc* sp. strain FSN was labeled with <sup>15</sup>N to gain further information about the subunit structures. By comparing the LC-MS results from labeled and unlabeled extracts containing 1 and 2, a shift in mass of five daltons was observed (Figure S3). This implied 1 and 2 contained five nitrogen atoms, consistent with the presence of Ile or Leu (one nitrogen atom) and argininal/argininol (four nitrogen atoms, Figure S3G). LC-MS analysis of Marfey derivatized acid hydrolysates of 1 and 2 demonstrated that they contained Ile in the L configuration (Figure S4A and B). Oxidation of 1 prior to acid hydrolysis and Marfey derivatization produced mainly L-Arg, showing that the native argininal of nostosin A (1) has an L configuration (Figure S4C). Leupeptin, which is reported to contain L-argininal (Table S1), gave an identical result.

The total structure of nostosins A (1) and B (2) were solved by NMR analysis. <sup>1</sup>H NMR, <sup>1</sup>H–<sup>1</sup>H DQF-COSY, <sup>1</sup>H–<sup>1</sup>H TOCSY, <sup>13</sup>C and <sup>15</sup>N gHSQC, and <sup>13</sup>C gHMBC spectra confirmed the presence of three substructures: 2-hydroxy-4-(4-hydroxyphenyl)butanoic acid (Hhpba), Ile, and Arg-CHO/OH (argininal/ol). The NMR data are presented in Table 1.

**Table 1.** <sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N NMR Spectral Data for Nostosins A (1) and B (2) in [D<sub>6</sub>] DMSO<sup>d</sup>

unit	nostosin A (1)					nostosin B (2)				
	C/H no.	$\delta_{C/N}$	$\delta_H$ , mult., J (Hz)	COSY	HMBC <sup>b</sup>	$\delta_{C/N}$	$\delta_H$ , mult., J (Hz)	COSY	HMBC <sup>b</sup>	
Hhpba	1	173.3				174.0				
	2	70.1	3.87, dd (4.0, 8.9)	3, 3'	–	70.1	3.88, m	2-OH, 3, 3'	1, 3, 4	
	3	36.3	1.65	2, 3', 4	4, 5	36.8	1.64, m	2, 3', 4	1, 2, 4, 5	
	3'		1.84	2, 3, 4	2		1.83, m	2, 3, 4	1, 2, 4, 5	
	4	29.8	2.50	3, 3'	2, 3, 5, 6/10	29.7	2.49	3, 3'	2, 3, 5, 6/10	
	5	131.7				132.2				
	6/10	128.9	6.95, d (8.4)	7/9	4, 6/10, 8	129.4	6.94, d (8.4)	7/9	4, 7/9, 6/10, 8	
	7/9	114.9	6.65, d (8.4)	6/10	5, 7/9, 8	115.3	6.65, d (8.4)	6/10	5, 7/9, 8	
	8	155.2				155.7				
	2-OH		5.69, brs	–	–		5.69, d (5.5)	2	1, 2, 3	
	8-OH		9.17, brs	–	–		9.17, s	–	7/9, 8	
Ile	1	170.4				171.2				
	2	55.9	4.27, dd (7.0, 9.5)	NH, 3	–	56.5	4.15, dd (9.0)	3	1, 3, 4, 6; Hhpba-1	
	3	37.2	1.70	2, 4, 6	–	37.2	1.69, m	2, 4, (4'), 6	–	
	4	24.1	1.03	3, 4', 5	–	24.5	1.03, m	3, 4', 5	3, 5, 6	
	4'		1.43	4, 5	–		1.42, m	3, 4, 5	5, 6	
	5	10.9	0.81	4, 4'	2	10.9	0.81	4, 4'	3, 4	
	6	15.1	0.82, d (6.2)	3	3, 4	15.1	0.82, d (7.0)	3	2, 3, 4	
	NH	113.2	7.60, d (9.5)	2	–	114.3	7.58, d (9.0)	2	2; Hhpba-1	
Arg <sup>c</sup>	1	76.0	5.28	2	–	63.2	3.21	2	2, 3 <sup>d</sup>	
	1'				–		3.35	– <sup>e</sup>	2, 3 <sup>d</sup>	
	2	48.9	3.75/3.77	1, 2-NH, (3), 3'	–	50.3	3.69	1, 1', 2-NH, 3, 3'	1; Ile-1 <sup>d</sup>	
	3	23.7	1.52	(2), 4'	–	27.9	1.28	2, 3', (4, 4')	1, 2, 4, 5 <sup>d</sup>	
	3'		1.69	2, 4	–		1.58	2, 3, (4, 4')	1, 2, 5 <sup>d</sup>	
	4	23.5	1.51	4', 5, (5')	4	25.3	1.40	3, 3', 5	2, 5 <sup>d</sup>	
	4'		1.70	4, (5), (5')	–		1.46	3', 5	2, 5 <sup>d</sup>	
	5	39.1	3.13	4, (4'), 5'	–	39.8	3.05	4, 4', 5-NH	3, 4, 6 <sup>d</sup>	
	5'		3.45	4, 5	–		–	–	–	
	6	–	–	–	–	156.9 <sup>d</sup>	–	–	–	
		1-OH	–	–	–	–	4.71, dd	1, 1'	1, 2	
	2-NH	124.4	7.88/7.98, d (8.4)	2	Ile-1	125.4	7.83, d (8.4)	2	2; Ile-1	
	5-NH	115.8	7.43, s	–	–	102.3 <sup>f</sup>	7.67, brs	5	–	

<sup>a</sup>Hhpba = 2-hydroxy-4-(4-hydroxyphenyl)butanoic acid, – = not detected, brs = broad singlet, () = weak signal. Nitrogen shifts match NH<sub>3</sub> reference. <sup>b</sup>HMBC correlations, optimized for 8 Hz, are from the proton(s) stated to the indicated carbon. <sup>c</sup>Argininal in nostosin A and argininol in nostosin B. <sup>d</sup>Present when optimized for 5 Hz. <sup>e</sup>Overlap with water signal. <sup>f</sup><sup>15</sup>N–<sup>1</sup>H correlation was shown aliased in <sup>15</sup>N HSQC spectrum due to finite <sup>15</sup>N decoupling field (low frequency 110.3 ppm), and hence the 118.3 ppm spectrum signal was corrected with the equation 110.3 – (118.3 – 110.3)

The overlapping elution of **1** and **2** made it difficult to obtain enough purified nostosin A (**1**) for NMR analysis. Therefore, some of the correlations are missing from the NMR data (Table 1). The nostosin A (**1**) NMR spectra are presented in Figures S5–S11. Almost identical NMR signals were recognized for the Hhpba and Ile in **1** and **2**.  $\delta_{\text{H}}$  and  $\delta_{\text{C}}$  signals of the argininal substructure were highly similar to those reported earlier and showed that argininal preferred the hemiaminal structure (Figure S12).<sup>10,18</sup>

A comprehensive NMR data set was obtained from nostosin B (**2**) in order to determine the complete structure. Proton signals for the *p*-oxygen-substituted phenyl ring were observed at  $\delta_{\text{H}}$  6.94 and 6.65 (Figure S13) and the corresponding characteristic carbons at  $\delta_{\text{C}}$  129.4 and 115.3 (Figure S14B). These methines and an aromatic oxygen-bearing carbon at  $\delta_{\text{C}}$  155.7 (C8) together with the hydroxyl proton at  $\delta_{\text{H}}$  9.17 and another quaternary carbon at  $\delta_{\text{C}}$  132.2 were connected by HMBC to form a hydroxyphenyl unit (Figure S15). COSY and HSQC correlations built up a carbon chain from C2 to C4 so that an HMBC carbonyl carbon signal at  $\delta_{\text{C}}$  172.8 and a proton signal at  $\delta_{\text{H}}$  5.69 (2-OH) were correlated with carbon C2 ( $\delta_{\text{H}}$  70.1). HMBC signals connected this structure to the hydroxyphenyl unit, completing the Hhpba substructure (Figures S14–S17). In the LC-MS analysis, a compound of the Marfey derivatized acid hydrolysate of **2** coeluted and had an identical mass (*m/z* 449) with the Marfey derivatized commercial Hhpba, proving the presence of Hhpba in **2**.

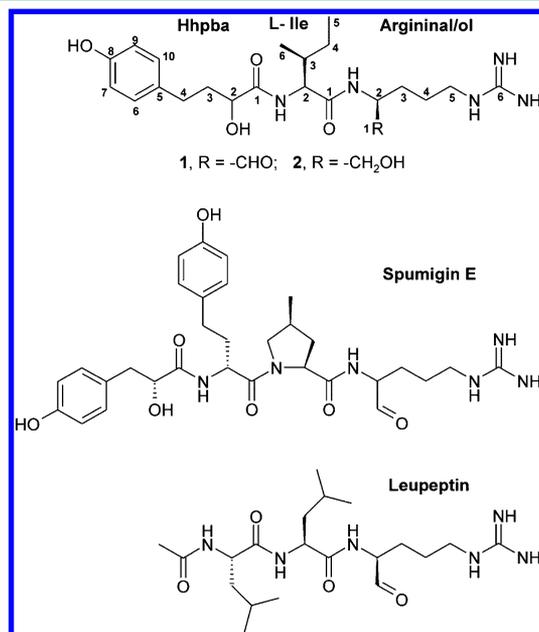
An amide proton  $\delta_{\text{H}}$  7.58 (NH) in the Ile substructure showed a COSY correlation to a typical alpha proton  $\delta_{\text{H}}$  4.15 (H2) (Figure S18), which showed an HMBC correlation to the carbonyl carbon  $\delta_{\text{C}}$  171.2 (CO<sup>Ile</sup>) and to another carbonyl carbon  $\delta_{\text{C}}$  174.0 (CO<sup>Hhpba</sup>), which links Ile to the Hhpba subunit (Figures S15, S17). H3<sup>Ile</sup> ( $\delta_{\text{H}}$  1.69) is vicinal to H6<sup>Ile</sup> ( $\delta_{\text{H}}$  0.82) and H4<sup>Ile</sup>, and 4'<sup>Ile</sup> ( $\delta_{\text{H}}$  1.03, 1.42) is vicinal to H5<sup>Ile</sup> ( $\delta_{\text{H}}$  0.81). All other connections observable from the COSY and HSQC spectra (Figure S19) clearly demonstrated that the second substructure was Ile.

Argininal (Arg-OH) could be recognized from the HMBC of amide proton  $\delta_{\text{H}}$  7.83 (2-NH) with CO<sup>Ile</sup> ( $\delta_{\text{C}}$  171.2) (Figures S15, S17). COSY, TOCSY, and <sup>15</sup>N HSQC spectra (Figure S18) showed that proton 2-NH belongs to a spin system of protons from H1,1' to 5-NH, and the COSY spectrum demonstrated the order of connections (Figure S20). HMBC correlations to guanidine quaternary carbon C6 were absent with <sup>n</sup>J<sub>CH</sub> values of 8 Hz, but a correlation from H5 to C6 with a typical  $\delta_{\text{C}}$  of 156.9 was seen with <sup>n</sup>J<sub>CH</sub> values of 5 Hz. Altogether these assignments showed that the third substructure was argininal. In conclusion, nostosin B (**2**) was composed of three substructure, Hhpba, L-Ile, and argininal. Elemental compositions of **1** (C<sub>22</sub>H<sub>35</sub>N<sub>5</sub>O<sub>5</sub>) and **2** (C<sub>22</sub>H<sub>37</sub>N<sub>5</sub>O<sub>5</sub>) obtained from the accurate mass measurements (Table S2) were in agreement with the aforementioned results.

Structures of the minor nostosin variants C, D, E, and F were evaluated from the mass spectrometric data after the chemical structures of **1** and **2** were solved (Table S2, Figure S2, and Table S3). The minor variants D and F were otherwise similar to **1** and **2**, but Val was present in position 2 instead of L-Ile. The chemical structure of nostosins suggests that they are produced by nonribosomal peptide synthesis. The production of minor peptide variants by the nonribosomal peptide synthetase enzymes is the result of incorporation of closely related amino acids in the end-product peptide and is a commonly observed phenomenon.<sup>19</sup> The neutral fragment

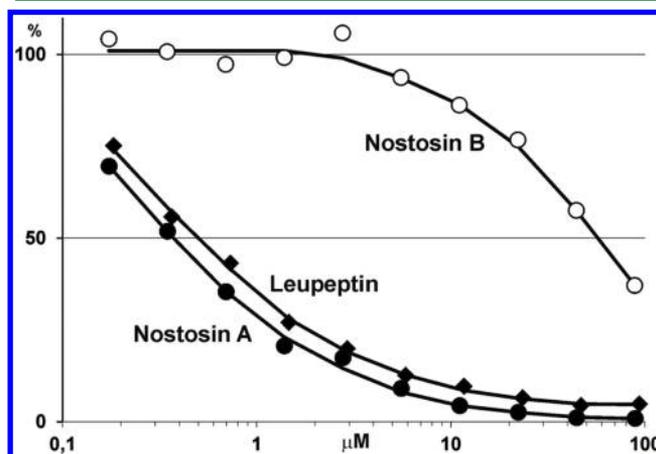
mass of Hhpba in position 1 is 178 Da (Figure S2B,F). However, in variants C and E the neutral fragment mass of subunit 1 was 162 Da (Figure S2D), 16 Da less than in Hhpba, suggesting that variants C and E contained deoxyHhpba (doHhpba). The observed retention times of the variants are in good agreement with these structures (Table S2).

The structures of nostosins (which contain Arg derivatives with a guanidino group being central for the binding of short linear inhibitors to the trypsin active site) suggested that they could be serine protease inhibitors. A trypsin inhibition assay showed that nostosin A (**1**) inhibited trypsin at a submicromolar level with an IC<sub>50</sub> value of 0.35  $\mu\text{M}$  (within a range of 0.2–0.8  $\mu\text{M}$ ) (Figure 3), while the IC<sub>50</sub> of nostosin B

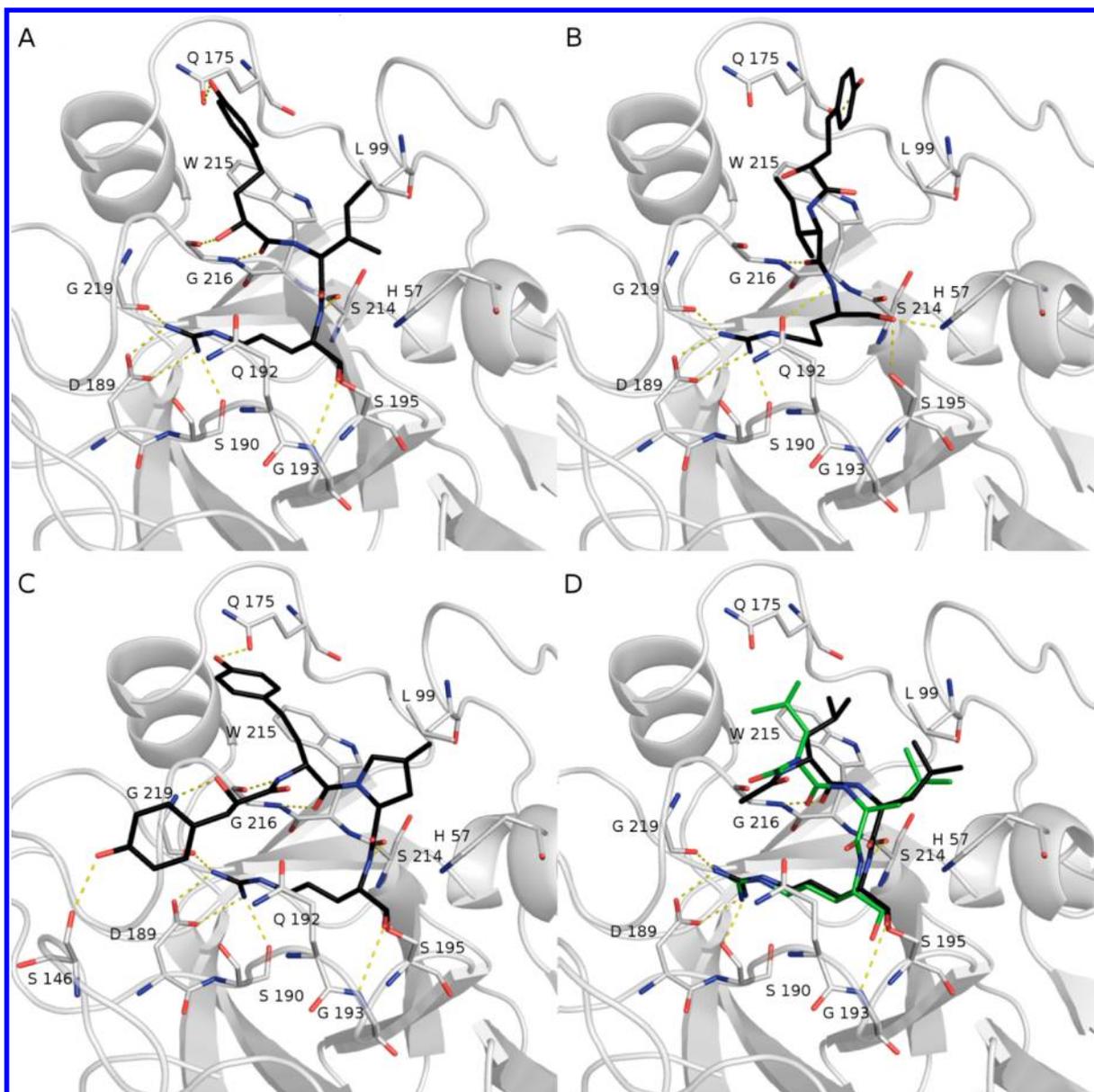


**Figure 2.** Structures of nostosin A (**1**), nostosin B (**2**), spumigin E, and leupeptin. The structure of spumigin E is taken from Fujii et al. and Fewer et al.<sup>9,2</sup> Hhpba = 2-hydroxy-4-(4-hydroxyphenyl)butanoic acid.

(**2**) was 55  $\mu\text{M}$ . Leupeptin, a commercial protease inhibitor discovered from actinomycetes and composed of acetic acid, two L-leucines, and L-argininal, was used as a positive control in



**Figure 3.** Dose-dependent trypsin inhibition curves of nostosin A (**1**), nostosin B (**2**), and leupeptin.



**Figure 4.** Binding poses of nostosins A (**1**) and B (**2**), spumigin E, and leupeptin at trypsin active site. (A) Most favorable binding orientation of **1** (D-L-L configuration) in binding site. (B) Most favorable binding orientation of **2** (D-L-L configuration) in binding site. (C) Most favorable binding orientation of spumigin E in binding site. (D) Most favorable binding orientation of leupeptin in binding site (carbon atom labeled with black) and extracted from the X-ray crystal structure with trypsin (pdb code: 1jrs; carbon atom labeled with green).

the assay.<sup>20</sup> Trypsin inhibition by leupeptin, with an  $IC_{50}$  of  $0.5 \mu\text{M}$  (within range of  $0.2\text{--}1 \mu\text{M}$  in this study;  $>0.2 \mu\text{M}$  with different substrates<sup>20</sup>), was at the same level as nostosin A (**1**). Leupeptin is a fast, tight-binding, and covalent complex forming inhibitor,<sup>21</sup> which provoked linear progress curves in an inhibition assay. **1** (and **2**) acted in a time-dependent manner, evidenced by the initially convex, but later linear progress curves (Figure S21). **1** behaves as a classical, reversible, and slow-binding protease inhibitor. The aldehyde-containing **1** had a different inhibition mechanism compared to the one subunit longer aldehydic spumigin E, which shows strong time-dependent inhibition of trypsin, indicating irreversible binding (Figure S22).<sup>3</sup>

The binding of nostosins A (**1**) and B (**2**) to the trypsin active site compared to other short linear peptidic trypsin inhibitors was evaluated with docking experiments. We were

especially interested to see how the small size (only three residues, Tables S1 and S4) of nostosins affects the binding. Covalent docking was set up for **1**, spumigin E, and leupeptin, all of which can covalently bind to the catalytic serine S195 with their aldehyde group. This covalent bond has been shown between leupeptin and trypsin.<sup>21</sup> Noncovalent “traditional” docking simulations were performed for **2**, which contains an alcohol group instead of an aldehyde group and therefore is not able to bind covalently. A covalent docking protocol was tested with leupeptin, and the binding pose close to the crystallographic pose of leupeptin in trypsin (PDB code 1JRS; Figure 4D) was successfully recreated, proving the validity of the docking method. Following this control experiment, the covalent docking protocol was applied to **1**. The stereochemistry of L-Ile and L-argininal of nostosin A (**1**) was experimentally proven, and the L-argininal configuration of

nostosin B (**2**) was used since probably both argininal and argininal have the same configuration. Since NMR was not able to discriminate the *L*- and *D*-Hhpba configurations in **1** and **2**, both configurations were included in the simulation. As a result, a range of putative high-scoring poses were found. Nostosin A (**1**) binding poses for *L*-*L*-*L* (labeled with cyan) and *D*-*L*-*L* (labeled with black) configurations are shown in Figure S23A. These poses were highly similar except the orientation of Hhpba 2-hydroxy and hydroxyphenyl groups, which vary even across one simulation run. The result that the nostosin A (**1**) binding pose *D*-*L*-*L* was better than the *L*-*L*-*L* configuration is more predictable because of the 39 trypsin inhibitors presented in Table S1, 27 contained the first residue in *D* configuration and only seven in *L* configuration. For nostosin B (**2**) highly similar docking scores of  $-9.932$ ,  $-9.914$ , and  $-9.560$  for the *L*-*L*-*L* configuration and  $-9.958$ ,  $-8.349$ , and  $-7.228$  for the *D*-*L*-*L* configuration were recorded (Figure S23B). Thus, this ensemble of poses suggests that both the *L*-Hhpba-*L*-Ile-*L*-argininol and *D*-Hhpba-*L*-Ile-*L*-argininol configurations of nostosin B (**2**) fit equally well.

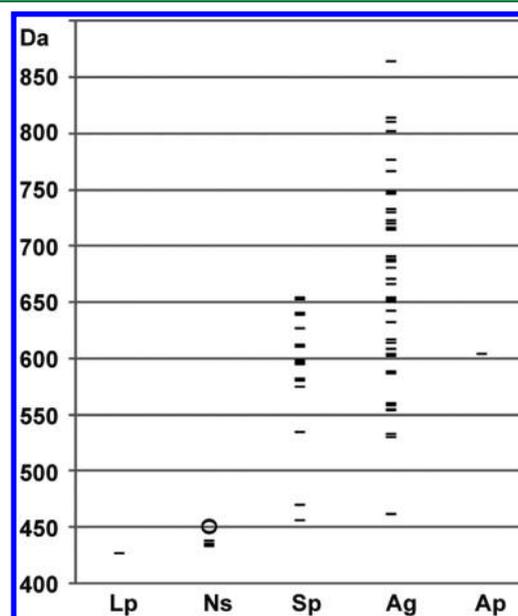
The docking poses of **1** and **2** presented in Figure 4A and B show the best fitting interactions with trypsin. The major interaction concerns the ionic interaction of the nostosin arginine derivative with G219, D189, and S190, where, in addition to ion-pairing to the carboxylate group of D189, the NH1 and NH2 of the guanidium group make charge-reinforced hydrogen bonds with the protein main-chain oxygen of S190 and G219. This interaction takes place in the so-called S1 or selectivity pocket and is a highly conserved feature of all trypsins. The aldehyde group of **1** most likely forms a covalent bond to the side chain oxygen of serine 195 of the catalytic triad (Figure 4A). The resulting hydroxyl reduced from the aldehyde carbonyl group formed a hydrogen bond with the backbone nitrogen of G193. In contrast, **2** contains an alcohol group (Figure 4B) and was free to explore the protein cavity space, resulting in a hydrogen bond with the side chain oxygen of S195.

The covalent bond between catalytic S195 of trypsin and **1** shifted compared to **2** toward the S1 pocket. This seems to have direct consequences on several interactions. The peptide bond amide nitrogen of the *L*-argininal residue of **1** acts as a hydrogen bond donor to the S214 backbone oxygen, but in **2**, the oxygen of the G192 residue acts as a hydrogen bond acceptor. In **1**, the *L*-Ile side chain is sandwiched between hydrophobic residues L99 and W215, which form a hydrophobic pocket for the amino acid. However, in **2**, an *L*-Ile side chain is positioned away from the hydrophobic pocket. In **1**, the 2-hydroxy group of Hhpba (when docked in the energetically favored *D* configuration) points to the G216 main chain oxygen, which leads to the formation of a hydrogen bond (Figure 4A). Another hydrogen bond is made between the hydroxyphenyl oxygen of Hhpba (in *D* configuration) and the Q175 side chain oxygen. In **2**, the same interaction between Hhpba (in *D* configuration) and G216 is absent, and the hydroxyphenyl oxygen of Hhpba hydrogen bonds to the Q175 backbone oxygen.

In summary, nostosin A (**1**) in *D*-*L*-*L* configuration binds to trypsin in a similar manner to leupeptin (Figure 4C and F), a result that fits with  $IC_{50}$  values of these peptides. Nostosin A (**1**) binding to trypsin was also compared to the one subunit larger spumigin E (Sp-E). Binding of *L*-argininal and hydrophobic amino acids *L*-Ile in **1** and *L*-argininal and (2*S*,4*S*)-4-methylproline in Sp-E was similar (Figure 4A and C). *D*-Hty of

Sp-E took up the position of Hhpba, forming similar hydrogen bonds. The extra subunit *D*-Hpla in Sp-E formed two extra hydrogen bonds compared to **1**. These two hydrogen bonds may account for the tight binding ( $IC_{50} < 0.1 \mu M$ ) of Sp-E to trypsin. The docking affinity value  $-8.315$  showed the best binding for Sp-E,  $-7.498$  for **1**, and  $-6.692$  for leupeptin, which followed the order of the  $IC_{50}$  values (Table S1).

Several families of small linear peptides that inhibit the serine protease trypsin have been described from cyanobacteria (e.g., aeruginosins, spumigins, nostosins) and actinomycetes (such as leupeptin and antipain) (Table S1, Figure 5, Table S4). Most of



**Figure 5.** Comparison of nostosin to other linear peptides (size in daltons) showing the small size of the three-subunit nostosins. – = variants of the peptides, ○ = nostosins A (**1**) and B (**2**). Lp = leupeptin, Ns = nostosins, Sp = spumigins, Ag = aeruginosins, Ap = antipain. Data from Table S4.

them are composed of four subunits and contain an arginine derivative at the C-terminal (Tables S1 and S4). Nostosins are compact, low molecular weight, and short chain length (three subunits) trypsin inhibitors compared to these other linear peptide groups (Figure 5, Table S4). To date, only a few small, three-subunit trypsin inhibitors such as nostosin A (**1**) have been described from cyanobacteria (Table S1). They are aeruginosins GH553, EI461, and 298B, which all are weak trypsin inhibitors most probably because the guanidino group of the Arg derivative is absent. In the small, three-subunit spumigins the Arg derivative is also absent, probably making these peptides poor inhibitors (Table S4). The four-subunit leupeptin from actinomycetes is even smaller in length and mass than nostosin A (**1**) but acts as a good inhibitor (Table S1). So nostosin A and leupeptin are the smallest trypsin inhibitors among the peptides discussed here, and despite the significant structural difference at the amino terminal, both were equally potent inhibitors. Nostosin B (**2**) inhibits trypsin with an  $IC_{50}$  value of  $55 \mu M$  (Figure 3); this suggests that the aldehyde group of the argininal residue plays a key role in the binding of nostosin to the active center of trypsin. Likewise, the *L*-arginine aldehyde in spumigins and aeruginosins results in potent trypsin inhibition (Table S1). However, the most potent inhibitors produced by actinobacteria and cyanobacteria

contain agmatine (Agma) or 1-(*N*-amidino- $\Delta^3$ -pyrrolino)ethyl (Adc) subunits, which lack the aldehyde functionality of *L*-argininyl (Table S1). Other features for the most potent trypsin inhibitors presented in Table S1 are that the amino acid in position 3 has to be in the *L* configuration, and the amino acid in position 2 is hydrophobic but with either *L* and *D* configuration. There is always a carboxylic acid, mostly in *D* configuration in position 1 but with high structural variation.

## EXPERIMENTAL SECTION

**General Experimental Procedures.** Optical rotations were measured on an Anton Paar MCP 200 modular circular polarimeter (Anton Paar GmbH, Germany). UV spectra were recorded on a Shimadzu UV-1800 spectrophotometer (Shimadzu Corp., Kyoto, Japan). IR spectra were recorded on a Bruker Vertex 70 FTIR spectrometer (Bruker Optics, Germany) equipped with a microplate HTS-XT accessory unit using a silicon plate for the solid sample platform. NMR spectra were recorded on Varian Inova 500 or 600 NMR spectrometers equipped with a (cryogenically cooled in Inova 600)  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$  triple-resonance probe head and an actively shielded *z*-axis gradient system in [ $\text{D}_6$ ] DMSO at 298 K. Mass spectra were recorded on an Agilent 1100 Series LC/MSD TRAP system HPLC (Agilent Technologies, USA) with an XCT Plus model ion trap mass detector or on a UPLC-ESI-QTOF Synapt G2 HDMS mass spectrometer (Waters, MA, USA). HPLC purifications were performed on an HP 1100 Series modular chromatograph (Agilent Technologies, Palo Alto, CA, USA) containing a binary pump, thermostated autosampler, six-port manual injector valve (Rheodyne), column oven, and diode array detector. Trypsin inhibition was measured with an Infinite M200 spectrofluorometer (Tecan Austria GmbH).

**Culture Conditions.** *Nostoc* sp. FSN was collected from a paddy field in Golestan Province of Iran in 2003.<sup>17</sup> The strain was grown at a photon irradiance of  $15 \mu\text{mol m}^{-2} \text{s}^{-1}$  in modified Z8IX medium at 20 °C for 21 days. The cells were collected by centrifugation for 10 min at 10000g. The biomass was lyophilized and used for LC-MS analysis and purification of nostosins.

**Stable Isotope Labeling.**  $^{15}\text{N}$  labeling was performed by growing the FSN strain in growth medium containing  $^{15}\text{N}$ -urea as a nitrogen source in a nitrogen-free atmosphere maintained by bubbling the medium with nitrogen-free argon with 20.9%  $\text{O}_2$  and 0.45%  $\text{CO}_2$  (quality 5.7; AGA Gas Ab, Sweden). The FSN strain was grown under a photon irradiance of  $15 \text{ mmol m}^{-2} \text{s}^{-1}$  for 21 days at 20 °C, at which point the biomass was collected and lyophilized. A 100 mg portion of  $^{15}\text{N}$ -labeled freeze-dried cells was extracted with MeOH, and the extract was analyzed with ESI-LC-MS.

**Nostosin Purification.** A total of 0.78 mg of nostosin A (1) was purified from 6 g of freeze-dried cells. Freeze-dried cells (in 2 g batches) were extracted with 120 mL of methanol and partitioned with water and dichloromethane. The water–methanol layer was separated and used in the HPLC purification of nostosin A (1). The extract was injected into a Luna  $\text{C}_8$  column (150 × 4.6 mm, 5  $\mu\text{m}$ , Phenomenex) in 1 mL batches. The column was eluted isocratically with 19% aqueous MeCN (LiChrosolv, Merck) at 30 °C and washed with 85% aqueous MeCN between the injections. The fractions, which contained nostosin A, were pooled and evaporated. The residue was dissolved in 19% aqueous MeCN. This solution was injected into a Luna  $\text{C}_{18}$  column (150 × 4.6 mm, 3.5  $\mu\text{m}$ , Phenomenex) in 50  $\mu\text{L}$  batches. The column was eluted isocratically with 19% aqueous MeCN (LiChrosolv, Merck) at 30 °C and washed with 85% aqueous MeCN between the injections. The peaks containing nostosin A were collected from several elutions, which were then pooled and evaporated.

Nostosin B (2) was purified from 3 g of freeze-dried cells, which were extracted twice with 120 mL of methanol. The extract was partitioned with water and dichloromethane in a 1:1:1 proportion. The upper water–methanol layer was evaporated using a rotary evaporator (Supermodulyo 12K freeze-dryer). The residue was resuspended in 4

mL of methanol and incubated with 70 mg of  $\text{NaBH}_4$  for 30 min to convert nostosin A (1) into nostosin B (2). The reaction solution was evaporated, and the residue was dissolved in 12.5% acetonitrile (MeCN). HPLC (an HP 1100 Series modular chromatograph, Agilent Technologies, Palo Alto, CA, USA) was used to purify nostosin B (2). The MeCN solution was injected into a Luna  $\text{C}_8$  column (150 × 10 mm, 5 mm, Phenomenex, Torrance, CA, USA) in batches of 1 mL. The column was eluted isocratically with 12.5% aqueous MeCN (LiChrosolv, Merck) at 30 °C and washed with 85% aqueous MeCN between injections. The fractions containing nostosin B were evaporated to dryness under a stream of air. The yield of nostosin B was 5.9 mg.

**Nostosin A (1):**  $[\alpha]_{\text{D}}^{20} -9.0$  (*c* 0.3,  $\text{H}_2\text{O}$ ); UV ( $\text{H}_2\text{O}$ )  $\lambda_{\text{max}}$  (log  $\epsilon$ ) 275 (3.04) nm; IR (solid on a silicon plate)  $\nu_{\text{max}}$  3334, 2966, 1660, 1606, 1515, 1456, 1357, 1228, 1122  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR (500 MHz,  $\text{DMSO}-d_6$ ) and  $^{13}\text{C}$  NMR, see Table 1 and Supporting Information for spectra; ITMS/MS precursor ion 450  $[\text{M} + \text{H}]^+$ , 432  $[\text{M} - \text{H}_2\text{O} + \text{H}]^+$ , 390  $[\text{M} - (\text{HN}=\text{C}=\text{NH} + \text{H}_2\text{O}) + \text{H}]^+$ , 292  $[\text{Hhpba-Ile} + \text{H}]^+$ , 264  $[\text{Hhpba-Ile} - \text{CO} + \text{H}]^+$ , see Table S3; HRMS (ESI-QTOF)  $m/z$  450.2712  $[\text{M} + \text{H}]^+$  ( $\Delta$  +0.23 ppm calcd for  $\text{C}_{22}\text{H}_{36}\text{N}_5\text{O}_5$ , 450.27110).

**Nostosin B (2):**  $[\alpha]_{\text{D}}^{20} -2.9$  (*c* 0.08,  $\text{H}_2\text{O}$ ); UV ( $\text{H}_2\text{O}$ )  $\lambda_{\text{max}}$  (log  $\epsilon$ ) 202 (4.04), 220 (3.65), 276 (2.96) nm; IR (solid on a silicon plate)  $\nu_{\text{max}}$  3286, 2966, 1656, 1602, 1515, 1459, 1377, 1353, 1249, 1086  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR (600 MHz,  $\text{DMSO}-d_6$ ) and  $^{13}\text{C}$  NMR, see Table 1 and Supporting Information for spectra; ITMS/MS precursor ion 452  $[\text{M} + \text{H}]^+$ , 435  $[\text{M} - \text{NH}_3 + \text{H}]^+$ , 418  $[\text{M} - 2\text{NH}_3 + \text{H}]^+$ , 410  $[\text{M} - \text{HN}=\text{C}=\text{NH} + \text{H}]^+$ , 393  $[\text{M} - (\text{HN}=\text{C}=\text{NH} + \text{NH}_3) + \text{H}]^+$ , 274  $[\text{Ile-argininol} + \text{H}]^+$ , 264  $[\text{Hhpba-Ile} - \text{CO} + \text{H}]^+$ , 161  $[\text{argininol} + \text{H}]^+$ , 144  $[\text{argininol} - \text{NH}_3 + \text{H}]^+$ , see Table S3; HRMS (ESI-QTOF)  $m/z$  452.2866  $[\text{M} + \text{H}]^+$  ( $\Delta$  -0.32 ppm calcd for  $\text{C}_{22}\text{H}_{38}\text{N}_5\text{O}_5$ , 452.28675).

**Mass Spectrometry.** LC-MS analysis was performed on an Agilent 1100 Series LC/MSD TRAP system HPLC (Agilent Technologies) with an XCT Plus model ion trap mass detector. Extracts were injected into a Luna  $\text{C}_8$  (2) column (2 × 150 mm, 5  $\mu\text{m}$ , Phenomenex) and eluted with different gradients of 2-propanol (+ 0.1% formic acid) in 0.1% formic acid at 40 °C. Mass spectra were acquired using electrospray ionization in the positive mode. Spectra were recorded using a scan range from  $m/z$  200 to  $m/z$  1100, and  $\text{MS}^2$  spectra were recorded as averages of three spectra. High-accuracy mass of nostosins A (1) and B (2) was measured by UPLC-ESI-QTOF mass spectrometry performed on Synapt G2 HDMS (Waters, MA, USA) in high-resolution mode.

**NMR.** NMR spectra of nostosin A (1) were collected on a Varian Inova 500 NMR spectrometer equipped with a conventional triple-resonance  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$  probehead and an actively shielded *z*-axis gradient system in [ $\text{D}_6$ ] DMSO at 298 K. A two-dimensional (2D)  $^1\text{H}$  TOCSY spectrum was measured using a mixing time of 60 ms. 2D  $^{13}\text{C}$ ,  $^1\text{H}$  correlation experiments, for both aliphatic and aromatic proton–carbon correlations, were measured using  $^{13}\text{C}$ -HSQC with polarization transfer adjusted for  $^1J_{\text{CH}} = 143$  Hz (aliphatic) and  $^1J_{\text{CH}} = 160$  Hz (aromatic). 2D  $^{15}\text{N}$ -HSQC, for one-bond  $^{15}\text{N}$ ,  $^1\text{H}$  correlations, was measured using heteronuclear coherence transfer delay optimized according to  $^1J_{\text{NH}} = 93$  Hz. Long-range  $^{13}\text{C}$ ,  $^1\text{H}$  correlations were established using the  $^{13}\text{C}$ -HMBC experiment with C–H transfer delay tuned according to an  $^nJ_{\text{CH}}$  value of 6 Hz.

NMR spectra of nostosin B (2) were recorded on a Varian Inova 600 spectrometer equipped with cryogenically cooled  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$  triple resonance probe heads and actively shielded *z*-axis gradient system in [ $\text{D}_6$ ] DMSO at 298 K.  $^{13}\text{C}$ -HSQC and  $^{15}\text{N}$ -HSQC experiments were optimized for  $^1J_{\text{CH}} = 150$  Hz and  $^1J_{\text{NH}} = 95$  Hz, respectively.  $^{13}\text{C}$ -HMBC spectra was measured using two different C–H transfer delays, i.e., using  $^nJ_{\text{CH}}$  values of 5 and 8 Hz. A  $^1\text{H}$  TOCSY spectrum was measured using a mixing time of 50 ms.

**Amino Acid Analysis.** Amino acid analysis of nostosins A (1) and B (2) was carried out by dissolving *L*- and *D*-Arg, *L*- and *D*-Leu, *L*- and *D*-Ile, *L*-*allo*-Ile (Sigma-Aldrich, MO, USA), and 2-hydroxy-4-(4-hydroxyphenyl)butanoic acid (Hhpba, order no. NP-017576, AnalytiCon Discovery GmbH, Germany) in water. Native and oxidized nostosin A (1) and nostosin B (2) and oxidized leupeptin (Peptide

Institute Inc., Osaka, Japan) were hydrolyzed with 6 M HCl at 110 °C for 19 h. The HCl acid was evaporated, and the residue was resuspended in 50  $\mu$ L of water. NaHCO<sub>3</sub> (20  $\mu$ L, 1 M) was added to hydrolysates and standard solutions. L-FDLA (*N*-( $\alpha$ )-(5-fluoro-2,4-dinitrophenyl)-L-leucinamide (1%), ABCR GmbH & Co., Germany) in acetone (100  $\mu$ L) was added to native and oxidized nostosin A (1) and oxidized leupeptin hydrolysates and standard solutions, and 1% L-FDAA (*N*-( $\alpha$ )-(5-fluoro-2,4-dinitrophenyl)-L-alaninamide, ABCR GmbH & Co., Germany) in acetone (100  $\mu$ L) was added to nostosin B (2) hydrolysate and standard solutions. The reaction was terminated with 1 M HCl (20  $\mu$ L) after incubation for 1 h at 37 °C. L-FDAA amino and hydroxy acid derivatives were analyzed with ESI-LC-MS using a Luna C<sub>18</sub> column (150 by 2 mm, 5  $\mu$ m; Phenomenex), which was eluted at 0.2 mL min<sup>-1</sup> with solvent A (0.1% HCOOH) and from 20% to 60% in 45 min with acetonitrile. L-FDLA amino acid derivatives were analyzed similarly with an isocratic elution of 32% acetonitrile in 0.1% formic acid. The detection wavelength for the Marfey derivatives was 340 nm. Oxidation of nostosin A (1) and leupeptin was carried out with the method presented by Ishida et al.<sup>10</sup>

**Protease Inhibition Assay.** Trypsin (Sigma-Aldrich, Type IX-S, porcine pancreas) activity was measured on a 96-well plate at 25 °C in a reaction mixture containing 50 mM Tris/HCl buffer, pH 8.0, 0.15 M NaCl, 1 mM CaCl<sub>2</sub>, 0.1 mg mL<sup>-1</sup> of bovine serum albumin, and leupeptin or 1 or 2. Two microliters of substrate Boc-Gln-Ala-Arg-MCA (Peptide Institute, Japan) in DMSO (100  $\mu$ M in assay) was added to 193  $\mu$ L of reaction mixture. The reaction was initiated with 5  $\mu$ L of trypsin solution, and hydrolysis was followed by measuring the fluorescence of the product 7-amino-4-methylcoumarin (MCA) with an Infinite M200 spectrofluorometer (Tecan Austria GmbH) with excitation at 380 nm and emission at 460 nm.<sup>2</sup>

**Computational Docking.** Binding of nostosins A (1) and B (2), spumigin E, and leupeptin to the trypsin active site was evaluated using the docking software Glide\_XP. This software looks for low-energy poses when flexibly searching the conformational space of the ligand while the protein is kept rigid. The possible ionic species states of nostosin A (1) and B (2), leupeptin, and spumigin E at a pH from 5.0 to 9.0 were generated by LigPre in Schrödinger Suite 2013. The docking experiment was conducted using the three-dimensional structure of bovine trypsin deposited in the PDB in complex with leupeptin at 1.80 Å resolution (PDB code 1JRS). The Schrödinger suite protein preparation wizard was used to assign atom type and side chain protonation states. Protein structure was minimized with force field OPLS 2005 prior to use. Docking simulations were run using default parameters, and Glide receptor grids were generated by defining a 10 Å box localized at the centroid of leupeptin. Leupeptin, nostosin A (1), and spumigin E were docked using a covalent docking procedure to serine 195, as suggested by the experimental binding evidence and by the argininal residue common to these molecules.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

Additional information including extracted ion chromatograms and product ion spectra of nostosins; LC-MS data of <sup>15</sup>N-labeled nostosin; chromatograms of Ile and Arg stereoisomers; NMR spectra of 1 and 2; trypsin inhibition curves of 1; estimation of IC<sub>50</sub> of spumigin E for trypsin; binding poses of nostosins A (1) and B (2); small linear peptide protease inhibitors with their subunit structures and IC<sub>50</sub> values for trypsin; assignments of the ions of the product ion spectra of the protonated nostosin variants; structures, ion mass, and abundance of nostosins; monoisotopic mass and number of subunits of 83 aeruginosins and aeruginosin-like protease inhibitors. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Tel: +358-9-191 59 270. Fax: +358-9-191 59 322. E-mail: kaarina.sivonen@helsinki.fi.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors thank L. Saari for helping to isolate the axenic *Nostoc* sp. FSN strain. This work was supported by grants (118637 and 258827) from the Academy of Finland to K.S. L.L. is a matching fund student at Viikki Graduate Programme in Bioscience and Microbiology and Biotechnology and financially supported by China Scholarship Council and a grant from the Academy of Finland and University of Helsinki (118637).

## ■ REFERENCES

- Murakami, M.; Ishida, K.; Okino, T.; Okita, Y.; Matsuda, H.; Yamaguchi, K. *Tetrahedron Lett.* **1995**, *36*, 2785–2788.
- 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* (9), e73618.
- Fewer, D. P.; Jokela, J.; Rouhiainen, L.; Wahlsten, M.; Koskeniemi, K.; Stal, L. J.; Sivonen, K. *Mol. Microbiol.* **2009**, *73*, 924–937.
- Pluotno, A.; Carmeli, S. *Tetrahedron* **2005**, *61*, 575–583.
- Elert, E. V.; Oberer, L.; Merkel, P.; Huhn, T.; Blom, J. F. J. *Nat. Prod.* **2005**, *68*, 1324–1327.
- Zafirir-Ilan, E.; Carmeli, S. *Tetrahedron* **2010**, *66*, 9194–9202.
- Taori, K.; Paul, V. J.; Luesch, H. J. *Nat. Prod.* **2008**, *71*, 1625–1629.
- Reshef, V.; Carmeli, S. *Tetrahedron* **2001**, *57*, 2885–2894.
- Reshef, V.; Carmeli, S. *Tetrahedron* **2006**, *62*, 7361–7369.
- Ishida, K.; Okita, Y.; Matsuda, H.; Okino, T.; Murakami, M. *Tetrahedron* **1999**, *55*, 10971–10988.
- Fujii, K.; Sivonen, K.; Adachi, K.; Noguchi, K.; Sano, H.; Hirayama, K.; Suzukia, M.; Harada, K. I. *Tetrahedron Lett.* **1997**, *38*, 5525–5528.
- 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.
- Radau, G.; Gebel, J.; Rauh, D. *Arch. Pharm. Pharm. Med. Chem.* **2003**, *336*, 372–380.
- Kaasalainen, U.; Fewer, D. P.; Jokela, J.; Wahlsten, M.; Sivonen, K.; Rikkinen, J. *PNAS* **2012**, *109*, 5886–5891.
- Murakami, M.; Sun, Q.; Ishida, K.; Matsuda, H.; Okino, T.; Yamaguchi, K. *Phytochemistry* **1997**, *45*, 1197–1202.
- Pluotno, A.; Carmeli, S. *Tetrahedron* **2002**, *58*, 9949–9957.
- Nowruzzi, B.; Nejad, R. A. K.; Sivonen, K.; Kazemi, B.; Najafi, F.; Nejadstatti, T. *Afr. J. Agric. Res.* **2012**, *7*, 3887–3897.
- Matsuda, H.; Okino, T.; Murakami, M.; Yamaguchi, K. *Tetrahedron* **1996**, *52*, 14501–14506.
- Fewer, D. P.; Rouhiainen, L.; Jokela, J.; Wang, H.; Wahlsten, M.; Laakso, K.; Sivonen, K. *BMC Evol. Biol.* **2007**, *7*, 183–193.
- Aoyagi, T.; Takeuchi, T.; Matsuzak, A.; Kawamura, K.; Kondo, S.; Hamada, M.; Maeda, K.; Umezawa, H. *J. Antibiot.* **1969**, *22*, 283–286.
- Kurinov, I. V.; Harrison, R. W. *Protein Sci.* **1996**, *5*, 752–758.