Ahmed, Muhammad N.; Wahlsten, Matti; Jokela, Jouni; Nees, Matthias; Stenman, Ulf-Håkan; Alvarenga, Danillo 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:

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

Muhammad N. Ahmed, Matti Wahlsten, Jouni Jokela, Matthias Nees, Ulf-Håkan Stenman, Danillo 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/acscambio.1c00611

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 IC₅₀ values of 4.7 and 11.5 nM, respectively, while trypsin-1 was inhibited with an IC₅₀ 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.¹ The human genome encodes three trypsin isoenzymes (trypsin-1, -2, and -3) that are structurally and functionally highly similar.¹ 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.²⁻⁵ Thus, trypsin-3 is considered as an emerging therapeutic target for these cancers.²⁻⁵ Although there are subtle differences between the substrate specificities of human trypsins,⁶ the high sequence identity between trypsin isoforms extends to the active sites and makes selective targeting of trypsin-3 challenging.⁷⁻¹¹ Thus, current small-molecule inhibitors of trypsin-3 target all three trypsin isoenzymes with equal potency.¹¹

Natural products have served as an inspiration in the development of new selective small-molecule inhibitors.¹² Indeed, most drugs that target proteases are based on or inspired by natural products.¹² Cyanobacteria produce an array of natural products with potent serine protease inhibition activity.¹³,¹⁴ Aeruginosins are a diverse family of linear peptides that are produced by a broad range of cyanobacteria.¹⁴ They are potent inhibitors of serine proteases in low micromolar to low nanomolar concentrations.¹⁵ 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, argininol, agmatine, 1-amino-2-(N-amidino-Δ⁴-pyrrolinyl)-ethyl (Aep) moiety¹⁶⁻¹⁹ (Figure 1, panel a). The N-terminus typically consists of phenyl lactic acid derivatives²⁰ but may also contain hexanoic acid²¹ or 2-O-methyl-3-sulfoglycric acid (Mgs)²¹,²² (Figure 1, panel a). Aeruginosins may also be glycosylated, sulfonated, and halogenated.¹³,¹⁸,¹⁹,⁵³ 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
shown by the X-ray crystallography of aeruginosin–protease complexes.15,24,25 Aeruginosins are synthesized on nonribosomal peptide synthetase (NRPS) enzyme complexes encoded in the aeruginosin (aer) biosynthetic gene cluster (BGC).18,20,22,23 Aeruginosin biosynthesis begins with the activation of a monocarboxylic acid or fatty acid by the loading module AerA.20,23 AerB is responsible for adding a hydrophobic D-amino acid.20,23 AerD, AerE, and AerF work in a cascading manner to synthesize and supply Choi to AerG.23,26–28 AerG adds a Choi
moiety and encodes an off-loading module.\textsuperscript{20,23} Variation in the loading and off-loading mechanisms of aeruginosin synthetases contributes to the observed structural diversity of the natural product family.\textsuperscript{18−20,22,23} Aeruginosin BGCs also encode an assortment of tailoring enzymes that promote structural diversification, including halogenation, acylation, and glycosylation.\textsuperscript{18,19,22,23}

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

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 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.
Suomilide and aeruginosins have been proposed to share a common biosynthetic logic based on the similarities in the chemical structure (Figure 1, panel a).\textsuperscript{15,29,32} 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 \( (\text{suom}) \) 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.\textsuperscript{20,23} 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 glyceraldehyde,\textsuperscript{33} 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 glyceraldehyde.\textsuperscript{22} 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.\textsuperscript{20,23} 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-acetyltransferase (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 Aep through a presently unknown biosynthetic mechanism.\textsuperscript{20}

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.\textsuperscript{30–32}

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.\textsuperscript{18,20,22,23} The off-loading modules allow the release of the tetrapeptides with C-terminal arginine derivatives that include argininal, argininol, agmatine, 1-amidino-2-ethoxy-3-aminopiperidine, or more rarely an Aep moiety.\textsuperscript{16–19} 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.\textsuperscript{15}

Suomilide Inhibits Human Trypsin Isoenzymes. Suomilide has been reported to inhibit an unspecified but most likely non-human trypsin isoform with an \( IC_{50} \) value of 1.8 \( \mu \text{M} \).\textsuperscript{29} 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.\textsuperscript{34} We purified 6.55 mg of suomilide from N. sphaerocarpa UHCC 0038 (Supporting Information Figures 2 and 3) and tested its ability to inhibit human trypsins different serine proteases could drive the pronounced chemical and genetic diversification of this variable family of protease inhibitors.\textsuperscript{15}
recombinant human trypsin-1, -2, and -3 using a chromogenic trypsin substrate. Suomilide inhibited trypsin-1, -2, and -3 with IC\textsubscript{50} values of 104 ± 7, 4.7 ± 0.6, and 11.5 ± 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.\textsuperscript{11} Indeed, suomilide 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.\textsuperscript{11} 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 suomilide and also potently inhibit plasma kallikrein.\textsuperscript{10} 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.\textsuperscript{8} When Gly193 in human trypsin-1 was mutated to Arg193, it conferred a trypsin-3-like resistance to inhibitors.\textsuperscript{9,35} On the other hand, proteinaceous inhibitors have been modified to effectively target trypsin-3.\textsuperscript{36} 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 

Figure 4. Molecular dynamics (MD) simulation of trypsin-1 and -3 interactions with suomilide 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 suomilide is highlighted in red and the Mgs moiety in gray. An almost identical interaction was obtained with trypsin-1.\textsuperscript{9,5} Lys224 in trypsin-1. (b) Root mean square deviation plot of suomilide 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 suomilide in MD simulation. Only amino acids whose interactions exceeded 0.5 or which are shown in (a) are shown. (d) Kinetic analysis of the suomilide interaction with trypsin-1 and trypsin-3 by surface plasmon resonance (Biacore).
trypsin-1 and -2. 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 suomilide also inhibits human plasmin, plasma kallikrein, and Factor Xa. These were inhibited at IC50 values of 7.6 ± 2.3, 82 ± 24, and 114 ± 2 nM, respectively. However, human urokinase was inhibited at an IC50 value of 7.1 ± 1.36 μM, and human thrombin was minimally inhibited (IC50 >12.5 μM). This contrasts somewhat with the previous study, showing that suomilide 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.

Suomilide and Trypsin-3 Interaction. To explain the selectivity of suomilide in inhibiting trypsin-2 and -3 over trypsin-1, we studied the potential binding mode of suomilide to trypsin-1 and trypsin-3, for which crystal structures are available.7,9 The suomilide structure was modeled, and molecular docking (Glide XP) was performed to obtain the initial docking poses. While suomilide 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 × 1 microseconds. Surprisingly, for both trypsin-1 and -3, docking poses with suomilide remained almost intact throughout the simulation. In both trypsins, the hydrogen bonds between Asp189, Ser190, and Gly219 and the Aeap moiety of suomilide were conserved.10−12 These were inhibited at an IC50 value of 7.1 ± 1.36 μM, and human thrombin was minimally inhibited (IC50 >12.5 μM). This contrasts somewhat with the previous study,29 showing that suomilide 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.

Suomilide 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.2−5 We tested the bioactivity of suomilide 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 μM suomilide significantly inhibited the invasion of trypsin-3-producing aggressive metastatic PC-3M prostate cancer cells.37,38 Suomilide 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.3,38 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.3,39 Of note, the knockdown of trypsin-1 in PC-3M cells did not have such an effect.3 Thus, cell invasion in this model appears to be regulated by the trypsin-3 activity. The functional studies support the potential of suomilide as a lead compound for the development of an anti-invasive cancer drugs. 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.31

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 suomilide 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 suomilide 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 suomilide 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 HKV) was grown in a photon irradiance of 15 μmol m−2 s−1 in 100 mL of 2XKS culture medium, obtained by adding 0.88% NaCl and 0.38% MgSO4·7H2O to Z8 media,42 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 7000 g 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 μL of 425−600 μm beads and 100 μL of 700−1180 μm of beads (Scientific Industries) were added to the pellet after addition of 600 μL of SP1 buffer (provided with the kit) to 50 μg 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−1 for 20 s, repeating this step twice. 5 μL 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 μg of genomic DNA. The quality and concentration of isolated DNA samples were assessed using an Agilent TapeStation (Agilent Technologies) and a NanoDrop 2000 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 126.1063 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 NRRPs and the function of the genes were also predicted by using the NCBI CD Database (CDD). The substrate specificity of the NRPS adenylation domains was predicted using the NRPSpredictor2 online version and CDD searches.

**Phylogenetic Trees.** Phylogenetic trees based on suomilide and aeruginosin biosynthetic proteins were constructed using PhyML to test the relationship between suomilide and aeruginosin. An aeruginosin biosynthetic proteins were constructed using PhyML to using a FastPrep cell disrupter (Bio 101, Thermo Electron Corporation, PTFE, VWR International). Sample analysis was performed with an Eppendorf AG) at room temperature and was stored at 4 °C, a desolvation gas flow of 1000 L h⁻¹, 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 Suomilide.** *N. sphaerocarpa* UHCC 0038 was grown in 210 L of Z8XS culture medium, centrifuged at 7 000 for 7 min and then lyophilized (Christ Beta 2–8 LS plus, 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 Silencer Crusher 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-Davidson 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 S 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 suomilide were dried under a nitrogen gas stream, and suomilide was purified by semi-preparative HPLC combined with MS using a Luna C8(2) (150 × 10.00 mm, 5 µm, 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 µL with a flow rate of 5 mL min⁻¹. 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 suomilide preparation contained only very minor amounts of other suomilide variants (Supporting Information Figure 2). The molecular formula of C₄₅H₇₃N₇O₁₉S was established based on the UPLC–QTOF analysis (Supporting Information Figure 2). *N. sphaerocarpa* UHCC 0038 does not synthetize secondary metabolites such as nodularins, spumigins, aeruginosins, and nodulaepetins, typical for *N. spumigena* strains and some other *Nodularia* species (Supporting Information Figure 3). To further assess the purity, suomilide was evaporated to dryness and subjected to nuclear magnetic resonance (NMR).

**Concentration and Purity Determination through NMR.** The purification yielded 7.7 mg of suomilide from *N. sphaerocarpa* UHCC 0038, as determined by weighing. Suomilide was dissolved in 0.5 mL of DMSO-d₆ (dimethyl sulfoxide). Commercial dipeptide Tyr–Val (Sigma-Aldrich) was dissolved to 1/10 of the volume of the suomilide solution in an equimolar ratio with suomilide. The comparison of integrated proton spectrum signals from suomilide and Tyr–Val dipeptide gave a 10% lower concentration for suomilide than that obtained by weighing. ¹³C HSQC (heteronuclear single quantum coherence) showed several minor signals in the carbohydrate region in addition to two minor signals in the amnoneric region (Supporting Information Figure 4). This indicates that the suomilide preparation contains suomilide variants in which fatty acid residues are different or located in different positions in the glucose unit. The comparison of suomilide and variant signal intensities yielded an estimate of 85% suomilide and 15% suomilide variants. The weight determined by NMR (6.55 mg) was used in this study.

One-dimensional ¹H and two-dimensional CH-edited ¹³C-HSQC NMR spectra were measured in a Bruker AVANCE III HD 800 NMR spectrometer, equipped with a cryogenically cooled triple-resonance (¹H, ¹³C, and ¹⁵N) probe head and z-gradient at 30 °C (303 K). For ¹H spectra, 16 complex points, corresponding to an acquisition time of 1.28 s, were collected using 16 transients. For ¹³C-HSQC spectra were collected with 24 transients per FID and 256/1024 complex points in t₁, t₂ 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* activated to trypsin-1, -2, and -3 as described previously. The enzyme inhibitory activity of suomilide
was determined in 96-well plates by adding 10 μL of diluted suomilide (0.21–3400 nM, diluted in ultra-purified water, containing up to 0.027% DMSO). 15 μL of 0.1% bovine serum albumin in 50 mM Tris-buffer saline (BSA/TBS), and 25 μL of each trypsin (1.88 nM). 10 μL 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 suomilide with trypsin isoenzymes prior to the addition of the substrate was required for maximal inhibition (data not shown). Thus, suomilide was preincubated with trypsins at room temperature for 30 min before the addition of 50 μL of 0.2 mM chromogenic substrate S-2222 (ChromogeniX). Except for urokinase from human kidney cells (ChromogeniX S-2302 (ChromogeniX); Factor Xa (Sigma) (3 nM) (ChromogeniX); plasma KLK (Sigma) (20 nM) with substrate which the absorbance of the control reaction increased linearly over 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 μL of BSA/TBS with 0.9% NaCl was used. Nine 4-fold dilutions of suomilide were prepared, with 12.5 μM being the highest final concentration. The IC50 values were estimated using Quest Graph, and the standard error was calculated using MS Excel 2013.

**Effect of Suomilide on the Invasion and Proliferation of PC-3M Cells.** PC-3M prostate cancer cells46 were cultured in Ham’s F12 Nutrient Mix with GlutaMAX (Gibco), supplemented with 10% fetal calf serum (FCS), 100 IU mL−1 penicillin, 100 μg mL−1 streptomycin, and 0.15% sodium bicarbonate at 37 °C in a humidified, CO2-controlled (5%) incubator. For the invasion assay, PC-3M cells (1.45 × 103) in a serum-free medium containing 0 or 3.3 μM suomilide were added into polycarbonate membrane Transwell inserts with 8.0 μm pore size (Thermo Fisher) covered with 100 μL 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 μL of the culture medium with 10% FCS and incubated for 4 days at 37 °C in a humidified, CO2-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 μM suomilide in 100 μL of culture media containing 1% FCS. After 3 days, 10 μL/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 suomilide-treated and control cells.

**Molecular Modeling.** All modeling was conducted with Maestro Small-Molecule Drug Discovery Suite 2021-1 (Schrodinger Release 2021-1: Maestro, Schrodinger, LLC, 2021) with an OPLS4 force field unless otherwise stated.47 The figures were prepared with PyMOL 2.4.2 (Schrodinger, LLC). For the docking and MD simulations, we used the PDB structures 2RA3 (trypsin-1) and 1H4W (trypsin-3), which were prepared using Protein Preparation Wizard (default settings with the “cap termini” option).48 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.49 The grid box was defined by SiteMap (the SiteMap region covering the active site was selected for GRID).50 Before docking, suomilide was prepared with LipPrep (default settings) using Epik, and the tautomeric state of suomilide was based on the QM Conformer & Tautomer Predictor results. The MD simulations were conducted with Desmond.53 The systems were solvated in a cubic box (edges 14 Å from the protein) and neutralized with counter ions (2 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 μs (5 × 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 suomilide 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 μL min−1. The kinetics of the suomilide interaction with the trypsins were measured by varying suomilide 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 suomilide concentration separately and overlaying the resulting sensorgrams for each ligand. Association and dissociation rate constants were obtained by the Langmuir global fit model (BioEvaluation 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 suomilide preparation, total ion chromatograms of positive and negative ionizations of the *N. sphaerocarpa* UHCC 0038 methanol extract,13C HSQC spectra of suomilide, antiSMASH analysis of the *N. sphaerocarpa* UHCC 0038 genome, and function prediction of the suomilide biosynthetic gene cluster from *N. sphaerocarpa* UHCC 0038 and comparison between suomilide and aeruginosin biosynthetic gene cluster proteins (PDF)

**Accession Codes**

The accession number of the complete genome sequence of *N. sphaerocarpa* UHCC 0038: CP006140.

**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.k.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; Email: david.fewer@helsinki.fi
We would like to thank A. Löfhjelm and L. Saari for excellent technical assistance. This work was supported by a Sigrid Juselius Foundation grant to H.K. and the Academy of Finland funding (321809) to T.S. We would also like to thank the Erkko 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.

## ACKNOWLEDGMENTS

The authors declare no competing financial interest.

## REFERENCES

(22) Schorn, MA; Jordan, PA; Poddell, 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 Dasyideae Sponges. mBio 2019, 10, e00821.