Discovery of Thanafactin A, a Linear Proline-containing Octa-Lipopeptide from *Pseudomonas* sp. SH-C52, Motivated by Genome Mining

Norbert Kirchner,†,‡ Carolina Cano-Prieto,†,‡ Anna-Christina Schulz-Fincke,‡ Michael Gütschow,‡ Nico Ortlieb,§,ǁ Julia Moschny,§ Timo H. J. Niedermeyer,§,¶ Jeannie Horak,¶,‖ Julia Moschny,§ Timo H. J. Niedermeyer,¶,‖ Harald Gross†,‡,*

Michael Lämmerhofer,¹ Menno van der Voort,○ Jos M. Raaijmakers,^,# Harald Gross†,‡,*

†Pharmaceutical Institute, Department of Pharmaceutical Biology, University of Tübingen, 72076 Tübingen, Germany

‡Pharmaceutical Institute, Pharmaceutical & Medicinal Chemistry, University of Bonn, 53121 Bonn, Germany

§Department of Microbiology and Biotechnology, Interfaculty Institute of Microbiology and Infection Medicine, Eberhard Karls University Tübingen, 72076 Tübingen, Germany

¶German Centre for Infection Research (DZIF), partner site Tübingen, 72076 Tübingen, Germany

¹Pharmaceutical Institute, Department of Pharmaceutical Analysis and Bioanalysis, University of Tübingen, 72076 Tübingen, Germany

‖Dr. von Hauner Children’s Hospital, Department of Metabolic and Nutritional Medicine, University of Munich Medical Center, Campus Innenstadt, 80337 Muenchen, Germany

○Laboratory of Phytopathology, Wageningen University, Netherlands

^Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands

#Institute of Biology, Leiden University, Leiden, The Netherlands

These authors contributed equally

*Corresponding Author: Harald Gross*  
Tel: +49-7071-2976970  
Fax: +49-7071-295250  
E-mail: harald.gross@uni-tuebingen.de
ABSTRACT

Genome mining of the bacterial strains *Pseudomonas* sp. SH-C52 and *Pseudomonas fluorescens* DSM 11579 showed that both strains contained a highly similar gene cluster encoding an octa-modular nonribosomal peptide synthetase (NRPS) system which was not associated with a known secondary metabolite. Insertional mutagenesis of a NRPS component followed by comparative profiling led to the discovery of the corresponding novel linear octalipopeptide thanafactin A, which was subsequently isolated and its structure determined by 2D NMR and further spectroscopic and chromatographic methods. In bioassays, thanafactin A exhibited weak protease inhibitory activity and was found to modulate swarming motility in a strain-specific manner.
Pseudomonas sp. SH-C52 is a plant beneficial bacterium that was isolated from soil that was naturally suppressive to Rhizoctonia solani, a fungal root pathogen of various economically important food crops. Originally, the strain attracted attention due to its antifungal properties, but subsequent chemical and biological screenings revealed an intriguingly diverse spectrum of other bioactivities against Gram-positive bacteria and oomycetes.

In order to understand the taxonomic position as well as the modes of action of SH-C52, its genome was sequenced in parallel in 2015. The sequencing data enabled genome-wide phylogenetic analyses which hinted that SH-C52 is taxonomically most closely related with Pseudomonas mediterranea and Pseudomonas corrugata strains (Figure S1). However, in pairwise comparisons, the digital DNA-DNA hybridization (dDDH) values (d4) between strain SH-C52 and its closest related type strains ranged from 32.0 to 52.9 % (Table S1). Because these values are below the species threshold of 70%, strain SH-C52 represents a new Pseudomonas species. The combined manual and automated bioinformatics analysis of the genome predicted 14 putative biosynthetic gene clusters (BGCs) coding for secondary metabolites (Table S2). Three orphan NRPS-based gene clusters thereof were already investigated in subsequent follow-up studies, employing mass spectrometry- and genomics-driven approaches, which led to the discovery of the lipopeptides thanapeptin and thanamycin as well as the cyclocarbamates brabantamides A-C. However, the chemical structure encoded by an octa-modular NRPS gene cluster remained elusive.

So far, 15 linear and 10 cyclic octalipopeptides have been reported from various Pseudomonads which can be categorized in seven subfamilies (Table 1). Concerning their biological activities, the linear forms were shown to be involved in virulence, swarming, biofilm formation, the increase of water availability and iron acquisition, while the cyclic forms attracted attention due to their cytotoxic, antifungal and anti-oomycetal properties.

The octa-modular NRPS-based BGC of this study appeared to be quite unique and only present in strain P. sp. SH-C52 and the closely related strain Pseudomonas fluorescens DSM 11579. The fact that it was predicted to have no similarity to other known Pseudomonas octalipopeptides prompted us to re-investigate these two strains, using a comparative metabolite screening approach.

Herein, we report the prediction, the discovery process, the isolation and the structure elucidation of the structurally novel octalipopeptide, designated thanafactin A, and its evaluation in bioassays.

**RESULTS AND DISCUSSION**
In silico analysis of the thanafactin gene cluster in P. sp. SH-C52 and P. fluorescens DSM 11579. The complete thanafactin (thf) NRPS gene cluster was not directly apparent in the SH-C52 genome. At first, we detected the start of the thf BGC on contig 239 (syn. 1_34). To complicate matters further, the region 81644-87657 of contig 239 was not recognized as an NRPS by automated web tools since it was misannotated as two hypothetical proteins and lengthy sequence gaps. Manual inspection of contig 239 revealed that it carried a starter (C₅-A-T) and four elongation (C-A-T) modules. However, the second gene (thfB) appeared to be truncated, as the 3'-end of the gene, belonging to the fifth module, coded only for a partial T-domain. Furthermore, the second gene lacked a stop codon. Due to these facts and the absence of a chain termination domain such as a thioesterase (TE) domain, a terminal condensation-like (C_T, C*) domain, a terminal reductase (R) domain or a stand-alone SurE-like PBP-type TE was absent, we hypothesized that this NRPS locus did not simply code for a pentalipopeptide, but rather represents only a portion of a larger lipopeptide. On contig 332 (syn. 1_35), we found a candidate NRPS cluster, containing three NRPS C-A-T modules including a terminal tandem TE domain, which represents a characteristic termination domain for Pseudomonas lipopeptide encoding BGCs. Notably, manual inspection revealed that the candidate gene lacks a start codon and that the length of the first encoded C/E domain was shorter than expected, which in turn pointed to a sequencing gap. Using PCR (Figure S2), we were able to bridge the gap between the two contigs and thus confirmed that SH-C52 possesses the potential to produce an octalipopeptide. The complete thf BGC consists of three genes, namely the tri-modular thfA (9.6 kbp), and the penta-modular thfB (17.8 kbp) (Figure 1). Downstream of thfB, a major facilitator transporter gene (thfC) is present. The thfABC BGC was also found in one contig of the recently sequenced genome of P. fluorescens DSM 11579. The genes are organized in the same order and showed a high level of identity at the amino acid level, ranging from 76-80 % (Figure 1). However, considering the genetic context of the BGC, it appears to be embedded in a different regulatory network than in strain SH-C52 (Table S3 and S4).
Figure 1. (a) Organization and comparison of the thanafactin biosynthetic gene cluster of *Pseudomonas fluorescens* DSM 11579 and *Pseudomonas* sp. SH-C52; structural NRPS genes are shown in red, while genes encoding transporters are depicted in blue color. The percentage of identity on the DNA and protein level, respectively, is indicated in grey bars. Triangles represent the positions of the disruptions in each thfB-gene obtained by mutagenesis. (b) Organization of the NRPSs encoding thanafactin A and predicted amino acid composition of the lipopeptide backbone.

Each of the NRPS modules consists of a condensation (C), adenylation (A) and thiolation (T) domain, while a terminal TE-TE domain\(^{15}\) was located at the carboxy terminal of ThfB. Manual phylogenetic (Figure S3) and automated\(^{16}\) analyses of the A domains of each module led to the prediction that thanafactin possesses the amino acid sequence Val\(^1\)-Ala\(^2\)-Gln\(^3\)-Ala\(^4\)-Val\(^5\)-Ala\(^6\)-Pro\(^7\)-Thr\(^8\). In order to predict the absolute configuration of each amino residue, the C domains were bioinformatically analyzed in a similar fashion (Figure S4). The analysis showed that the C domains of modules 5, 7 and 8 represent conventional \(^1\)C\(_L\) domains, while the remaining C domains are all classified as combined C/E domains.\(^{17}\) This suggests that Ala\(^4\), Ala\(^6\), Pro\(^7\) and Thr\(^8\) are \(L\)-configured, while all other amino acids appear to be \(D\)-configured. However, as a number of deviations from sequence-based prediction of epimerization from C/E-domains have been reported in the literature,\(^{18}\) the *in silico* prediction provides only a tentative assignment, which is later on validated by chiral amino acid analysis. Furthermore, a lipo-initiation (C\(_{\text{Start}}\))\(^{19}\) domain in module I of ThfA was identified, suggesting that a fatty acid is attached to Val\(^1\). However, because the biosynthesis genes of the lipid moiety are not genetically encoded in the thf BGC, the fatty acid is provided by primary metabolism, most likely in form of a 3-OH fatty acid.\(^{20}\) The exact length and saturation degree could not be predicted, but ranges empirically from 10 to 16 carbon atoms in length in related lipopeptides (Table 1).\(^{21}\) The termination module of ThfB consists of a tandem
TE domain, which is typical for most of the *Pseudomonas* lipopeptides.\(^2\) A phylogenetic analysis (Figure S5) showed that the first TE domain clusters with type I TE domains, while the second TE forms a clade with type II TE domains. Although the type I TE of ThfB clusters with TE domains, which are enabled to perform a macrocyclization by forming a depsipeptide bond between the last amino acid and a Ser/Thr residue,\(^2\) thanafactin was assumed to be a linear compound, particularly because serine or further threonine residues were absent in the backbone of thanafactin.

Overall, the above bioinformatic analysis strongly suggests the formation of a novel linear lipodepsipeptide with the following sequence: 3-OH-FA-\(\text{D/L-Val1-\(\text{D/L-Ala2-\(\text{D/L-Gln3-\(\text{L-Ala4-\(\text{D/L-Val5-\(\text{L-Ala6-\(\text{L-Pro7-\(\text{L-Thr8.}

Depending on the nature of the lipid side chain, the target compound thanafactin was expected to range from 925-993 Da. Due to the unprecedented peptide sequence (Table 1) of the resultant lipopeptide we embarked on the isolation and structural identification of thanafactin to validate our predictions and to determine its bioactivities.

**Table 1.** Fatty Acid and Amino Acid Sequence Alignments of Thanafactin A (I) with Reported Linear and Cyclic *Pseudomonas* Octa-lipopeptides.\(^7,8,23,24\) Dab = 2,4-diaminobutyric acid; cDab = cyclized Dab (cyclic amidine formation with the carbonyl function of the previous amino acid).
Target Compound Identification by Comparative Metabolite Profiling. In order to identify the product of the orphan *thf* BGC, gene disruption followed by comparative metabolite profiling, using LC-MS, was employed. Specifically, the NRPS gene *thfB* was disrupted in both strains by single homologous recombination (Figure 1). Two independent mutants of each strain were selected and the mutations were confirmed by PCR. The resulting four mutants (SHC52-ΔNRPS A1 and B1 / DSM11579-ΔNRPS C2 and E7) showed in the chromatograms the absence of a minor peak in comparison with the corresponding wild type strains (Figure S6). The peak showed a [M+H]⁺ ion at *m/z* 926.6 in the LRESIMS spectrum which was in full agreement with the expected mass range.

Isolation and Structural Analysis of Thanafactin A. Although the target compound (1) was located in the chromatogram and the production of 1 was now apparent from both producer strains, all attempts to isolate 1 in sufficient amounts were initially unsuccessful. To increase the titer of 1, several media formulations, cultivation conditions in combination with three extraction methods (EtOAc, BuOH and Diaion HP-20) were tested and the resultant crude analyzed by LC-MS. SRM medium in combination with Diaion HP-20 extraction turned out to yield the best results. Application of the optimized conditions and procedures enabled the reliable and reproducible isolation of 1 from the fermentation broths of both producer strains, while DSM 11579
showed a slightly higher production. To purify 1, DSM 11579 was fermented under the optimized conditions on a large scale (28.8 L). After sequential column chromatography on Diaion HP-20 followed by RP-HPLC, pure 1 (4.5 mg) was obtained. The upscaling process unveiled also that beside one major thanafactin peak four minor congeners (tentatively assigned as thanafactins B, C, D and E) were present (Figure S7), whose occurrence are commonly attributed to the relaxed substrate specificity of the C₈- or A-domains. Consequently, in the following, the tracked down major peak will be therefore termed thanafactin A.

Thanafactin A (1) showed a [M+H]⁺ ion at m/z 926.5553 in the HRESIMS spectrum (Figure S8), which is consistent with the formula C₄₃H₇₅N₉O₁₃ and required 11 degrees of saturation. Initial evaluation of the ¹H and ¹³C NMR spectra supported the bioinformatics-driven hypothesis that 1 was a lipopeptide. A methylene envelope at δₕ 1.23-1.26 and a terminal methyl group at δₕ 0.85 were indicative of a fatty acid moiety. The peptidic nature of the molecule was evident from resonances for α-CH groups, observed at δₕ 4.10 - 4.70 and by the presence of a cluster of exchangeable deshielded resonances at δₕ 7.6 - 8.3 in the ¹H NMR spectrum (Figure S12). Furthermore, the ¹³C NMR spectrum showed 10 carbonyl signals at δc 170.0 - 174.0 (Figure S13). The planar structure of 1 was assigned after extensive 1D- and 2D NMR studies (Table 2). Initial interpretation of the HSQC-TOCSY and COSY NMR spectra revealed the presence of ten spin systems (Figure 2, S16 and S17). This included the amino acids 1× Thr, 1×Pro, 3×Ala, 2×Val, 1×Glx and a fatty acid. HMBC cross correlations between the side chain hydrogens and their corresponding carbonyl group completed each amino acid moiety.
Table 2. NMR Data (400 MHz, d<sub>6</sub>-DMSO) for Compound 1 (δ in ppm, J in Hz)

<table>
<thead>
<tr>
<th>position</th>
<th>δ&lt;sub&gt;C&lt;/sub&gt;, type</th>
<th>δ&lt;sub&gt;H&lt;/sub&gt;</th>
<th>position</th>
<th>δ&lt;sub&gt;C&lt;/sub&gt;, type</th>
<th>δ&lt;sub&gt;H&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>3OH-FA</td>
<td></td>
<td></td>
<td>Ala4</td>
<td>α</td>
<td>45.9, CH</td>
</tr>
<tr>
<td>1</td>
<td>171.4, C</td>
<td></td>
<td>β</td>
<td>17.2, CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>1.18, m</td>
</tr>
<tr>
<td>2</td>
<td>43.3, CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>2.27, d (5.9)</td>
<td>C=O</td>
<td>170.4, C</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>67.7, CH</td>
<td>3.78, m</td>
<td>NH</td>
<td></td>
<td>8.24, d (7.5)</td>
</tr>
<tr>
<td>4</td>
<td>36.8, CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1.35, m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5a</td>
<td>24.9, CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1.23, m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5b</td>
<td>1.33, m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>28.7, CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1.23, m</td>
<td>Val5</td>
<td>α</td>
<td>57.4, CH</td>
</tr>
<tr>
<td>7</td>
<td>29.1, CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1.24, m</td>
<td>β</td>
<td>30.2, CH</td>
<td>1.98, m</td>
</tr>
<tr>
<td>8</td>
<td>31.2, CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1.24, m</td>
<td>γ&lt;sub&gt;1&lt;/sub&gt;</td>
<td>18.04&lt;sup&gt;a&lt;/sup&gt;, CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>0.82, m</td>
</tr>
<tr>
<td>9</td>
<td>22.1, CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1.26 m</td>
<td>γ&lt;sub&gt;2&lt;/sub&gt;</td>
<td>19.1&lt;sup&gt;b&lt;/sup&gt;, CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>0.86, m</td>
</tr>
<tr>
<td>10</td>
<td>14.0, CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>0.85, m</td>
<td>C=O</td>
<td>171.0, C</td>
<td></td>
</tr>
<tr>
<td>OH</td>
<td>4.68, d (4.5)</td>
<td></td>
<td>NH</td>
<td></td>
<td>7.90,_brdd</td>
</tr>
<tr>
<td>Val1</td>
<td>α</td>
<td>57.2, CH</td>
<td>C=O</td>
<td>171.9, C</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>30.8, CH</td>
<td>1.93, m</td>
<td>NH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>γ&lt;sub&gt;1&lt;/sub&gt;</td>
<td>18.8&lt;sup&gt;a&lt;/sup&gt;, CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>0.86, m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>γ&lt;sub&gt;2&lt;/sub&gt;</td>
<td>19.3&lt;sup&gt;b&lt;/sup&gt;, CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>0.86, m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C=O</td>
<td>171.7, C</td>
<td></td>
<td>Pro7</td>
<td>α</td>
<td>59.1, CH</td>
</tr>
<tr>
<td>NH</td>
<td>7.90 brd</td>
<td></td>
<td>β</td>
<td>29.0, CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1.88, m</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>γ</td>
<td>24.4, CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1.87, m</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>δ</td>
<td>46.5, CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>3.55, m</td>
</tr>
<tr>
<td>Ala2</td>
<td>α</td>
<td>48.3, CH</td>
<td>C=O</td>
<td>172.1, C</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>17.96&lt;sup&gt;a&lt;/sup&gt;, CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>1.20, m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C=O</td>
<td>172.1, C</td>
<td></td>
<td>NH</td>
<td></td>
<td>7.90, brd</td>
</tr>
<tr>
<td>Thr8</td>
<td>α</td>
<td>57.7, CH</td>
<td>C=O</td>
<td>170.7, C</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>66.4, CH</td>
<td>1.42, m</td>
<td>NH</td>
<td></td>
<td>7.69, d (8.6)</td>
</tr>
<tr>
<td>γ</td>
<td>20.4, CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>1.05, d (6.3)</td>
<td>OH</td>
<td></td>
<td>n.o.</td>
</tr>
<tr>
<td>Gln3</td>
<td>α</td>
<td>52.3, CH</td>
<td>COOH</td>
<td>170.7, C</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>28.0, CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>1.72, m</td>
<td>NH</td>
<td></td>
<td>7.69, d (8.6)</td>
</tr>
<tr>
<td>γ</td>
<td>31.4, CH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>2.07, dd (7.6, 8.1)</td>
<td>OH</td>
<td></td>
<td>n.o.</td>
</tr>
<tr>
<td>C=O</td>
<td>170.2, C</td>
<td></td>
<td>NH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C=O&lt;sub&gt;γ&lt;/sub&gt;</td>
<td>173.7, C</td>
<td>7.83, d (9.2)</td>
<td>NH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>6.75, brs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.22, brs</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>OH-FA = 3-hydroxy fatty acid; n.o.: not observed; <sup>a,b</sup>assignments may be interchangeable.
The NOESY cross correlations between 3-OH-FA-H2-2/Val1-NH; Val1-Hα/Ala2-NH; Ala2-Hα/Gln3-NH; Gln3-Hα/Ala4-NH; Val5-Hα/Ala6-NH; Pro7-Hα/Thr8-NH in combination with the interpretation of the tandem mass spectrum (Figure S18), revealed the sequence of the amino acids to be 3-OH-FA-Val1-Ala2-Gln3-Ala4-Val5-Ala6-Pro7-Thr8, which was in full agreement with the sequence proposed based on bioinformatics.

**Figure 2.** Key NMR correlations of 1. Bold lines indicate COSY or HSQC-TOCSY correlations, while arrows and dashed lines show key HMBC and NOESY cross correlations, respectively.

Subtraction of the C, N, and O atoms accounted for by the eight identified amino acid residues from the molecular formula of thanafactin A, showed that the 3-OH-fatty acid had to consist of 10 carbons and two oxygen atoms. The $^{13}$C and HSQC NMR data revealed that the remaining 10 carbon atoms consist of one methyl ($\delta_C$ 14.0), one carbinol methine ($\delta_C$ 67.7), one ester carbonyl ($\delta_C$ 171.4), and seven aliphatic methylene carbons, suggesting a linear hydroxy-decanoic acid (Table 2). Correlations in the COSY spectrum delineated a connected spin system for protons at $\delta_H$ 2.27 (H$_2$-2) and $\delta_H$ 3.78 (H-3). The observed long-range $^1$H-$^{13}$C coupling between H$_2$-2 and C-1 subsequently identified the fatty acid fragment as 3-hydroxy decanoic acid (HDA) (Figure 2). The assigned residues accounted for the required 11 double bond equivalents, thus thanafactin A has to be linear.

With the planar structure of 1 determined, the absolute configuration of the amino acids and at C-3 of the lipid side chain required resolution. Analysis of the absolute configuration of the chiral residues of 1 were both accomplished by enantioselective LC-QTOF-ESI-MS analysis using the same Chiralpak ZWIX (+) column. 3-Hydroxy decanoic acid was confirmed to be $R$-configured (Figure S19). The amino acid composition and configuration of the 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) derivatized hydrolysate was determined by comparison of the retention times and mass spectra of the amino acids in the sample with those of authentic standards (Figure S20) and found to be 1 L-Thr, 1 L-Pro, 1 D-Ala, 2 L-Ala, 2 D-Val and 1 D-Gln (detected as D-Glu).
In order to assign the relative positions of the D- and L-Ala residues, we turned our attention to the information available from the thf BGC. As pointed out above in the bioinformatics analyses, we deduced that Ala4 and Ala6 had to be L-configured and only for Ala2 the possibility for epimerization existed driven by the C/E-domain of module 3. Thus, it was concluded that Ala2 has to be D- and Ala4 and Ala6 to be L-configured. It is noteworthy to mention that, in contrast to many other examples of Pseudomonas lipopeptide BGCs, all C/E domains appeared to be fully functional. Finally, the configuration of the peptide bond preceding the proline unit was established by the difference in the chemical shifts of the β and γ carbons (Δδβ,γ). The difference Δδβ,γ = 3.8 ppm (Table 2) indicated a trans configuration of the proline residue (Pro7) in 1. The resultant complete 3D-structure of thanafactin A is shown in Figure 3.

![Figure 3. Chemical structure of thanafactin A (1).](image)

Thus, the experimentally validated structure of thanafactin A (1) was in full agreement with the predicted structure. It is noteworthy to mention that thanafactin cannot be assigned to any existing group of Pseudomonas octalipopeptides because its amino acid sequence is distinctly different (Table 1). Consequently, thanafactin A represents the founding member of a new class of linear octalipopeptides, characterized by a 3-hydroxy decanoic acid connected to the N-terminus of an eight amino acid linear peptide which comprises a proline and multiple alanine residues as well as a C-terminal threonine moiety. The latter is rather a rare feature, because taking into account all known Pseudomonas lipopeptides, a C-terminal Thr residue is so far only given in the syringomycin group but in a chlorinated version and in a cyclic context. Proline and multiple alanine containing Pseudomonas lipopeptides are so far only known from the considerably larger cyclic lipopeptides of the tolaasin and the syringopeptin group. Thus, is it tempting to speculate if the thf BGC can be delineated from the larger thanapeptin BGCs by inter gene cluster recombination, as recently suggested by Stallforth and coworkers.
A survey of public genome databases to identify BGCs related to thfABC revealed that similar gene clusters are existent. One example is the so far only partially in silico predicted octalipopeptide brasamide, which also shows an Ala-Val-Ala-Pro motif and might be, upon experimental validation, a further aspirant of the thanafactin compound family.

**Biological Activity.** Linear octalipopeptides such as syringafactins and chichofactins were so far foremost reported to be involved in virulence, swarming motility and biofilm formation. Because the surfactant activity of lipopeptides commonly assists swarming, we tested the swarming behavior of the wild type producer strains on soft agar in comparison with the corresponding KO-mutants. Notably, the wild type strains were able to swarm on soft agar plates, while the KO strains showed a strain-dependent behavior (Figure S21). The deletion mutants DSM11579-ΔNRPS C2 and E7 exhibited in comparison with the wild type reduced swarming, thus advocating for a physiological role of 1. However, intriguingly the interruption of thanafactin production in strain SH-C52 produced no difference concerning the swarming radius.

We were further interested in the potential for 1 to act on bacterial quorum sensing (QS) systems and proteases. Thus, the inhibitory activity towards a panel of 10 proteases (Table S6) was investigated, however 1 exhibited only a weak inhibitory activity against human cathepsin K. In the QS inhibition assay thanafactin A (1) was found to be inactive at the tested concentration (1 mg/mL).

**EXPERIMENTAL SECTION**

**General Experimental Procedures.** Optical rotation values were measured on a Jasco P-2000 polarimeter, using a 3.5 mm x 10 mm cylindrical quartz cell. UV spectra were taken on a Perkin-Elmer Lambda 25 UV/VIS spectrometer. ECD spectra were conducted using a Jasco J-720 spectropolarimeter. Infrared spectra were obtained employing a Jasco FT/IR 4200 spectrometer, interfaced with a MIRacle ATR device (ZnSe crystal). 1D and 2D NMR spectra were acquired on a 400 MHz Bruker AVANCE III NMR spectrometer operating at 400.17 and 100.63 MHz, respectively, and equipped with a 5 mm broadband SmartProbe and AVANCE III HD Nanobay console. All spectra were recorded in d6-DMSO and calibrated to the residual solvent signals (resonances at δH 2.50 and δC 39.5 ppm). High-resolution mass spectra were acquired on an HR-ESI-TOF-MS Bruker maXis 4G mass spectrometer. HPLC was performed with a Waters system, consisting of a Waters 600S controller and a 616 pump, a Waters 996 photodiode array detector, a
Rheodyne 7725i injector and a PerkinElmer vacuum degasser series 200, operated by the software Millenium. For LC-MS analysis, an 1100 Series Agilent HPLC system was fitted with a G1322A degasser, a G1312A binary pump, a G1329A autosampler, and a G1315A diode array detector. The Agilent HPLC components were connected with an ABSCIEX 3200 QTRAP LC/MS/MS mass spectrometer. All solvents were purchased as HPLC or LC-MS grade, respectively. Steam sterilization of medium components and waste were performed at 121 °C for 15 min at 2.1 bar using a Systec VX-150 autoclave, equipped with air exhaust filtration.

**Bioinformatics.** The detection of biosynthetic gene clusters encoding secondary metabolites was performed using the online software antiSMASH 4.0. Every open reading frame, detected by antiSMASH, was further investigated by protein-protein BLAST and Pfam analyses. This enabled the assignment of putative roles for the proteins within the clusters. Automated prediction of adenylation and condensation domain specificity was done using antiSMASH 4.0. The manual phylogenetic analyses of the A-, C- and TE domains were conducted as previously described, employing the software packages Mega v7.0 or Clustal Omega in combination with Geneious 10.0.6. The comparison between the non-ribosomal peptide synthetases and the Major Facilitator Superfamily transporter of both strains was performed using ClustalW. Genome-wide phylogenetic and digital DNA-DNA hybridization analyses were conducted with the web-based software Type Strain Genome Server (TYGS) of the DSMZ.

**Strains.** *Pseudomonas* sp. SH-C52 was isolated by one of the authors (J.M.R) from disease suppressive soil, while *Pseudomonas fluorescens* DSM 11579 was obtained from the closed repository of the DSMZ – German Collection of Microorganisms and Cell Cultures GmbH. *Escherichia coli* DH10B, employed for molecular cloning procedures, was purchased from Thermo Fisher Scientific.

**Recombinant DNA Methods and Generation of ΔNRPS Mutants.** Standard methods were used for the preparation of *E. coli* DH10B competent cells and heat shock transformation. *Pseudomonas fluorescens* SH-C52 and *Pseudomonas fluorescens* DSM 11579 transformations were performed using the ‘sucrose method’. The fragments for the disruption of NRPSs were amplified using the Herculase II fusion DNA polymerase (Agilent Technologies) from genomic DNA of *Pseudomonas fluorescens* SH-C52 and *Pseudomonas fluorescens* DSM 11579,
respectively, and with the pairs of primers SH-C52 NRPS (forward and reverse) and DSM NRPS (forward and reverse) (Table S5). The fragments were purified using the innPREP DOUBLEpure kit (analytikjena). Every fragment was cloned into the plasmid pCR Blunt (Invitrogen) using the Zero BluntTM PCR cloning kit (Invitrogen). The plasmids were transformed into *E. coli* DH10B by application of the heat shock method and subsequently transferred to *P. fluorescens* SH-C52 and *P. fluorescens* DSM 11579 by electroporation. Transformants were selected on Luria-Bertani agar (LA) with kanamycin (25 µg/mL) and grown for 10 days at 30 °C. Next, putative transformants were cultivated in 96-wells plate with 200 µL of Luria-Bertani Broth (LB) and kanamycin (25 µg/mL) at 30 °C overnight. Disruptions were confirmed by PCR using the corresponding primers (Table S5). Finally, two transformants were selected for every strain: *P. fluorescens* SH-C52 ΔNRPS A1 and B1 and *P. fluorescens* DSM 11579 ΔNRPS C2 and E7.

**Cultural Conditions for Isolation of Thanafactin A.** Seed cultures of *P. fluorescens* DSM 11579 were grown in 12.0 mL of Davis Minimal Broth without dextrose but containing 20 mM glycerol (DMBgly) in 50 mL Falcon tubes for 2 days employing a Gerhardt horizontal shaker at 110 rpm. The main cultivation was performed in six batches. Each batch (4.8 L) consisted of twelve 1.8 L Fernbach flasks containing 0.4 L of SRM production medium which were each inoculated with 0.4 mL of seed culture. Cultures were incubated for 120 h at 28 °C in an INFORS HT Multitron Pro orbital incubator shaker without shaking.

**Extraction and Isolation.** The broth from a 28.8 L fermentation of *P. fluorescens* DSM 11579 was centrifuged at 4000 rpm and 4 °C for 10 min. Cells were discarded, while 20g/L Diaion HP-20 resin was added to the supernatant. The mixture was transferred into a fritted funnel and drained of excess supernatant. The adsorbent resin was first washed with five bed volumes of purified H₂O and then eluted with eight bed volumes of 100% MeOH. The fraction containing 1 was finally purified by semipreparative HPLC on a Phenomenex Aeris Peptide XB-C18 column (3.6 µm, 250 x 2.6 mm). A linear gradient of 20:80 to 100:0 MeOH-H₂O (0.1% formic acid) over a period of 20 min, followed by isocratic elution at 100:0 for an additional 10 min (1 mL/min flow rate; UV monitoring at 210 nm) was applied. Under these conditions, 1 was eluted with a retention time of 34.8 min and obtained as a white powder (4.5 mg).
Thanafactin A (1): amorphous, white opaque powder; \([\alpha]^{25}_D\) -26.3 (c 1.0, MeOH); FT-IR (ATR) \(\nu_{max}\) 3663, 3454, 2969, 2948, 2921, 1738 cm\(^{-1}\); \(^1\)H NMR and \(^{13}\)C NMR data, Table 1; HRESIMS \(m/z\) 926.5553 [M+H]+ (calcld for C\(_{43}\)H\(_{76}\)N\(_9\)O\(_{13}\), 926.5557; \(\Delta -0.8\) ppm).

**Enantioselective HPLC-MS Analysis of the 3-Hydroxy-Decanoic Acid Portion of 1.** 1 mg of thanafactin A was hydrolysed using 250 \(\mu\)L of 6N deuterated hydrochloric acid (DCl/D\(_2\)O) at 110 °C for 18 h. Afterwards DCl/D\(_2\)O was removed using a Thermo Savant ISS110 SpeedVac (Thermo Scientific) at 43 °C to complete dryness. The residue was re-suspended in 100 \(\mu\)L H\(_2\)O and extracted twice with 200 \(\mu\)L CHCl\(_3\). The aqueous layer was used for chiral amino acid analysis, while the combined CHCl\(_3\) fraction was employed for the determination of the absolute configuration of 3-hydroxy decanoic acid. After the complete removal of CHCl\(_3\) on the Speed Vac, the residue was dissolved in 100 \(\mu\)L MeOH. Chiral 3-hydroxy decanoic acid separation were subsequently performed on a Chiralpak ZWIX (+) (150 × 4 mm i.d.) and a Chiralpak ZWIX (-) column (150 × 4 mm i.d.) column from Chiral Technologies Europe. Both columns were thermostatted at 5 °C. The isocratic run of 30 min was performed with a MeCN-MeOH-CH\(_3\)COOH (95:5:0.025) mixture as a mobile phase, a flow rate of 0.3 mL/min and an injection volume of 10 \(\mu\)L. Sample injections were performed with a Pal HTC-XS autosampler from CTC. Chromatographic separations were performed on an Agilent 1290 UHPLC system equipped with a binary gradient pump and a thermostatted column compartment from Agilent Technologies. Mass spectrometric detection was performed with a SCIEX TripleTOF 5600+ MS equipped with a Turbo V ion source from SCIEX, which was operated in the negative ionization mode at a source temperature of 350 °C and ion spray voltage floating (ISVF) of 4500 V. Analyses were performed in scan mode, using a scan range of \(m/z\) 30-2000, and an accumulation time of 250 ms, collision energy (CE) of -10 V, a declustering potential (DP) of -80 V. The gas settings were curtain gas (CUR) 35 psi, ion source gas 1 (GS1) 60 and GS2 60. As a second experiment, information dependent acquisition (IDA) with an accumulation time of 100 ms, CE of -20V, collision energy spread (CES) of 10 and DP pf -80 V was employed. Analysis of the authentic standards of racemic and 3\(R\)-hydroxy-decanoic acid ([M-H]$: \(m/z\) 187.1340) showed that the \(R\)-enantiomer elutes earlier on the Chiralpak ZWIX (+) column at a retention time of 14.15 min, while the \(S\)-enantiomer elutes at 15.57 min (Figure S19a). As expected, the elution order of the two enantiomers were reversed on the Chiralpak ZWIX (-) column with the \(S\)-enantiomer eluting at 13.89 min and the \(R\)-enantiomer at 14.88 min (Figure S19b). The enantiomers of \(R\)- and \(S\)- were verified by comparing their retention times (TOF survey scan) and their mass spectra (IDA) with that of the corresponding
RS- and R-3-hydroxy decanoic acid standard compounds. The (R)-3-hydroxydecanoic acid standard was obtained by acidic hydrolysis of a R95-rhamnolipid from Sigma-Aldrich according to the literature.\textsuperscript{34} Note that the small amount of S-enantiomer shown in Figure S19 originated from a slight racemization of the R-enantiomer during sample preparation.

**Enantioselective LC-MS Analysis of the Peptide Moiety of Thanafactin A (1).** After fatty acid extraction with CHCl\(_3\), the aqueous fraction was derivatized with 6-aminoquinolyl-N-hydroxy succinimidyl carbamate. Preparation of AQC and amino acid derivatization with AQC were performed as described in the literature.\textsuperscript{35} Enantioselective separation of amino acid-AQC derivatives were performed according to Horak et al.\textsuperscript{36} on a Chiralpak ZWIX (+) column using an Agilent 1290 UHPLC coupled with a SCIEX TripleTOF 5600+ ESI-MS instrument in positive ionisation modes. Note that due to acidic hydrolysis with deuterated solvents, glutamine was deaminated to glutamic acid with HDX exchange to \([\text{Glu-}[^2\text{H}]_2 + \text{H}]^+: m/z \ 320.1210]\) (Figure S20).

**Swarming Motility Experimental Section.** To evaluate the swarming motility, Kuiper’s method\textsuperscript{37} and the variation described by Pauwelyn et al.\textsuperscript{38} were used. The wild type strains and Thf-mutants were cultured overnight in 5 mL of LB media at 30˚C. Later, the pre-cultures were diluted to obtain OD\(_{600}\) 0.53 (±0.05). 5 µL of the sample was spotted on the center of the surface of the LB plates solidified with 0.5% agar. Each experiment was triplicated. The plates were incubated at 30 °C for 24 h.

**Protease Inhibition Assay.** Human cysteine proteases (cathepsin K, cathepsin B, cathepsin L), human serine proteases (leukocyte elastase, thrombin, matriptase-2), bovine serine proteases (chymotrypsin, trypsin, factor Xa) and porcine pancreatic elastase were assayed with chromogenic and fluorogenic peptide substrates, respectively, as described.\textsuperscript{39} The formation of \(p\)-nitroaniline (pNA) and 7-amino-4-methylcoumarin (AMC), respectively, was followed over 20 min (matriptase-2) or 60 min (all other proteases).

**Assay for Inhibition of Quorum Sensing.** Chromobacterium violaceum CV026 and Staphylococcus aureus PC322 were used to assess the potential influence of the compound on acyl homoserinelactone (\(C. \ violaceum\)) or autoinducing peptide (\(S. \ aureus\)) based quorum sensing.

Assay with \(C. \ violaceum\) CV026. 1.25 mL 1 M MOPS solution, 2 µL C6-HSL solution (Cayman Chemical; 100 µg/mL) and 2 mL of an over-night culture of \(C. \ violaceum\) in LB broth were added to 20 mL liquefied LB agar (Sigma-Aldrich) at 50 °C. The agar was poured into a petri dish. After solidification, 50 µL of a 1 mg/mL test solution of the compound were given into round wells punched into the agar (7 mm diameter). After incubation at room temperature for 1 h, the
plates were incubated at 30 °C overnight. 50 µL of a coumarin solution (3 g/L in MeOH) were used as positive control.\textsuperscript{40} QS-regulated violacein production was visually examined after incubation.

Assay with \textit{S. aureus} PC322. 200 µL 5-bromo-4-chloro-3-indoxyl-β-D-galactopyranoside (X-Gal) [20 mg/mL in DMSO], 2.5 µL erythromycin solution [50 mg/mL in EtOH], and 2 mL of an over-night culture of \textit{S. aureus} PC322 (OD\textsubscript{600} adjusted to 0.1) were added to 20 mL liquefied TSA agar, as previously described.\textsuperscript{41} The agar was poured into a petri dish. After solidification, 50 µL of a 1 mg/mL test solution of the compound were given into round wells punched into the agar (7 mm diameter). After incubation at room temperature for 1 h, the plates were incubated for at least 12 h at 37 °C. 50 µL of a hamamelitannin solution (4 mg/mL in DMSO) were used as positive control.\textsuperscript{42} QS-regulated X-Gal metabolism was visually examined after incubation.

\textbf{Nucleotide Sequence Accession Numbers.} The nucleotide sequence of the whole genome sequence of the strains \textit{Pseudomonas} sp. SH-C52\textsuperscript{2} and \textit{P. fluorescens} DSM 11579\textsuperscript{9} can be found in the GenBank database under accession numbers CBLV000000000 and JAAOIQ000000000, respectively. Furthermore, the complete and corrected gene cluster coding for thanafactin A has been also deposited separately in the GenBank database under accession number MT43159.

\textbf{ASSOCIATED CONTENT}

\textbf{Supporting Information}

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jnat-prod.XXXXXXX.

Bioinformatics analyses, spectroscopic data (HRESIMS, IR, UV, 1D and 2D NMR spectra), enantioselective analyses, and biological assay results of compound 1 (PDF)

\textbf{AUTHOR INFORMATION}

*Tel: +4970712976970. Fax: +497071295250. E-Mail: harald.gross@uni-tuebingen.de
ACKNOWLEDGMENTS
We like to thank Dr. D. Wistuba and her team (Mass Spectrometry Department, Institute for Organic Chemistry, University of Tübingen, Germany) for HR-MS measurements. Funding is gratefully acknowledged from the Deutsche Forschungsgemeinschaft (DFG) who supported H.G. (GR2672/2–1) within the frame of the “Research Unit FOR854-Post-Genomic Strategies for New Antibiotic Drugs and Targets”. The contribution of H.G. and T.H.J.N was also financially supported by the German Centre for Infection Research (DZIF). We gratefully acknowledge M. Bird, Director of Operations and Portfolio, Molecular Discovery Research, GlaxoSmithKline Research & Development Limited, Stevenage, United Kingdom, for granting the permission to obtain and work with the strain DSM 11579 from the nonpublic collection of the DSMZ.
REFERENCES


