

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Ahmed, Muhammad N.; Wahlsten, Matti; Jokela, Jouni; Nees, Matthias; Stenman, Ulf-Håkan; Alvarenga, Danilo O.; Strandin, Tomas; Sivonen, Kaarina; Poso, Antti; Permi, Perttu; Metsä-Ketelä, Mikko; Koistinen, Hannu; Fewer, David P.

**Title:** Potent Inhibitor of Human Trypsins from the Aeruginosin Family of Natural Products

**Year:** 2021

**Version:** Published version

**Copyright:** © XXXX The Authors. Published by American Chemical Society

**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

Ahmed, M. N., Wahlsten, M., Jokela, J., Nees, M., Stenman, U.-H., Alvarenga, D. O., Strandin, T., Sivonen, K., Poso, A., Permi, P., Metsä-Ketelä, M., Koistinen, H., & Fewer, D. P. (2021). Potent Inhibitor of Human Trypsins from the Aeruginosin Family of Natural Products. *ACS Chemical Biology*, 16(11), 2537-2546. <https://doi.org/10.1021/acschembio.1c00611>

# Potent Inhibitor of Human Trypsins from the Aeruginosin Family of Natural Products

Muhammad N. Ahmed, Matti Wahlsten, Jouni Jokela, Matthias Nees, Ulf-Håkan Stenman, Danilo O. Alvarenga, Tomas Strandin, Kaarina Sivonen, Antti Poso, Perttu Permi, Mikko Metsä-Ketelä, Hannu Koistinen,\* and David P. Fewer\*



Cite This: <https://doi.org/10.1021/acscchembio.1c00611>



Read Online

ACCESS |



Metrics & More

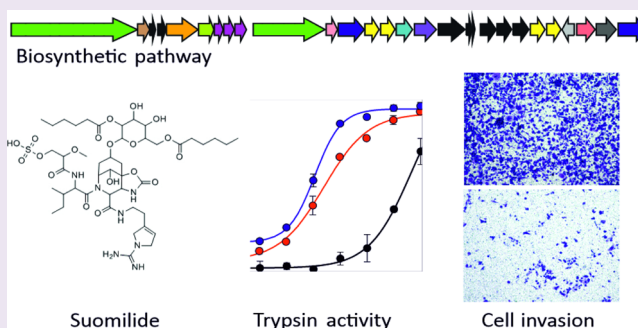


Article Recommendations



Supporting Information

**ABSTRACT:** Serine proteases regulate many physiological processes and play a key role in a variety of cancers. Aeruginosins are a family of natural products produced by cyanobacteria that exhibit pronounced structural diversity and potent serine protease inhibition. Here, we sequenced the complete genome of *Nodularia sphaerocarpa* UHCC 0038 and identified the 43.7 kb suomilide biosynthetic gene cluster. Bioinformatic analysis demonstrated that suomilide belongs to the aeruginosin family of natural products. We identified 103 complete aeruginosin biosynthetic gene clusters from 12 cyanobacterial genera and showed that they encode an unexpected chemical diversity. Surprisingly, purified suomilide inhibited human trypsin-2 and -3, with  $IC_{50}$  values of 4.7 and 11.5 nM, respectively, while trypsin-1 was inhibited with an  $IC_{50}$  of 104 nM. Molecular dynamics simulations suggested that suomilide has a long residence time when bound to trypsins. This was confirmed experimentally for trypsin-1 and -3 (residence times of 1.5 and 57 min, respectively). Suomilide also inhibited the invasion of aggressive and metastatic PC-3M prostate cancer cells without affecting cell proliferation. The potent inhibition of trypsin-3, together with a long residence time and the ability to inhibit prostate cancer cell invasion, makes suomilide an attractive drug lead for targeting cancers that overexpress trypsin-3. These results substantially broaden the genetic and chemical diversity of the aeruginosin family and suggest that aeruginosins may be a source of selective inhibitors of human serine proteases.



## INTRODUCTION

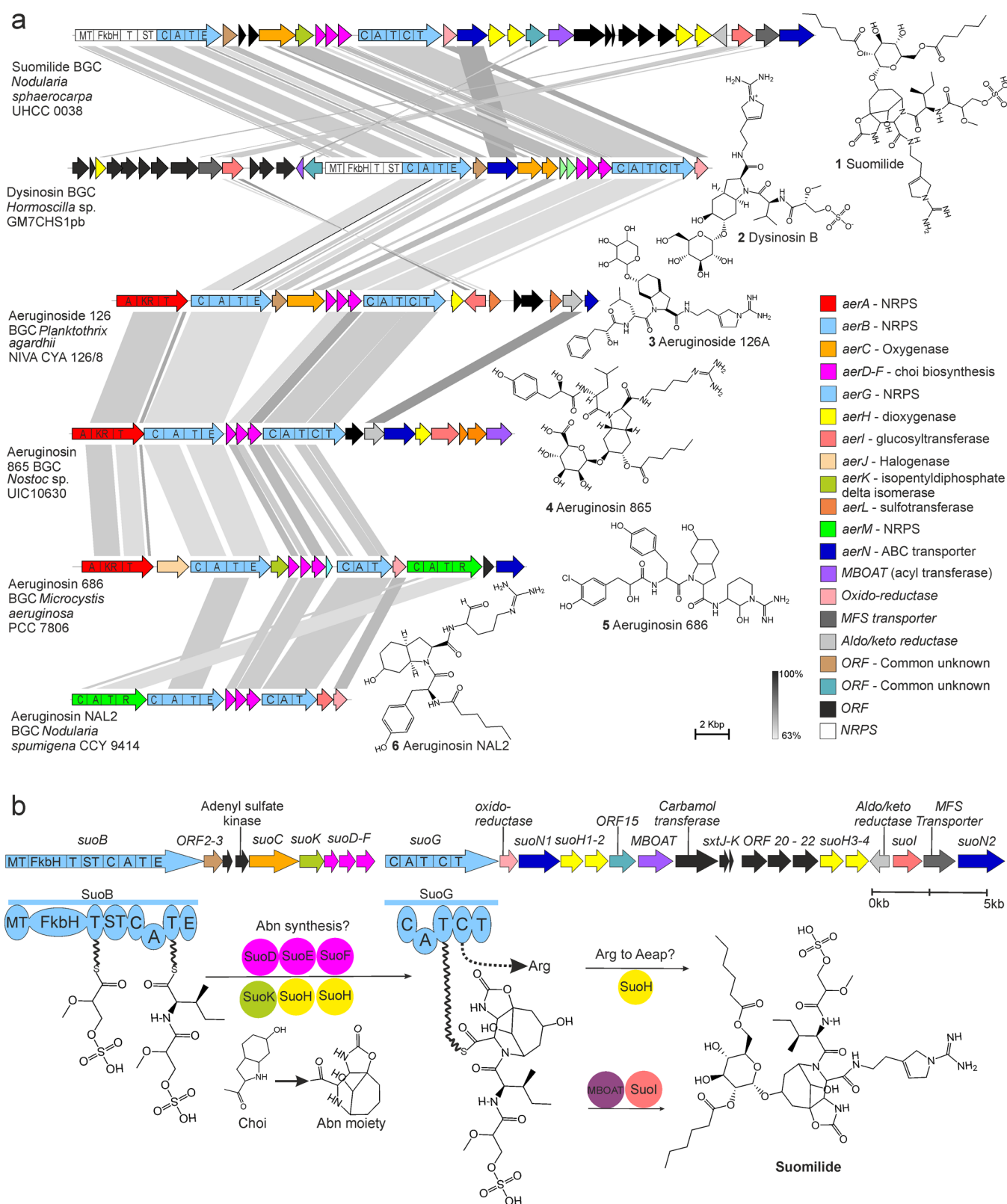
Trypsins are involved in the digestion of food, but they also have important functions in many tissues outside the digestive system.<sup>1</sup> The human genome encodes three trypsin isoenzymes (trypsin-1, -2, and -3) that are structurally and functionally highly similar.<sup>1</sup> Trypsin-3 (also known as mesotrypsin or PRSS3) has recently been shown to promote cancer progression and metastatic dissemination in prostate, breast, pancreatic, and lung cancer.<sup>2–5</sup> Thus, trypsin-3 is considered as an emerging therapeutic target for these cancers.<sup>2–5</sup> Although there are subtle differences between the substrate specificities of human trypsins,<sup>6</sup> the high sequence identity between trypsin isoforms extends to the active sites and makes selective targeting of trypsin-3 challenging.<sup>7–11</sup> Thus, current small-molecule inhibitors of trypsin-3 target all three trypsin isoenzymes with equal potency.<sup>11</sup>

Natural products have served as an inspiration in the development of new selective small-molecule inhibitors.<sup>12</sup> Indeed, most drugs that target proteases are based on or inspired by natural products.<sup>12</sup> Cyanobacteria produce an array of natural products with potent serine protease inhibition activity.<sup>13,14</sup> Aeruginosins are a diverse family of linear peptides

that are produced by a broad range of cyanobacteria.<sup>14</sup> They are potent inhibitors of serine proteases in low micromolar to low nanomolar concentrations.<sup>15</sup> Aeruginosins have a complex chemical structure characterized by the presence of an unusual 2-carboxy-6-hydroxyoctahydroindole (Choi) moiety and the C-terminal arginine derivatives argininal, argininal, agmatine, 1-amidino-2-ethoxy-3-aminopiperidine, or more rarely the 1-amino-2-(*N*-amidino- $\Delta^3$ -pyrrolinyl)-ethyl (Aeap) moiety<sup>16–19</sup> (Figure 1, panel a). The N-terminus typically consists of phenyl lactic acid derivatives<sup>20</sup> but may also contain hexanoic acid<sup>18</sup> or 2-O-methyl-3-sulfoglyceric acid (Mgs)<sup>21,22</sup> (Figure 1, panel a). Aeruginosins may also be glycosylated, sulfonated, and halogenated.<sup>18,19,23</sup> This array of structural features is responsible for their differing affinities to the catalytic binding pockets of trypsin, thrombin, and other serine proteases, as

Received: August 5, 2021

Accepted: October 4, 2021

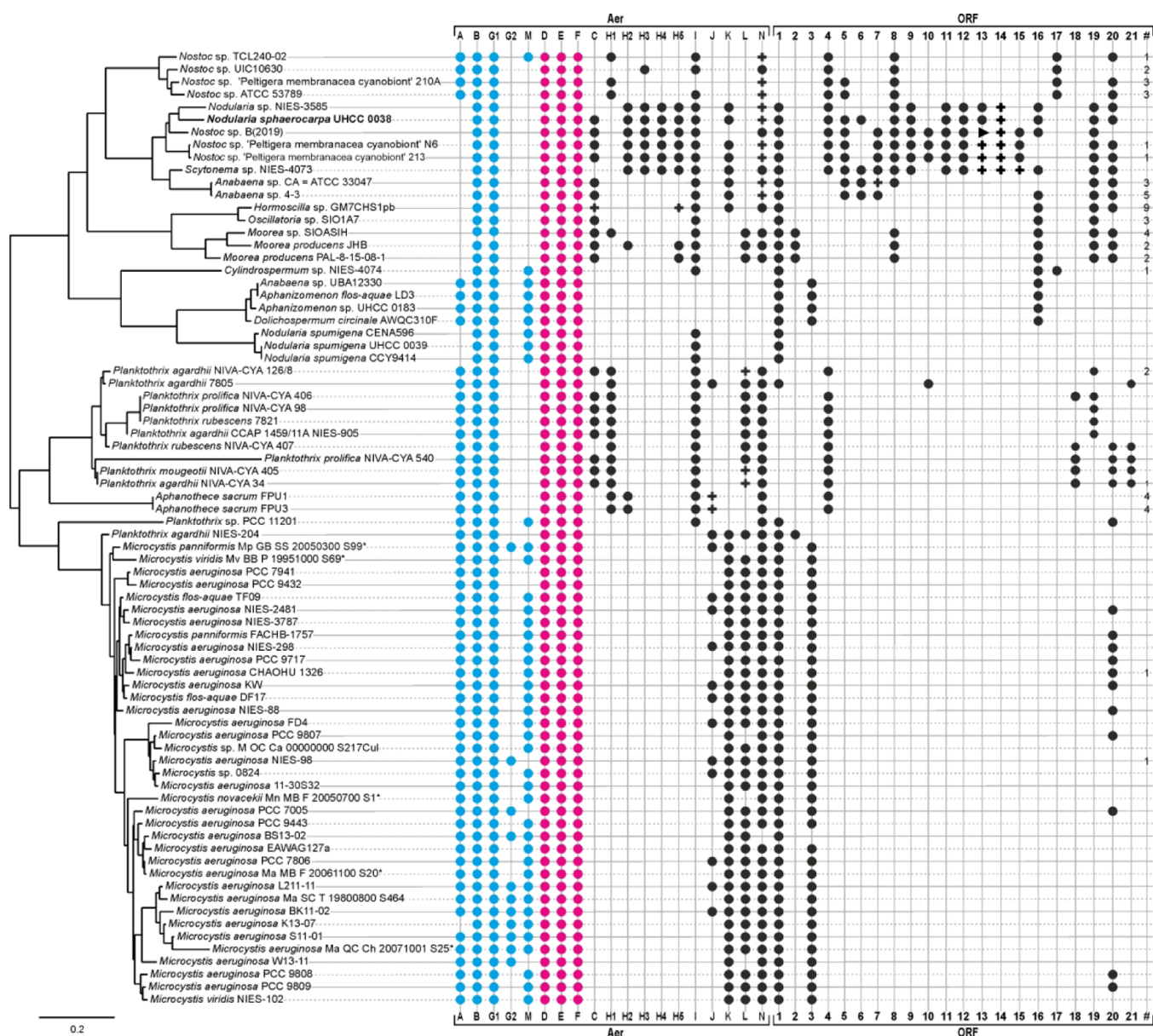


**Figure 1.** Chemical and genetic diversity in the aeruginosin family of protease inhibitors and discovery of the suomilide biosynthetic gene cluster (BGC). (a) Aeruginosin BGCs in comparison with the suomilide BGC. (b) Tentative biosynthetic scheme for the biosynthesis of suomilide.

shown by the X-ray crystallography of aeruginosin–protease complexes.<sup>15,24,25</sup>

Aeruginosins are synthesized on nonribosomal peptide synthetase (NRPS) enzyme complexes encoded in the aeruginosin (*aer*) biosynthetic gene cluster (BGC).<sup>18,20,22,23</sup>

Aeruginosin biosynthesis begins with the activation of a monocarboxylic acid or fatty acid by the loading module AerA.<sup>20,23</sup> AerB is responsible for adding a hydrophobic D-amino acid.<sup>20,23</sup> AerD, AerE, and AerF work in a cascading manner to synthesize and supply Choi to AerG.<sup>23,26–28</sup> AerG adds a Choi



**Figure 2.** Comprehensive representation of biosynthetic gene clusters (BGCs) from the aeruginosin family of protease inhibitors. The PhyML phylogenetic tree is based on AerB, AerG, AerD, AerE, and AerF proteins from aeruginosin BGC with the scale bar representing amino acid substitutions/site. Blue dots indicate NRPS, purple dots indicate Choi-making enzymes, black dots indicate a single copy of tailoring enzymes, plus sign represents two copies of the enzymes, and a triangle represents three copies of the enzymes. The last column represents the enzymes of unknown functions and are not homologous to each other.

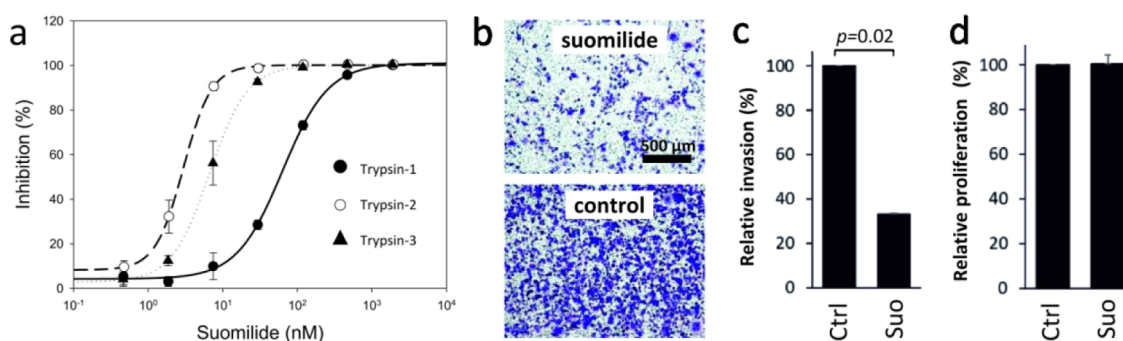
moiety and encodes an off-loading module.<sup>20,23</sup> Variation in the loading and off-loading mechanisms of aeruginosin synthetases contributes to the observed structural diversity of the natural product family.<sup>18–20,22,23</sup> Aeruginosin BGCs also encode an assortment of tailoring enzymes that promote structural diversification, including halogenation, acylation, and glycosylation.<sup>18,19,22,23</sup>

Suomilide is a complex tetrapeptide that contains unusual Mgs and Aeap moieties common in aeruginosins.<sup>29</sup> Although suomilide lacks Choi and instead contains a densely functionalized tricyclic azabicyclononane (Abn) moiety,<sup>29</sup> these structural features suggest that suomilide is a part of the aeruginosin family.<sup>30–32</sup> The recent inclusion of dysinosins into the aeruginosin family further suggests that suomilide is a derived member of the aeruginosin family.<sup>22</sup> However, the

relationship between the suomilide and aeruginosin biosynthetic pathways has remained unclear. Suomilide, similar to many aeruginosins, has been reported to inhibit thrombin, plasmin, and trypsin at low micromolar concentrations.<sup>29</sup>

Here, we identify the suomilide BGC and demonstrate that suomilide belongs to the aeruginosin family of protease inhibitors. Surprisingly, we show that suomilide potently inhibits human trypsin-2 and -3, while human trypsin-1 is inhibited to a lesser degree. Importantly, our results indicate that suomilide inhibits the invasion of prostate cancer cells, making it an attractive lead molecule for the development of a drug targeting cancers associated with high trypsin-3 activity. This information should facilitate the identification of aeruginosin variants, which may facilitate future drug development studies aiming to develop even more selective trypsin-3 inhibitors.





**Figure 3.** Suomilide inhibits human trypsin isoforms and PC-3M prostate cancer cell invasion. (a) Suomilide inhibits trypsin-1, -2, and -3 at IC<sub>50</sub> values of 104 ± 1.1 ± 6.7, 4.7 ± 0.6, and 11.5 ± 2.4 nM, respectively. (b) Effect of suomilide (3.3 μM) on PC-3M prostate cancer cell invasion. (c) Representative example of staining of invaded cells (blue) (scale: 500 μm) through Matrigel basement membrane preparation in the Transwell invasion model. (d) Cell proliferation. Buffer controls (Ctrl) contained the same amount of DMSO as suomilide preparations. Average + SE of two separate experiments with four–six replicates (*n* = 2).

## RESULTS AND DISCUSSION

**Suomilide BGC.** Suomilide and aeruginosins have been proposed to share a common biosynthetic logic based on the similarities in the chemical structure (Figure 1, panel a).<sup>15,29,32</sup> Whole-genome sequencing of *Nodularia sphaerocarpa* UHCC 0038 yielded a complete 5.45 Mb genome organized in a single chromosome (Supporting Information Figure 1). Bioinformatics analysis identified 16 secondary metabolite BGCs (Supporting Information Figure 1 and Supporting Information Table 1), one of which showed a strong similarity to aeruginosin BGCs. The putative 43.7 kb suomilide (*suo*) BGC consisted of 28 predicted proteins, of which 14 shared similarity to gene products residing in aeruginosin BGCs (Figure 1 panel a and Supporting Information Table 2).

We propose the following biosynthetic scheme for suomilide based on earlier studies on the biosynthesis of aeruginosins.<sup>20,23</sup> SuoB and SuoG encode two bimodular NRPS enzymes, which would form the suomilide backbone (Figure 1, panel b). The FkbH domain in the loading module of SuoB is likely responsible for the loading of glycerate,<sup>33</sup> which would then be methylated and sulfonated (Figure 1, panel b). An identical loading module from the dysinosin biosynthetic pathway is also proposed to synthesize Mgs through the activation of glyceric acid.<sup>22</sup> The second module of SuoB is predicted to activate and racemize L-Ile (Supporting Information Table 2). The first module of SuoG would activate a Choi-like compound (Figure 1 panel b and Supporting Information Table 2). The suomilide BGC encodes SuoD, SuoE, and SuoF, which are homologues of AerD, AerE, and AerF, respectively, involved in the biosynthesis of Choi.<sup>20,23</sup> The suomilide BGC also encodes oxygenases (*suoH1-4*) and carbamoyltransferase (ORF17), which likely participate in the biosynthesis of the unique Abn moiety (Figure 1 panel b and Supporting Information Table 2). The decoration of the Abn moiety with sugar is achieved by SuoI, and the addition of hexanoic acids is likely catalyzed by membrane-bound O-acyltransferase (Figure 1 panel b and Supporting Information Table 2). The terminal off-loading module of SuoG is responsible for the addition of arginine, which is converted to Aeap through a presently unknown biosynthetic mechanism.<sup>20</sup>

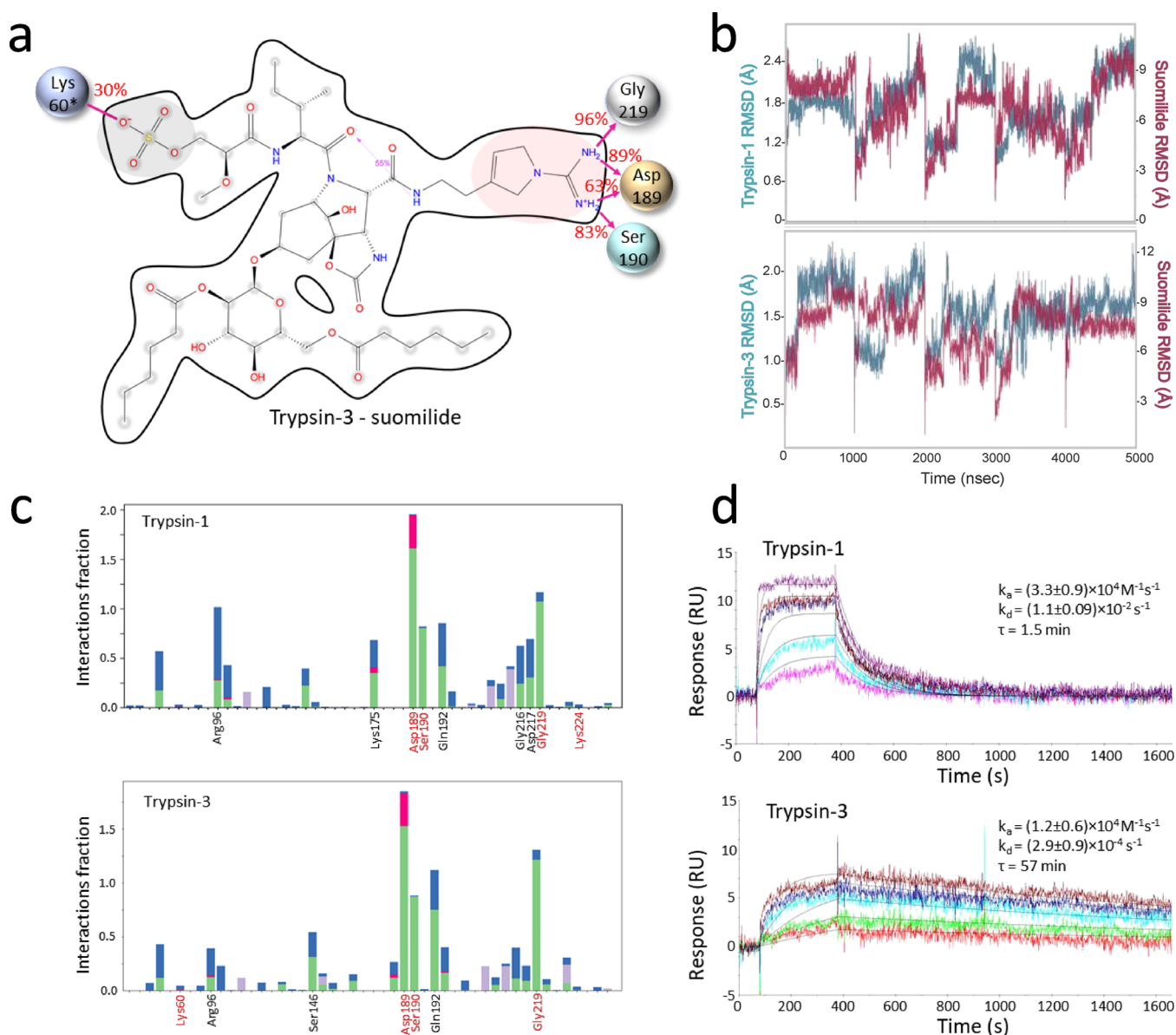
**Genetic Diversity of the Aeruginosin Family.** We examined all publicly available cyanobacterial genomes to gain insights into the genetic diversity of the aeruginosin biosynthetic pathways at the phylum level in cyanobacteria (Figure 2 and Supporting Information data set 1). We identified 103 BGCs from 12 morphologically and genetically diverse genera of

cyanobacteria (Figure 2 and Supporting Information data set 1). The aeruginosin BGCs range in size from 14 to 60 kb and encode 10–31 proteins (Figure 2). We observed that AerB, AerG, AerD, AerE, and AerF proteins, which are responsible for the assembly of the peptide intermediate and synthesis of the Choi precursor, were common to all aeruginosin BGCs (Figure 2). We constructed a phylogenetic tree from concatenated sequences of five core proteins encoded in 78 BGCs (Figure 2). The suomilide biosynthetic enzymes did not form a separate clade, thus confirming the earlier predictions that this compound is part of the aeruginosin family of natural products.<sup>30–32</sup>

Collectively, the aeruginosin biosynthetic pathways encode three different loading mechanisms and two off-loading mechanisms (Figures 1 and 2). These allow the incorporation of diverse starter units, including phenyl lactic acid, Mgs, or hexanoic acid.<sup>18,20,22,23</sup> The off-loading modules allow the release of the tetrapeptides with C-terminal arginine derivatives that include argininal, argininal, argininal, 1-amidino-2-ethoxy-3-aminopiperidine, or more rarely an Aeap moiety.<sup>16–19</sup> In addition, the aeruginosin BGCs encode a plethora of tailoring enzymes for glycosylation, sulfonation, oxygenation, halogenation, reduction, and carbamoylation of the final products (Figure 2 and Supporting Information data set 1). Our bioinformatic analyses identified many novel putative enzymes and gene combinations, suggesting that unique aeruginosin derivatives have yet to be discovered (Figure 2 and Supporting Information data set 1). This phylogenetic and genomic examination of the aeruginosin BGCs broadens the definition of the aeruginosin family of protease inhibitors to include compounds such as suomilide, dysinosin, and banyasides and suggests that the chemical diversity of the aeruginosin family is larger than currently anticipated. The selectivity of aeruginosins for different serine proteases could drive the pronounced chemical and genetic diversification of this variable family of protease inhibitors.<sup>15</sup>

### Suomilide Inhibits Human Trypsin Isoenzymes.

Suomilide has been reported to inhibit an unspecified but most likely non-human trypsin isoenzyme with an IC<sub>50</sub> value of 1.8 μM.<sup>29</sup> Compounds derived from cyanobacteria have most often been tested using bovine and porcine trypsins, whose primary structures differ from that of human trypsins more than human trypsins differ from each other.<sup>34</sup> We purified 6.55 mg of suomilide from *N. sphaerocarpa* UHCC 0038 (Supporting Information Figures 2 and 3) and tested its ability to inhibit



**Figure 4.** Molecular dynamics (MD) simulation of trypsin-1 and -3 interactions with suamilide and residence time. (a) Schematic of detailed ligand atom interactions with the protein residues of trypsin-3. Interactions that occur more than 30% of the simulation time (0–5004 ns) in the selected trajectory are shown. The Aeap moiety of suamilide is highlighted in red and the Mgs moiety in gray. An almost identical interaction was obtained with trypsin-1. \*, Lys224 in trypsin-1. (b) Root mean square deviation plot of suamilide interaction with trypsin-1 and trypsin-3 during MD simulations. (c) Histogram of amino acids of trypsin-1 and -3 that are in contact with suamilide in MD simulation. Only amino acids whose interactions exceeded 0.5 or which are shown in (a) are shown. (d) Kinetic analysis of the suamilide interaction with trypsin-1 and trypsin-3 by surface plasmon resonance (Biacore).

recombinant human trypsin-1, -2, and -3 using a chromogenic trypsin substrate. Suamilide inhibited trypsin-1, -2, and -3 with  $\text{IC}_{50}$  values of  $104 \pm 7$ ,  $4.7 \pm 0.6$ , and  $11.5 \pm 2.4$  nM, respectively (Figure 3, panel a). The preferential inhibition of trypsin-2 and -3 is notable, given the high sequence similarity between human trypsin isoenzymes that extends to the active sites, making selective targeting of individual trypsins with small molecules challenging.<sup>11</sup> Indeed, suamilide shows better selectivity and potency for human trypsin-3 than the previously described small-molecule human trypsin inhibitors, such as diminazene, which equally inhibits all three human trypsins and only at micromolar concentrations.<sup>11</sup> With a phage-display approach, we previously identified peptides of about 10 amino acids that show some selectivity among human trypsins, preferentially

targeting trypsin-1 and trypsin-3. However, these peptides have lower potency than suamilide and also potentially inhibit plasma kallikrein.<sup>10</sup> As trypsin-3 has some unique features, such as Arg193 instead of Gly193 (as in bovine trypsin), it can proteolytically cleave and inactivate most natural proteinaceous trypsin inhibitors.<sup>8</sup> When Gly193 in human trypsin-1 was mutated to Arg193, it conferred a trypsin-3-like resistance to inhibitors.<sup>9,35</sup> On the other hand, proteinaceous inhibitors have been modified to effectively target trypsin-3.<sup>36</sup> Although all human trypsins have highly similar substrate specificities, the specificity of trypsin-3 is somewhat different from those of trypsin-1 and -2, which has also allowed the creation of substrates that are preferentially cleaved either by trypsin-3 or

trypsin-1 and -2.<sup>6</sup> This may form a basis for the identification of more selective trypsin-3 inhibitors.

Many trypsin inhibitors also target several other trypsin-like proteases. Accordingly, we tested whether suamilide also inhibits human plasmin, plasma kallikrein, and Factor Xa. These were inhibited at IC<sub>50</sub> values of  $7.6 \pm 2.3$ ,  $82 \pm 24$ , and  $114 \pm 2$  nM, respectively. However, human urokinase was inhibited at an IC<sub>50</sub> value of  $7.1 \pm 1.36$   $\mu$ M, and human thrombin was minimally inhibited (IC<sub>50</sub> >12.5  $\mu$ M). This contrasts somewhat with the previous study,<sup>29</sup> showing that suamilide inhibits undefined thrombin and plasmin with similar micromolar concentrations. Although high selectivity would be preferred for drug leads, the inhibition of plasmin may be a favorable property as plasmin is also involved in tumor progression.<sup>37</sup>

**Suamilide and Trypsin-3 Interaction.** To explain the selectivity of suamilide in inhibiting trypsin-2 and -3 over trypsin-1, we studied the potential binding mode of suamilide to trypsin-1 and trypsin-3, for which crystal structures are available.<sup>7,9</sup> The suamilide structure was modeled, and molecular docking (Glide XP) was performed to obtain the initial docking poses. While suamilide easily docked to the trypsin-3 active site (Figure 4), trypsin-1 docking was only successful manually by using the identical binding mode as in trypsin-3. The docked complexes were then simulated for  $5 \times 1$  microseconds. Surprisingly, for both trypsin-1 and -3, docking poses with suamilide remained almost intact throughout the simulation. In both trypsins, the hydrogen bonds between Asp189, Ser190, and Gly219 and the Aep moiety of suamilide in particular stayed intact during the whole molecular dynamics (MD) simulation, indicating that they are important for the interaction (Figure 4, panel a and b). Asp189, which is located at the base of the specificity pocket of trypsins, has previously been found to be important also for the interaction between trypsin-3 and the trypsin inhibitor diminazene.<sup>11</sup> The Mgs moiety also appeared to be important for the interaction with Lys60 in trypsin-3 or Lys224 in trypsin-1 (Figure 4, panel c). However, this interaction was not as stable as that with Aep. These results indicate a very stable trypsin–suamilide complex (i.e., long residence time).

To experimentally validate the MD simulations, we used surface plasmon resonance, in which trypsins were immobilized on chips and binding of suamilide was followed over time. We observed that suamilide had a residence time in the range of minutes (1.5 min for trypsin-1 and 57 min for trypsin-3) with both trypsins (Figure 4, panel d). The longer residence time with trypsin-3 may explain why it was inhibited more effectively. Long residence time suggests a long-lasting drug effect, which is a desired property for many drugs, including those inhibiting the invasion of cancer cells.<sup>38</sup> While molecular docking and molecular docking simulation did not provide a clear explanation for why suamilide inhibits trypsin-2 and -3 more potently than trypsin-1, these analyses were partially able to explain the strong inhibitory potency of suamilide.

**Suamilide Inhibits the Invasion of Prostate Cancer Cells.** Trypsin-3 has gained considerable interest as a potential drug target. Trypsin-3 is overexpressed in various cancers and appears to drive cancer progression.<sup>2–5</sup> We tested the bioactivity of suamilide using a Transwell cell invasion assay, in which the invasion of cells through a Matrigel basement membrane preparation was studied. In this model, 3.3  $\mu$ M suamilide significantly inhibited the invasion of trypsin-3-producing aggressive metastatic PC-3M prostate cancer cells.<sup>39</sup> Suamilide

did not affect the proliferation of cells grown on a plastic surface (Figure 3, panels b–d). The model was chosen based on previous studies, indicating that that trypsin-3 is an important mediator of prostate cancer progression and metastasis.<sup>3,40</sup> In these previous studies, the role of trypsin-3 in prostate cancer was shown using a mouse orthotropic model and different cell models. These included PC-3M cells, whose invasion was reduced upon the knockdown of trypsin-3 expression by RNA interference and a modified proteinaceous trypsin-3 inhibitor.<sup>3,40</sup> Of note, the knockdown of trypsin-1 in PC-3M cells did not have such an effect.<sup>3</sup> Thus, cell invasion in this model appears to be regulated by the trypsin-3 activity. The functional studies support the potential of suamilide as a lead compound for the development of an anti-invasive cancer drug. Targeting invasion is important as it is a prerequisite of the metastatic dissemination of cancer cells, and most cancer-associated deaths in patients with solid cancers (such as prostate cancer) are caused by metastases.<sup>41</sup>

## CONCLUSIONS

We explored the biosynthetic diversity encoded in aeruginosin biosynthetic pathways using whole-genome sequencing and bioinformatics analyses and confirmed the long-standing hypothesis that aeruginosins and suamilide are members of the same family of natural products. The overall genetic diversity of the aeruginosin biosynthetic gene clusters suggests that the chemical and genetic diversity is greater than currently appreciated. We observed that suamilide is a potent inhibitor of human trypsin-2 and -3 isoenzymes and also inhibits the invasion of PC-3M prostate cancer cells without affecting proliferation. Therefore, in addition to showing that suamilide is a potential lead for the development of anti-invasive drugs, in a broader context, this suggests that aeruginosins are a good source of lead compounds for the further development of protease inhibitors.

## METHODS

**Genome Sequencing.** The cyanobacterial strain *N. sphaerocarpa* UHCC 0038 (formerly HKVV) was grown in a photon irradiance of 15  $\mu$ mol m<sup>−2</sup> s<sup>−1</sup> in 100 mL of Z8XS culture medium, obtained by adding 0.88% NaCl and 0.38% MgSO<sub>4</sub>·7H<sub>2</sub>O to Z8 media,<sup>42</sup> for 25–30 days at 21–25 °C. Genomic DNA was isolated using an E.Z.N.A. SP Plant DNA extraction and purification kit (Omega Bio-Tek). The culture was centrifuged at 7000g for 7 min (Sorvall LYNX 6000 superspeed centrifuge, Thermo Fisher Scientific). The supernatant was discarded, and the pellet was transferred to a 2 mL tube. 100  $\mu$ L of 425–600  $\mu$ m beads and 100  $\mu$ L of 700–1180  $\mu$ m of beads (Scientific Industries) were added to the pellet after addition of 600  $\mu$ L of SP1 buffer (provided with the kit) to 50 mg of cells. The cells were mechanically broken using a Fastprep cell disrupter (Bio 101, Thermo Electron Corporation, Qbiogene, Inc.) at a speed of 6.5 ms<sup>−1</sup> for 20 s, repeating this step twice. 5  $\mu$ L of RNase buffer (provided with the kit) was added prior to vortexing. The remainder of the procedure was followed as indicated in the kit manual. The purification yielded 1616  $\mu$ g of genomic DNA. 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, both reaching an average of 77× coverage. Nextera adapters were removed from Illumina data by cutadapt v1.14, keeping reads with a minimum of 50 bp length and Phred 25 quality. PacBio data were assembled with HGAP3, implemented in SMRTportal 2.3. Illumina data were assembled with SPAdes v3.11.1 with the “careful” option. Gap4 was used to combine assemblies and check sequence circularity. Illumina reads were mapped



against the combined assembly with bwa 0.7.12-r1039. The genome was then polished using Pilon v1.16 to fix homopolymer errors. Finally, the complete sequence was annotated using Prokka v1.13, and protein functional annotation was performed with PANNZER2. All sequencing, assembly, and annotation were performed at the DNA Sequencing and Genomics laboratory, Institute of Biotechnology, University of Helsinki. The complete genome sequence of *N. sphaerocarpa* UHCC 0038 was deposited in the GenBank database under the accession number CP060140.

**Identification of Aeruginosin BGCs.** AerD, AerE, AerB, and AerG from the aeruginosin BGC of *Planktothrix agardhii* NIVA-CYA 126/11 were used to query the non-redundant (*nr*) database at NCBI in BLASTp searches. The identified BGCs were manually annotated using Artemis version 18.1. BLASTp searches were performed on the GenBank database (release no. 242.0) to predict the function of genes present in the BGCs; the BGCs were then annotated manually. The conserved domains (CD) in the NRPS and the function of the genes were also predicted by using the NCBI CD Database (CDD). The substrate specificity of the NRPS adenylating domains was predicted using the NRPSpredictor2 online version and CDD searches. antiSMASH 6.0 analysis was performed to identify the BGCs in the genome of *N. sphaerocarpa* UHCC 0038. Easyfig (version 2.2.2.) was used to compare and draw the BGCs of *N. sphaerocarpa* UHCC 0038, *Hormoscilla* sp. GM7CHS1pb, *P. agardhii* NIVA-CYA 126/8, *Nostoc* sp. UIC 10630, *Microcystis aeruginosa* PCC 7806, and *Nodularia spumigena* CCY9414.

**Phylogenetic Trees.** Phylogenetic trees based on suimilide and aeruginosin biosynthetic proteins were constructed using PhyML to test the relationship between suimilide and aeruginosin.<sup>43,44</sup> An alignment was constructed combining the AerB (CATE domains only), AerG (CAT domains only), AerD, AerE, and AerF proteins obtained from the suimilide and aeruginosin BGCs. The ClustalW multiple sequence alignment was constructed using BioEdit (version: 7.0.5.3). The phylogenetic trees, constructed using the “advanced” model, relied on sequence homology.<sup>43</sup> Using this model, we skipped the multiple sequence alignment as this was performed using BioEdit. The curation of the aligned sequences was performed with the Gblocks program, which excluded the ambiguous regions in the aligned sequences. PhyML program was used for the construction of the trees. Visualization of the phylogenetic trees was performed by TreeDyn software.

**Chemical Analysis.** The biomass of *N. sphaerocarpa* UHCC 0038 was harvested from 40 mL cultures through centrifugation at 7000g for 7 min (Sorvall LYNX 6000 superspeed centrifuge, Thermo Fisher Scientific) and was subjected to freeze-drying (Christ Beta 2–8 LSCplus, LyoCube 4–8) for 72 h, yielding 10–50 mg of the dried sample. The freeze-dried sample was extracted with 1 mL of 70% methanol [HiperSolv for high-performance liquid chromatography (HPLC), BDH Laboratory supplies], 200  $\mu$ L of 0.5 mm glass beads (Scientific Industries) was added, and cells were mechanically broken using a FastPrep cell disrupter (Bio 101, Thermo Electron Corporation, Qbiogene, Inc.) at a speed of 6.5 ms<sup>-1</sup> for 30 s. The suspension was centrifuged at 20 000 rpm for 5 min (Eppendorf Centrifuge 5804R, Eppendorf AG) at room temperature and was stored at 4 °C for LC–mass spectrometry (LC–MS) and quadrupole time-of-flight (QTOF) analyses.

**Liquid Chromatography–Mass Spectrometry.** 100  $\mu$ L of the MeOH extract was filtered through a 0.2  $\mu$ m filter (13 mm syringe filter, PTFE, VWR International). Sample analysis was performed with an ultra-performance liquid chromatograph interfaced to a high-definition electrospray ionization mass spectrometer [Acquity ultra-performance LC (UPLC) I-Class-Synapt G2-Si Q-ToF, Waters]. A Kinetex C8 (50  $\times$  2.1 mm, 1.7  $\mu$ m 100 Å, Phenomenex) column was used for separation. The mobile phase consisted of a 0.1% aqueous solution of formic acid (A) and 0.1% formic acid in a 1:1 mixture of acetonitrile and 2-propanol (B). The following gradient was run: from 5 to 100% B in 5 min, maintained at 100% B for 2 min, and then back to 5% B in 30 s where it was maintained for 2.5 min. The total run time was 10 min. The injection volume was 1  $\mu$ L. The instrument parameters for MS and MS/MS analysis were a capillary voltage of 1.5 kV, a source offset of 80.0, a

source temperature of 120 °C, a sampling cone of 40.0, a desolvation temperature of 600 °C, a desolvation gas flow of 1000 L h<sup>-1</sup>, an ion energy of 1.0, a nebulizer gas pressure of 6.5 bar, a trap wave velocity of 300.0 V, and a trap collision energy of 30.0.

**Purification of Suimilide.** *N. sphaerocarpa* UHCC 0038 was grown in 210 L of Z8XS culture medium, centrifuged at 7 000g for 7 min and then lyophilized (Christ Beta 2–8 LSCplus, LyoCube 4–8) for 72 h, yielding 30 g of dried biomass. The dried cells were lyophilized, and 5 g was lysed in 60 mL of methanol using a SilentCrusher M homogenizer (Heidolph) for 30 s at 15 000 rpm. After the centrifugation at 10 000g for 5 min, the supernatant was collected, and the extraction was repeated. The crude extract (120 mL) was rotary-evaporated (Büchi) to dryness in 5 g of octadecylsilyl silica gel (Fuji-Davison Chemical Ltd.) at 150 mbar at 25 °C. The silica gel-bound sample was fractionated in two batches using a solid-phase Strata SI-1 silica cartridge 5 g/20 mL (Phenomenex). The sample was loaded on a preconditioned cartridge and sequentially eluted with *n*-heptane, acetonitrile, acetone, ethyl acetate, and methanol. The fractions were analyzed with the aforementioned UPLC–QTOF system. The methanol fractions that contained suimilide were dried under a nitrogen gas stream, and suimilide was purified by semi-preparative HPLC combined with MS using a Luna C8(2) (150  $\times$  10.00 mm, 5  $\mu$ m, Phenomenex) column and a Waters Auto Purification System. The positive-ion mode was used in the electrospray ionization. The sample was dissolved in 2 mL of the eluent, and the injection volume was 150  $\mu$ L with a flow rate of 5 mL min<sup>-1</sup>. The mobile phase consisted of 0.1% aqueous formic acid (solvent A) and acetonitrile/isopropanol (1:1) (solvent B). A gradient from 30 to 100% of solvent B over 7 min was used, and the total run time was 14 min. The organic solvents were evaporated from the fractions under a nitrogen gas stream. The fractions were dissolved in methanol and pooled together to be analyzed with the aforementioned UPLC–QTOF system. Analysis showed that the suimilide preparation contained only very minor amounts of other suimilide variants (Supporting Information Figure 2). The molecular formula of C<sub>45</sub>H<sub>73</sub>N<sub>7</sub>O<sub>19</sub>S was established based on UPLC–QTOF analysis (Supporting Information Figure 2). *N. sphaerocarpa* UHCC 0038 does not synthesize secondary metabolites such as nodularins, spumigins, aeruginosins, and nodulapeptins,<sup>18,44</sup> typical for *N. spumigena* strains and some other *Nodularia* species (Supporting Information Figure 3). To further assess the purity, suimilide was evaporated to dryness and subjected to nuclear magnetic resonance (NMR).

**Concentration and Purity Determination through NMR.** The purification yielded 7.7 mg of suimilide from *N. sphaerocarpa* UHCC 0038, as determined by weighing. Suimilide was dissolved in 0.5 mL of DMSO-*d*<sub>6</sub> (dimethyl sulfoxide). Commercial dipeptide Tyr–Val (Sigma-Aldrich) was dissolved to 1/10 of the volume of the suimilide solution in an equimolar ratio with suimilide. The comparison of integrated proton spectrum signals from suimilide and Tyr–Val dipeptide gave a 10% lower concentration for suimilide than that obtained by weighing. <sup>13</sup>C HSQC (heteronuclear single quantum coherence) showed several minor signals in the carbohydrate region in addition to two minor signals in the anomeric region (Supporting Information Figure 4). This indicates that the suimilide preparation contains suimilide variants in which fatty acid residues are different or located in different positions in the glucose unit. The comparison of suimilide and variant signal intensities yielded an estimate of 85% suimilide and 15% suimilide variants. The weight determined by NMR (6.55 mg) was used in this study.

One-dimensional <sup>1</sup>H and two-dimensional CH<sub>3</sub>-edited <sup>13</sup>C-HSQC NMR spectra were measured in a Bruker-AVANCE III HD 800 NMR spectrometer, equipped with a cryogenically cooled triple-resonance (<sup>1</sup>H, <sup>13</sup>C, and <sup>15</sup>N) probe head and z-gradient at 30 °C (303 K). For <sup>1</sup>H spectra, 16 k complex points, corresponding to an acquisition time of 1.28 s, were collected using 16 transients. <sup>13</sup>C-HSQC spectra were collected with 24 transients per FID and 256/1024 complex points in *t*<sub>1</sub>/*t*<sub>2</sub>, corresponding to acquisition times of 7.1 and 92 ms, respectively.

**Serine Protease Inhibition.** Recombinant trypsinogens-1, -2, and -3 were produced in *Escherichia coli* and activated to trypsin-1, -2, and -3 as described previously.<sup>45</sup> The enzyme inhibitory activity of suimilide



was determined in 96-well plates by adding 10  $\mu\text{L}$  of diluted suamilide (0.21–3400 nM, diluted in ultra-purified water, containing up to 0.027% DMSO), 15  $\mu\text{L}$  of 0.1% bovine serum albumin in 50 mM Tris-buffer saline (BSA/TBS), and 25  $\mu\text{L}$  of each trypsin (1.88 nM). 10  $\mu\text{L}$  of ultra-purified water was used as a control. DMSO at the concentrations used did not have any effect on the activity of trypsins (data not shown). Preincubation of suamilide with trypsin isoenzymes prior to the addition of the substrate was required for maximal inhibition (data not shown). Thus, suamilide was preincubated with trypsins at room temperature for 30 min before the addition of 50  $\mu\text{L}$  of 0.2 mM chromogenic substrate S-2222 (ChromogeniX) in ultra-purified water. The change of absorbance at 405 nm was followed for 15 min (VICTOR X4 multilabel plate reader, PerkinElmer). The experiment was performed three times, each with two replicates ( $n = 3$ ). The absorbance change during the phase of the substrate reaction in which the absorbance of the control reaction increased linearly over time was determined. The same experiment was performed with plasmin (Sigma) (15 nM) with chromogenic substrate S-2251 (ChromogeniX); plasma KLK (Sigma) (20 nM) with substrate ChromogeniX S-2302 (ChromogeniX); Factor Xa (Sigma) (3 nM) with substrate S-2222; and thrombin (Sigma) (1 nM) with substrate S-2238 (ChromogeniX). Except for urokinase from human kidney cells (Sigma) (125 IU), 15  $\mu\text{L}$  of BSA/TBS with 0.9% NaCl was used. Nine 4-fold dilutions of suamilide were prepared, with 12.5  $\mu\text{M}$  being the highest final concentration. The  $\text{IC}_{50}$  values were estimated using Quest Graph, and the standard error was calculated using MS Excel 2013.

**Effect of Suamilide on the Invasion and Proliferation of PC-3M Cells.** PC-3M prostate cancer cells<sup>46</sup> were cultured in Ham's F12 Nutrient Mix with GlutaMAX (Gibco), supplemented with 10% fetal calf serum (FCS), 100 IU  $\text{mL}^{-1}$  penicillin, 100  $\mu\text{g mL}^{-1}$  streptomycin, and 0.15% sodium bicarbonate at 37  $^{\circ}\text{C}$  in a humidified,  $\text{CO}_2$ -controlled (5%) incubator. For the invasion assay, PC-3M cells ( $1.45 \times 10^5$ ) in a serum-free medium containing 0 or 3.3  $\mu\text{M}$  suamilide were added into polycarbonate membrane Transwell inserts with 8.0  $\mu\text{m}$  pore size (Thermo Fisher) covered with 100  $\mu\text{L}$  of the phenol red-free Matrigel basement membrane matrix (Corning) diluted 1:30 with phosphate-buffered saline. The inserts were placed into a 24-well plate containing 750  $\mu\text{L}$  of the culture medium with 10% FCS and incubated for 4 days at 37  $^{\circ}\text{C}$  in a humidified,  $\text{CO}_2$ -controlled (5%) incubator. Matrigel and the noninvasive cells were removed using cotton swabs, and the membranes were stained with 0.09% crystal violet for 10 min. The experiment was performed twice, both with four replicates ( $n = 2$ ). The invasion was quantitated from microscopic images (three panels from each of the replicates) using Fiji ImageJ. For the proliferation assay, 7500 cells/well were cultured as above on a 96-well plate with 0 or 3.3  $\mu\text{M}$  suamilide in 100  $\mu\text{L}$  of culture media containing 1% FCS. After 3 days, 10  $\mu\text{L}$ /well CCK-8 solution (Cell Counting Kit-8, Dojindo) was added, and absorbance at 450 nm was measured after 1.5 h incubation. The experiment was performed twice, both with six replicates ( $n = 2$ ). *T*-Test (two-tailed, Microsoft Excel) was used to assess the differences in the invasion and proliferation between suamilide-treated and control cells.

**Molecular Modeling.** All modeling was conducted with Maestro Small-Molecule Drug Discovery Suite 2021-1 (Schrödinger Release 2021-1: Maestro, Schrödinger, LLC, 2021) with an OPLS4 force field unless otherwise stated.<sup>47</sup> The figures were prepared with PyMOL 2.4.2 (Schrödinger, LLC). For the docking and MD simulations, we used the PDB structures 2RA3 (trypsin-1)<sup>7</sup> and 1H4W (trypsin-3),<sup>9</sup> which were prepared using Protein Preparation Wizard (default settings with the "cap termini" option).<sup>48</sup> The amino acid numbering used here is based on the numbering in these PDB structures. The initial coordinates for the ligand–protein complexes were obtained by Glide XP docking using default settings.<sup>49–51</sup> The grid box was defined by SiteMap (the SiteMap region covering the active site was selected for GRID).<sup>52</sup> Before docking, suamilide was prepared with LipPrep (default settings) using Epik, and the tautomeric state of suamilide was based on the QM Conformer & Tautomer Predictor results. The MD simulations were conducted with Desmond.<sup>53</sup> The systems were solvated in a cubic box (edges 14 Å from the protein) and neutralized with counter ions (2  $\text{Cl}^-$ ) with 0.15 M NaCl salt. The water was described with the SPC

water model. The final trypsin-1 and trypsin-3 systems consisted of 33 840 and 36 545 atoms, respectively. The default relaxation protocol of Desmond was used before the 1000 ns production simulations, which were conducted in the NPT ensemble (300 K, thermostat: Nosé–Hoover chain; 1.01325 bar, barostat: Martyna–Tobias–Klein). The default time step of 2 fs and cutoff radius of 9.0 Å for Coulombic interactions were used. The total simulation time was 5  $\mu\text{s}$  ( $5 \times 1000$  ns) for both systems. All the MD setup and trajectory files and films are available at Zenodo (<https://zenodo.org/record/5512048#.YVF1w-dRVaQ>).

**Surface Plasmon Resonance.** The binding of suamilide to trypsin-1 and -3 was analyzed with surface plasmon resonance in Biacore 2000 instrumentation (Cytiva). The flow cells of a CMD700M biosensor chip (Xantec bioanalytics) were covalently coated with trypsin-1, trypsin-3, or a proteolytically inactive trypsin-3 mutant (negative control) via standard amine coupling. The coating resulted in a resonance unit increase of 5600–7200. The binding was analyzed using Tris-buffered saline (pH 7.4) containing 0.01% DMSO as the running buffer and a flow rate of 30  $\mu\text{L min}^{-1}$ . The kinetics of the suamilide interaction with the trypsins were measured by varying suamilide concentrations from 250 to 4000 nM with a contact time of 5 min and a dissociation phase of 20 min. After the completion of the dissociation phase, the flow cells were regenerated with 10 mM Glycine, pH 1.7. The data were evaluated by first subtracting the sensorgrams obtained from the trypsin-3 mutant control flow cell from those obtained from the flow cells containing trypsin-1 or -3 for each suamilide concentration separately and overlaying the resulting sensorgrams for each ligand. Association and dissociation rate constants were obtained by the Langmuir global fit model (BiaEvaluation Software, Cytiva). The residence time was calculated as a reciprocal of the dissociation constant.

## ■ ASSOCIATED CONTENT

### ■ Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acschembio.1c00611>.

103 aeruginosin BGCs identified through the bioinformatic analysis. A link for molecular dynamic simulations: <https://zenodo.org/record/5512048#.YVF1w-dRVaQ> (XLSX)

Analysis of polyketide synthetases and NRPS BGCs in the complete genome of *N. sphaerocarpa* UHCC 0038, UPLC–QTOF purity analysis of suamilide preparation, total ion chromatograms of positive and negative ionizations of the *N. sphaerocarpa* UHCC 0038 methanol extract,  $^1\text{H}$  and edited  $^{13}\text{C}$  HSQC spectra of suamilide, antiSMASH analysis of the *N. sphaerocarpa* UHCC 0038 genome, and function prediction of the suamilide biosynthetic gene cluster from *N. sphaerocarpa* UHCC 0038 and comparison between suamilide and aeruginosin biosynthetic gene cluster proteins (PDF)

### Accession Codes

The accession number of the complete genome sequence of *N. sphaerocarpa* UHCC 0038: CP060140.

## ■ AUTHOR INFORMATION

### Corresponding Authors

Hannu Koistinen – Department of Clinical Chemistry and Haematology, Faculty of Medicine, University of Helsinki and Helsinki University Hospital, Helsinki FIN-00014, Finland; Email: [hannu.koistinen@helsinki.fi](mailto:hannu.koistinen@helsinki.fi)

David P. Fewer – Department of Microbiology, Faculty of Agriculture and Forestry, University of Helsinki, Helsinki FIN-00014, Finland;  [orcid.org/0000-0003-3978-4845](https://orcid.org/0000-0003-3978-4845); Email: [david.fewer@helsinki.fi](mailto:david.fewer@helsinki.fi)

## Authors

**Muhammad N. Ahmed** – Department of Microbiology, Faculty of Agriculture and Forestry, University of Helsinki, Helsinki FIN-00014, Finland; Department of Clinical Chemistry and Haematology, Faculty of Medicine, University of Helsinki and Helsinki University Hospital, Helsinki FIN-00014, Finland

**Matti Wahlsten** – Department of Microbiology, Faculty of Agriculture and Forestry, University of Helsinki, Helsinki FIN-00014, Finland; [orcid.org/0000-0002-4107-1695](https://orcid.org/0000-0002-4107-1695)

**Jouni Jokela** – Department of Microbiology, Faculty of Agriculture and Forestry, University of Helsinki, Helsinki FIN-00014, Finland; [orcid.org/0000-0001-5096-3575](https://orcid.org/0000-0001-5096-3575)

**Matthias Nees** – Department of Biochemistry and Molecular Biology, Medical University in Lublin, Lublin 20-093, Poland; Institute of Biomedicine and Western Cancer Centre FICAN West, University of Turku, Turku 20101, Finland

**Ulf-Håkan Stenman** – Department of Clinical Chemistry and Haematology, Faculty of Medicine, University of Helsinki and Helsinki University Hospital, Helsinki FIN-00014, Finland

**Danillo O. Alvarenga** – Department of Microbiology, Faculty of Agriculture and Forestry, University of Helsinki, Helsinki FIN-00014, Finland; Department of Biology, Faculty of Science, University of Copenhagen, Copenhagen DK-2100, Denmark; [orcid.org/0000-0002-3666-2220](https://orcid.org/0000-0002-3666-2220)

**Tomas Strandin** – Department of Virology, Faculty of Medicine, University of Helsinki, Helsinki FIN-00014, Finland

**Kaarina Sivonen** – Department of Microbiology, Faculty of Agriculture and Forestry, University of Helsinki, Helsinki FIN-00014, Finland; [orcid.org/0000-0002-2904-0458](https://orcid.org/0000-0002-2904-0458)

**Antti Poso** – School of Pharmacy, University of Eastern Finland, Kuopio FIN-70211, Finland; Dept. of Internal Medicine VIII, University Hospital Tübingen, Tübingen DE-72076, Germany; [orcid.org/0000-0003-4196-4204](https://orcid.org/0000-0003-4196-4204)

**Perttu Permi** – Department of Biological and Environmental Science and Department of Chemistry, Nanoscience Center, University of Jyväskylä, Jyväskylä FI-40014, Finland

**Mikko Metsä-Ketelä** – Department of Biochemistry, University of Turku, Turku FIN-20014, Finland; [orcid.org/0000-0003-3176-2908](https://orcid.org/0000-0003-3176-2908)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acscchembio.1c00611>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We would like to thank A. Löfhjelm and L. Saari for excellent technical assistance. This work was supported by a Sigrid Jusélius Foundation grant to H.K. and the Academy of Finland funding (321809) to T.S. We would also like to thank the Erkkö Foundation and Nordforsk Nordic center of Excellency NordAqua (project number #82845) and University of Helsinki's Doctoral Programme in Microbiology and Biotechnology funding to M.N.A. D.O.A. was supported by a postdoctoral research fellowship from the São Paulo Research Foundation (FAPESP #2018/01563-2). We thank Biocenter Kuopio for the use of their facilities for molecular modeling and MD simulations. We thank the DNA Sequencing and Genomics laboratory, Institute of Biotechnology, University of Helsinki for assistance in sequencing, assembly, and annotation of DNA. We thank the Language Center of University of Helsinki for their help in language revisions.

## REFERENCES

- (1) Nyberg, P.; Ylipalosaari, M.; Sorsa, T.; Salo, T. Trypsins and their role in carcinoma growth. *Exp. Cell Res.* **2006**, *312*, 1219–1228.
- (2) Hockla, A.; Radisky, D. C.; Radisky, E. S. Mesotrypsin promotes malignant growth of breast cancer cells through shedding of CD109. *Breast Cancer Res. Treat.* **2010**, *124*, 27–38.
- (3) Hockla, A.; Miller, E.; Salameh, M. d. A.; Copland, J. A.; Radisky, D. C.; Radisky, E. S. PRSS3/mesotrypsin is a therapeutic target for metastatic prostate cancer. *Mol. Cancer Res.* **2012**, *10*, 1555–1566.
- (4) Jiang, G.; Cao, F.; Ren, G.; Gao, D.; Bhakta, V.; Zhang, Y.; Cao, H.; Dong, Z.; Zang, W.; Zhang, S.; et al. PRSS3 promotes tumour growth and metastasis of human pancreatic cancer. *Gut* **2010**, *59*, 1535–1544.
- (5) Ma, H.; Hockla, A.; Mehner, C.; Coban, M.; Papo, N.; Radisky, D. C.; Radisky, E. S. 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*, 1844.
- (6) Schilling, O.; Biniossek, M. L.; Mayer, B.; Elsässer, B.; Brandstetter, H.; Goettig, P.; Stenman, U.-H.; Koistinen, H. Specificity profiling of human trypsin-isoenzymes. *Biol. Chem.* **2018**, *399*, 997–1007.
- (7) Katona, G.; Berglund, G. I.; Hajdu, J.; Gráf, L.; Szilágyi, L. Crystal structure reveals basis for the inhibitor resistance of human brain trypsin. *J. Mol. Biol.* **2002**, *315*, 1209–1218.
- (8) Sahin-Tóth, M. Human mesotrypsin defies natural trypsin inhibitors: from passive resistance to active destruction. *Protein Pept. Lett.* **2005**, *12*, 457–464.
- (9) Salameh, M. d. A.; Soares, A. S.; Hockla, A.; Radisky, E. S. Structural basis for accelerated cleavage of bovine pancreatic trypsin inhibitor (BPTI) by human mesotrypsin. *J. Biol. Chem.* **2008**, *283*, 4115–4123.
- (10) Wu, P.; Weisell, J.; Pakkala, M.; Peräkyä, M.; Zhu, L.; Koistinen, R.; Koivunen, E.; Stenman, U. H.; Närvi, A.; Koistinen, H. Identification of novel peptide inhibitors for human trypsins. *Biol. Chem.* **2010**, *391*, 283–293.
- (11) Kayode, O.; Huang, Z.; Soares, A. S.; Caulfield, T. R.; Dong, Z.; Bode, A. M.; Radisky, E. S. Small molecule inhibitors of mesotrypsin from a structure-based docking screen. *PLoS One* **2017**, *12*, No. e0176694.
- (12) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770–803.
- (13) Dittmann, E.; Gugger, M.; Sivonen, K.; Fewer, D. P. Natural Product Biosynthetic Diversity and Comparative Genomics of the Cyanobacteria. *Trends Microbiol.* **2015**, *23*, 642–652.
- (14) Jones, M. R.; Pinto, E.; Torres, M. A.; Dörr, F.; Mazur-Marzec, H.; Szubert, K.; Tartaglione, L.; Dell'Aversano, C.; Miles, C. O.; Beach, D. G.; et al. CyanoMetDB, a comprehensive public database of secondary metabolites from cyanobacteria. *Water Res.* **2021**, *196*, 117017.
- (15) Ersmark, K.; Del Valle, J. R.; Hanessian, S. Chemistry and biology of the aeruginosin family of serine protease inhibitors. *Angew. Chem., Int. Ed. Engl.* **2008**, *47*, 1202–1223.
- (16) Murakami, M.; Ishida, K.; Okino, T.; Okita, Y.; Matsuda, H.; Yamaguchi, K. Aeruginosins 98-A and B, trypsin inhibitors from the blue-green alga *Microcystis aeruginosa* (NIES-98). *Tetrahedron Lett.* **1995**, *36*, 2785–2788.
- (17) Ishida, K.; Okita, Y.; Matsuda, H.; Okino, T.; Murakami, M. Aeruginosins, protease inhibitors from the cyanobacterium *Microcystis aeruginosa*. *Tetrahedron* **1999**, *55*, 10971–10988.
- (18) Fewer, D. P.; Jokela, J.; Paukku, E.; Österholm, J.; Wahlsten, M.; Permi, P.; Aitio, O.; Rouhiainen, L.; Gomez-Saez, G. V.; Sivonen, K. New structural variants of aeruginosin produced by the toxic bloom forming cyanobacterium *Nodularia spumigena*. *PLoS One* **2013**, *8*, No. e73618.
- (19) May, D. S.; Crnkovic, C. M.; Krunić, A.; Wilson, T. A.; Fuchs, J. R.; Orjala, J. E. 15N Stable Isotope Labeling and Comparative Metabolomics Facilitates Genome Mining in Cultured Cyanobacteria. *ACS Chem. Biol.* **2020**, *15*, 758–765.



- (20) Ishida, K.; Welker, M.; Christiansen, G.; Cadel-Six, S.; Bouchier, C.; Dittmann, E.; Hertweck, C.; Tandeau de Marsac, N. Plasticity and evolution of aeruginosin biosynthesis in cyanobacteria. *Appl. Environ. Microbiol.* **2009**, *75*, 2017–2026.
- (21) Carroll, A. R.; Buchanan, M. S.; Edser, A.; Hyde, E.; Simpson, M.; Quinn, R. J. Dysinosins B–D, Inhibitors of Factor VIIa and Thrombin from the Australian Sponge *Lamellodysidea chlorea*. *J. Nat. Prod.* **2004**, *67*, 1291–1294.
- (22) Schorn, MA; Jordan, PA; Podell, S.; Blanton, JM; Agarwal, V.; Biggs, JS; Allen, EE; Moore, BS; McFall-Ngai, M. J.; Donia, M.; et al. Comparative Genomics of Cyanobacterial Symbionts Reveals Distinct, Specialized Metabolism in Tropical Dysideidae Sponges. *mBio* **2019**, *10*, No. e00821.
- (23) Ishida, K.; Christiansen, G.; Yoshida, W. Y.; Kurmayer, R.; Welker, M.; Valls, N.; Bonjoch, J.; Hertweck, C.; Börner, T.; Hemscheidt, T.; et al. Biosynthesis and Structure of Aeruginoside 126A and 126B, Cyanobacterial Peptide Glycosides Bearing a 2-Carboxy-6-Hydroxyoctahydroindole Moiety. *Chem. Biol.* **2007**, *14*, 565–576.
- (24) Sandler, B.; Murakami, M.; Clardy, J. Atomic Structure of the Trypsin-Aeruginosin 98-B Complex. *J. Am. Chem. Soc.* **1998**, *120*, 595–596.
- (25) Hanessian, S.; Tremblay, M.; Petersen, J. F. W. The N-acyloxyiminium ion aza-Prins route to octahydroindoles: total synthesis and structural confirmation of the antithrombotic marine natural product oscillarin. *J. Am. Chem. Soc.* **2004**, *126*, 6064–6071.
- (26) Mahlstedt, S.; Fielding, E. N.; Moore, B. S.; Walsh, C. T. Prephenate decarboxylases: A new prephenate-utilizing enzyme family that performs nonaromatizing decarboxylation en route to diverse secondary metabolites. *Biochemistry* **2010**, *49*, 9021–9023.
- (27) Qiu, X.; Zhu, W.; Wang, W.; Jin, H.; Zhu, P.; Zhuang, R.; Yan, X. Structural and functional insights into the role of a cupin superfamily isomerase in the biosynthesis of Choi moiety of aeruginosin. *J. Struct. Biol.* **2019**, *205*, 44–52.
- (28) Qiu, X.; Wei, Y.; Zhu, W.; Fu, J.; Duan, X.; Jin, H.; Zhu, P.; Zhou, C.; Yan, X. Structural and functional investigation of AerF, a NADPH-dependent alkenal double bond reductase participating in the biosynthesis of Choi moiety of aeruginosin. *J. Struct. Biol.* **2020**, *209*, 107415.
- (29) Fujii, K.; Sivonen, K.; Adachi, K.; Noguchi, K.; Shimizu, Y.; Sano, H.; Hirayama, K.; Suzuki, M.; Harada, K.-i. Comparative study of toxic and non-toxic cyanobacterial products: A novel glycoside, suomilide, from non-toxic *Nodularia spumigena* HKVV. *Tetrahedron Lett.* **1997**, *38*, 5529–5532.
- (30) Welker, M.; Von Döhren, H. Cyanobacterial peptides — Nature's own combinatorial biosynthesis. *FEMS Microbiol. Rev.* **2006**, *30*, 530–563.
- (31) Pluotno, A.; Carmeli, S. Banyasin A and banyasides A and B, three novel modified peptides from a water bloom of the cyanobacterium *Nostoc* sp. *Tetrahedron* **2005**, *61*, 575–583.
- (32) Schindler, C. S.; Stephenson, C. R. J.; Carreira, E. M. Enantioselective Synthesis of the Core of Banyaside, Suomilide, and Spumigin HKVV. *Angew. Chem., Int. Ed.* **2008**, *47*, 8852–8855.
- (33) Auerbach, D.; Yan, F.; Zhang, Y.; Müller, R. Characterization of an Unusual Glycerate Esterification Process in Vioprolide Biosynthesis. *ACS Chem. Biol.* **2018**, *13*, 3123–3130.
- (34) Roach, J. C.; Wang, K.; Gan, L.; Hood, L. The molecular evolution of the vertebrate trypsinogens. *J. Mol. Evol.* **1997**, *45*, 640–652.
- (35) Alloy, A. P.; Kayode, O.; Wang, R.; Hockla, A.; Soares, A. S.; Radisky, E. S. Mesotrypsin Has Evolved Four Unique Residues to Cleave Trypsin Inhibitors as Substrates\*. *J. Biol. Chem.* **2015**, *290*, 21523–21535.
- (36) Cohen, I.; Coban, M.; Shahar, A.; Sankaran, B.; Hockla, A.; Lacham, S.; Caulfield, T. R.; Radisky, E. S.; Papo, N. Disulfide engineering of human Kunitz-type serine protease inhibitors enhances proteolytic stability and target affinity toward mesotrypsin. *J. Biol. Chem.* **2019**, *294*, S105–S120.
- (37) Andreassen, P. A.; Egelund, R.; Petersen, H. H. The plasminogen activation system in tumor growth, invasion, and metastasis. *Cell. Mol. Life Sci.* **2000**, *57*, 25–40.
- (38) Copeland, R. A. The drug–target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discovery* **2016**, *15*, 87–95.
- (39) Kozłowski, J. M.; Fidler, I. J.; Campbell, D.; Xu, Z. L.; Kaighn, M. E.; Hart, I. R. Metastatic behavior of human tumor cell lines grown in the nude mouse. *Cancer Res.* **1984**, *44*, 3522.
- (40) Cohen, I.; Kayode, O.; Hockla, A.; Sankaran, B.; Radisky, D. C.; Radisky, E. S.; Papo, N. Combinatorial protein engineering of proteolytically resistant mesotrypsin inhibitors as candidates for cancer therapy. *Biochem. J.* **2016**, *473*, 1329–1341.
- (41) Dillekås, H.; Rogers, M. S.; Straume, O. Are 90% of deaths from cancer caused by metastases? *Cancer Med.* **2019**, *8*, 5574–5576.
- (42) Kotai, J. *Instructions for Preparation of Modified Nutrient Solution Z8 for Algae*; Norwegian Institute for Water Research: Oslo, 1972; Vol. 11, p 5.
- (43) Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J. F.; Guindon, S.; Lefort, V.; Lescot, M.; et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **2008**, *36*, W465.
- (44) Fewer, D. P.; Jokela, J.; Rouhiainen, L.; Wahlsten, M.; Koskeniemi, K.; Stal, L. J.; Sivonen, K. The non-ribosomal assembly and frequent occurrence of the protease inhibitors spumigins in the bloom-forming cyanobacterium *Nodularia spumigena*. *Mol. Microbiol.* **2009**, *73*, 924–937.
- (45) Koistinen, H.; Koistinen, R.; Zhang, W.-M.; Valmu, L.; Stenman, U.-H. Nexin-1 inhibits the activity of human brain trypsin. *Neuroscience* **2009**, *160*, 97–102.
- (46) Pettaway, C. A.; Pathak, S.; Greene, G.; Ramirez, E.; Wilson, M. R.; Killion, J. J.; Fidler, I. J. Selection of highly metastatic variants of different human prostatic carcinomas using orthotopic implantation in nude mice. *Clin. Cancer Res.* **1996**, *2*, 1627.
- (47) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296.
- (48) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221–234.
- (49) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
- (50) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.
- (51) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.
- (52) Halgren, T. A. Identifying and characterizing binding sites and assessing drug ability. *J. Chem. Inf. Model.* **2009**, *49*, 377–389.
- (53) Bowers, K. J.; Chow, D. E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D. Scalable algorithms for molecular dynamics simulations on commodity clusters. *SC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*; ACM/IEEE, 2006; Vol. 43.