A Non-Canonical NRPS Is Involved in the Synthesis of Fungisporin and Related Hydrophobic Cyclic Tetrapeptides in Penicillium chrysogenum

Ali, Hazrat; Ries, Marco I.; Lankhorst, Peter P.; van der Hoeven, Rob A. M.; Schouten, Olaf L.; Noga, Marek; Hankemeier, Thomas; van Peij, Noel N. M. E.; Bovenberg, Roel A. L.; Vreeken, Rob J.

Published in:
PLoS ONE

DOI:
10.1371/journal.pone.0098212

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2014

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Ali, H., Ries, M. I., Lankhorst, P. P., van der Hoeven, R. A. M., Schouten, O. L., Noga, M., Hankemeier, T., van Peij, N. N. M. E., Bovenberg, R. A. L., Vreeken, R. J., & Driessen, A. J. M. (2014). A Non-Canonical NRPS Is Involved in the Synthesis of Fungisporin and Related Hydrophobic Cyclic Tetrapeptides in Penicillium chrysogenum. PLoS ONE, 9(6), [e98212]. https://doi.org/10.1371/journal.pone.0098212

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 15-05-2021
A Non-Canonical NRPS Is Involved in the Synthesis of Fungisporin and Related Hydrophobic Cyclic Tetrapeptides in *Penicillium chrysogenum*

Hazrat Ali1,2*, Marco I. Ries3*, Peter P. Lankhorst4, Rob A. M. van der Hoeven4, Olaf L. Schouten4, Marek Noga3,5, Thomas Hankemeier3,5, Noël N. M. E. van Peij4, Roel A. L. Bovenberg4,6, Rob J. Vreeken3,5, Arnold J. M. Driessen1,2*

1 Molecular Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, Zernike Institute for Advanced Materials, University of Groningen, Groningen, The Netherlands, 2 Kluyver Centre for Genomics of Industrial Fermentations, Delft, The Netherlands, 3 Division of Analytical Biosciences, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands, 4 DSM Biotechnology Center, Delft, The Netherlands, 5 Netherlands Metabolomics Centre, Leiden University, Leiden, The Netherlands, 6 Synthetic Biology and Cell Engineering, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Groningen, The Netherlands

**Abstract**

The filamentous fungus *Penicillium chrysogenum* harbors an astonishing variety of nonribosomal peptide synthetase genes, which encode proteins known to produce complex bioactive metabolites from simple building blocks. Here we report a novel non-canonical tetra-modular nonribosomal peptide synthetase (NRPS) with microheterogeneity of all involved adenylation domains towards their respective substrates. By deleting the putative gene in combination with comparative metabolite profiling various unique cyclic and derived linear tetrapeptides were identified which were associated with this NRPS, including fungisporin. In combination with substrate predictions for each module, we propose a mechanism for a ‘trans-acting’ adenylation domain.

**Citation:** Ali H, Ries MI, Lankhorst PP, van der Hoeven RAM, Schouten OL, et al. (2014) A Non-Canonical NRPS Is Involved in the Synthesis of Fungisporin and Related Hydrophobic Cyclic Tetrapeptides in *Penicillium chrysogenum*. PLoS ONE 9(6): e98212. doi:10.1371/journal.pone.0098212

**Editor:** Mikael Rardam Andersen, Technical University of Denmark, Denmark

**Received** November 5, 2013; **Accepted** April 29, 2014; **Published** June 2, 2014

**Copyright:** © 2014 Ali et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** HA and MR were supported by STW (Stichting Technische Wetenschappen) in the framework of the Genbiotics Perspectief Programme. HA was also supported by Higher education commission Pakistan (HEC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** AD is a PLOS ONE Editorial Board member. PL, RvdH, OS, NvP and RB are employed by the DSM. There are no patents, products in development or marketed products to declare. This does not alter the authors’ adherence to PLOS ONE policies on sharing data and materials.

* E-mail: a.j.m.driessen@rug.nl

† These authors contributed equally to this work.

‡ Current address: Center for Biotechnology and Microbiology, University of Swat, Swat, Pakistan

**Introduction**

Fungal non-ribosomal peptides contribute a large variety of secondary metabolites with remarkable properties such as antibacterial, antifungal, antiparasitic, antiviral and immunosuppressive activities. These metabolites are produced by large, multifunctional protein complexes, called nonribosomal peptide synthetases (NRPS). These enzymes catalyze the stepwise condensation of simple amino acid building blocks to complex molecules. NRPSs have a modular architecture, with each module responsible for one discrete chain-elongation step. Every single module can be subdivided into domains that carry all essential information for recognition, activation and modification of the corresponding substrate. At a minimum, a typical NRPS module consists of an adenylation (A) domain, responsible for amino acid activation, a thiolation domain, also known as peptidyl carrier protein (PCP), which binds the activated amino acid and a condensation (C) domain that catalyzes peptide-bond formation. The common arrangements of these domains follow a (C-A-PCP)n organization. Additionally, a variety of optional domains have been described such as methyltransferase (MT) and epimerization (E) domains [1].

The number of modules and their domain organization within NRPS enzymes controls the structures of the final product(s) [1–3]. Thus, the order of modules usually corresponds to the sequence of amino acids in the peptide. Many NRPS systems adhere to this mechanistic paradigm, which is often referred to as the “co-linearity rule” [4]. Also exceptions to this rule have been discovered, including iterative NRPSs, which incorporate multiple residues of the same amino acid iteratively into the peptide structure and the so called nonlinear NRPSs, which deviate completely from the standard domain organization leading to unexpected products [3,5,6].

The impact of non-ribosomal peptide metabolites on the quality of human life raised the interest of pharmaceutical industries to invest in identification, engineering and heterologous expression of NRPS genes and pathways to ensure the rational production of novel compounds [7–9]. To understand the basic mechanisms of the biosynthesis of these complex NRPSs, detailed studies have been performed during the past few decades. These included the structural analysis of adenylation domains, mutational analysis of substrate specificity of these modules, the fusion of unrelated modules to produce new products and the identification of helper proteins for optimal activation of adenylation domains [10–12].
Although this has led to detailed insights into catalytic mechanisms, so far a structure of a complete NRPS is lacking that would reveal how modules cooperate to facilitate product formation. The availability of genome sequencing data and sophisticated bioinformatics analysis of various fungi revealed the presence of many NRPS genes that have not been associated with known secondary metabolites [13–15]. Moreover, most of these genes are not expressed when the fungi are grown under laboratory conditions, implying that many more secondary metabolites await discovery.

The filamentous fungus *Penicillium chrysogenum* is well known for the production of the antibiotic penicillin G that is synthesized by the tri-modular NRPS d-L-2-amino-3-oxopropionyl-L-cysteinyl-D-valine synthetase. In addition, other NRPS derived secondary metabolites like the roquefortines and meleagrin have been reported from filamentous fungi, including the previously identified metabolite fungisporin [19,20], and the discovery of a tetra-modular NRPS with an unusual domain organization with adenylation domains showing microheterogeneity. It is proposed to term this NRPS HcpA [19,20].

**Materials and Methods**

**A. Chemicals**

6-aminoquinolyl-N-hydroxy succinimidyl carbamate (AQC) reagent and borate buffer were obtained as part of AccQ-Tag Reagent Kit from Waters (Waters, Milford, MA, USA). cyclic-(d-Tyr-r-Phe-α-Val-l-Val) was obtained from Celtek Peptides (Nashville, TN).

**B. Host strains, media, grown condition and plasmid construction**

Deletion of the *hcpA* gene was carried out in *P. chrysogenum* strain DS54555, which lacks penicillin cluster genes and the *ka70* gene [18]. This strain was kindly provided by the DSM Biotechnology Center (Delft, Netherlands). A deletion plasmid was constructed by amplifying the flanking regions of the targeted gene with the Multisite Gateway® Three-Fragment Vector Construction Kit according to the procedure described by Invitrogen using pDEST R4-R3p as template. Primers used for the construction of the deletion plasmid pDEST R4-R3p PcHcpA (Figure S1) are listed in Table S1. *Escherichia coli* DH15α [F- Φ80lacZΔM15 Δ(lacZΔM15T4-argF)] U169 recA1 endA1 hsdR17 (rK-, mK+) phi80 supE44 λ- thi-1 gyrA96 mlA) was used as host strain for high frequency transformation and plasmid DNA amplification [21]. All the strains were grown on yeast nitrogen base-glucose-yeast extract (YGG)-medium for protoplasts formation and transformation [22]. Both mutant and host strains of *P. chrysogenum* were grown on secondary metabolite production medium as described previously [18].

**C. Transformation procedure**

The deletion plasmid pDes R4-R3p PcHcpA was transformed to the protoplasts of *P. chrysogenum* DS54555 [23] yielding the ΔhcpA derivative of strain DS54555. The phleomycin resistance gene was used as selection marker for the deletion of the *HcpA* gene [22,24].

**D. Genomic DNA extraction, total RNA extraction, cDNA amplification and qPCR analysis**

Genomic DNA (gDNA) was isolated after 96 hours of growth on SMP medium (secondary metabolite production medium) using the modified yeast gDNA isolation protocol [25] in which the fungal mycelium is broken in a FastPrep FP120 system (Qbiogene). gDNA of the host and various deletion strains was isolated using the E.Z.N.A. Fungal DNA kit (Omega Bio-tek). Total RNA of the host strain was isolated after 48 hours of growth in SMP medium for the first time and then with the interval of 24 hours up to 216 hours of growth using the Trizol reagent (Invitrogen), with additional DNase treatment using the Turbo DNA-free kit (Ambion). Total RNA was measured using a NanoDrop ND-1000 and a concentration of 500 ng per cDNA reaction was used. cDNA was synthesized using the iScript cDNA synthesis kit (Bio-Rad) in a 10 μl end volume. The primers used to analyze the expression of the *hcpA* gene were designed around an intron to avoid amplification on gDNA (Table S2). For expression analyses, the γ-actin gene was used as a control for normalization (Table S2). A negative reverse transcriptase (RT) control was used to determine the gDNA contamination in isolated total RNA. The expression levels were determined as described [18].

**E. Southern blotting**

Southern blotting was carried out by digesting gDNA (5 μg) with the indicated restriction enzymes. Digested DNA fragments were separated on a 0.8 % agarose gel, blotted onto a Zeta-Probe membrane (Biorad) as described earlier [26], and hybridized with the indicated DIG labeled probes.

![Figure 1. Identified secondary metabolites.](https://doi.org/10.1371/journal.pone.0098212.g001) Structures of cyclic tetrapeptides identified in *P. chrysogenum*.
F. Metabolite profiling and analysis

1. Sample preparation. Host and deletion strains of P. chrysogenum strains used for gene assignments were grown in quinupticate according to the procedure described above. Samples for acquisition of the metabolite profiles from the growth curves were from five replicates. Metabolite profiling was carried out with modifications as described earlier [10]. Briefly, 4 μL of internal standard mixture (0.55 nmol/mL raniidine, 0.65 nmol/mL reserpine and 1.14 nmol/mL ampicillin) was added to 100 μL fermentation broth followed by the addition of 400 μL methanol for protein precipitation. The samples were vortexed and spun down at 14,000 g for 10 minutes. 300 μL supernatant was evaporated for 30 minutes in a speedvac (Thermo Scientific, San Jose, CA) and re-dissolved in 100 μL water. LC-UV-MS analysis was performed on an Agilent 1200 Capillary pump (Agilent, Santa Clara, CA) coupled in-line to a Surveyor PDA detector (Thermo Scientific, San Jose, CA) and LTQ-FT mass spectrometer (Thermo Scientific, San Jose, CA) using electrospray ionization and operated in a scan range between m/z 110 and m/z 2000 in positive/negative ion switching mode. Separation was performed on a Waters Atlantis T3 column (2.1×100 mm, 3 μm) (Waters, Milford, MA) starting with 98 % of solvent A (1 % acetonitrile and 0.1 % formic acid in water) and 2 % solvent B (1 % water and 0.1 % formic acid in acetonitrile) for 1.5 minutes at a flow rate of 300 μL/min. 40 % B were reached after 22 minutes and 100 % B at 25 minutes. The column was flushed with 100 % B and re-equilibrated to initial conditions. Peak detection and integration were performed using an in-house tool followed by statistical tests to discover significant different features. Finally, discovered features were integrated using LCQuant v.2.6 (Thermo Scientific, San Jose, CA). The non-related non-endogenous compound reserpine was used as internal standard.

2. Identification of cyclic tetrapeptides. The identity of cyclic tetrapeptides was determined using samples from liquid cultures of P. chrysogenum and a crude spontaneous precipitate obtained from P. chrysogenum cultures, containing primarily 1 and 2 next to various minor abundant cyclic tetrapeptides in relative concentrations of 70 and 15 %, respectively, LC-MS™ experiments for the determination of consecutive amino acid losses were performed according the metabolite profiling section with normalized collision energies of 35 %, an isolation width of 1 amu and an activation Q of 0.30. NMR spectra were recorded on a Bruker Avance III 700 MHz NMR spectrometer (Bruker, Billerica, MA), equipped with a 5 mm TCI probe. 2 mg of each sample was dissolved in 0.6 mL anhydrous DMSO. NMR spectra were acquired at 340 K.

3. Identification of linear tetrapeptides. Linear tetrapeptides were identified according their multiple-stage fragmentation after AQC derivatization [27]. Methanol (400 μL) was added to an aliquot of 100 μL fermentation broth for protein precipitation. Samples were vortexed for 10 minutes, spun down for 10 minutes and 300 μL of the supernatant was evaporated to dryness in a speedvac (Thermo Scientific, San Jose, CA). Derivatization was done according to the supplier’s procedure by re-dissolving the sample in 40 μL water, 40 μL borate buffer (pH 8.5) and 20 μL AQC solution. The mixture was vortexed for 10 minutes and heated for 10 minutes at 55°C.

LC-MS™ experiments were conducted on an Agilent 1200 Capillary pump (Agilent, Santa Clara, CA) coupled to a LTQ-FT mass spectrometer (Thermo Scientific, San Jose, CA) using electrospray ionization. Separation was performed on a Waters Atlantis T3 column (2.1×100 mm, 3 μm) (Waters, Milford, MA) starting with 72 % of solvent A (1 % acetonitrile and 0.1 % formic acid in water) and 28 % of solvent B (1 % water and 0.1 % formic acid in acetonitrile) for 1.5 minutes at a flow rate of 300 μL/min. After 8 minutes the gradient reached 60 % of solvent B. Subsequently, the column was flushed with 100 % B before it was re-equilibrated to initial conditions. The peptide sequences were elucidated using multiple-stage collision-induced dissociation (CID) of the protonated molecule following consecutive cleavages of amino acids residues, starting from the C-terminus of the derivatized linear tetrapeptide. CID was performed with normalized collision energies of 35 %, an isolation width of 1 amu and an activation Q of 0.30.

Results

Bioinformatic analysis of a tetrapeptide NRPS

Genome sequencing revealed that P. chrysogenum encodes 11 NRPS genes [15]. Microarray expression analysis under glucose-limited chemostat culture conditions as well as quantitative PCR under shake flask culture condition showed that Pc160g04690 (genbank protein identifier CAP93139.1) (hcpA) is highly expressed (Figure 2A) [15]. The hcpA gene encodes a large multimodular non-ribosomal peptide synthetase enzyme (Figure 3) with 6064 amino acids and a calculated molecular mass of about 670 kDa. HcpA, which shows 54% sequence identity to the orthologous An08g02310 in A. niger, has the domain architecture A1-PCP1-E-C2-A2-PCP2-C3-A3-PCP3-E-C4-PCP4-C-PCP (A = adenylation, C = condensation, PCP = thiolation, and E = epimerization) (Figure 3A) [15]. A similar domain architecture was deduced for the orthologous protein of A. niger except for an insertion of a 177 amino acid long sequence between the adenylation domain A1 and A2 which shows homology to conserved motifs of an incomplete domain (C) in the C-terminal part of HcpA. The HcpA sequence analysis shows a similarity with the signature sequence of the adenylation domain of the bacitracine-producing NRPS that activates phenylalanine as a substrate. The signature sequences for the A1 domain (DAVCAGVAK) and DAVSAGVAAK and DAIIIAVAVK as the signature sequences for the A2 domain in the P. chrysogenum and A. nidulans HcpA proteins. This motif exhibits similarity with the signature sequence of the adenylation domain of a bacitracine-producing NRPS that activates phenylalanine as a substrate. The overall architecture of both synthetases is unusual, as the A1 and A2 domains occur adjacent to each other, flanked by a single C and PCP domain in a C1-A1-A2-PCP1 pattern. On the other hand, an incomplete module (C1-PCP1) without an adjacent A domain is found at the N-termini of these NRPSs (Figure 3A).

Genetic deletion of the tetrapeptide NRPS and secondary metabolite identification

In order to identify the secondary metabolites synthesized by HcpA, the corresponding gene was deleted and comparative metabolite profiling was performed on the culture supernatant of the host and deletion strain. As host strain, P. chrysogenum DS4555 was used, which is derived from the industrial DS17690 strain and that lacks the kat70 gene to make it competent for homologous recombination. The DS4555 strain also lacks the multiple penicillin biosynthetic genes clusters in order to facilitate the detection of unknown secondary metabolites in the culture broth
as the profile is no longer dominated by β-lactams. The *hcpA* gene was resequenced from the genome of *P. chrysogenum* DS5455 and the nucleotide sequence of the open reading frame (genbank KJ679502) including the promotor region was found to be identical to the *hcpA* gene present in the sequenced genome of *P. chrysogenum* Wisconsin54-1255 [15]. The gene deleted by homologous recombination using the deletion plasmid pDEST R3-R4 (Figure S1) containing the flanking regions of *hcpA* and the phleomycin resistance gene. Colonies were selected on phleomycin containing agar plates where the mutant colonies showed smooth phenotypic characteristics compared to the wrinkled surface of the colonies of the parental strain (Figure 4). The deletion of the *hcpA* gene was confirmed by southern blot hybridization (Figure 5A).

The host and *hcpA* strain were grown for 168 hours in SMP Medium followed by comparative metabolite analysis using HPLC-UV-MS. Several secondary metabolites were found to be present in the host but absent in the deletion strain (Figure 6, Table S3). These compounds could be classified into two groups according their chemical structure. The first group consists of ten cyclic tetrapeptides, which were identified using HPLC-MS<sup>n</sup>, NMR and a synthetic standard (Figure 1). Upon excitation, cyclic tetrapeptides undergo ring opening in the mass spectrometer resulting in four linear tetrapeptides, which can be sequenced in a similar fashion as linear peptides. By following the sequential loss of amino acids from b-ions of each generated linear tetrapeptide, the sequence of their cyclic origin could be determined (Figure S2). The identities of the involved amino acids, corresponding to the losses in the mass spectrometer, as well as their sequence were additionally confirmed by <sup>1</sup>H-NMR and <sup>13</sup>C-NMR experiments for the compounds 1 and 2 (Table S4 and S5). To discriminate the amino acid isoleucine from its isomer leucine, which is represented by a loss of 113 Da (C<sub>6</sub>H<sub>11</sub>ON) in the mass spectra of compounds 5–10, <sup>1</sup>H-, <sup>13</sup>C- and various 2D-NMR experiments were conducted (Figure S3, Table S6). Overall, cyclic tetrapeptides obtained from metabolic profiling contain various combinations of the five amino acids valine, isoleucine, phenylalanine, tyrosine and tryptophan making them extremely hydrophobic. They are arranged in a common sequence in which two aromatic amino acids are followed by two aliphatic amino acids. The absolute stereochemistry of compound 2 was confirmed by spiking its synthetic standard to a natural extract, which did not lead to additional signals in the <sup>1</sup>H-NMR spectra, whereas the intensity of the main signals increased as compared to the impurities (Figure S4A and B). In addition, superimposing the HMBC spectra of the natural and the synthetic sample of peptide 2 illustrated identical correlations and shifts (Figure S5). Furthermore, retention time and MS<sup>2</sup> fragmentation did not show differences between the extracted compound and the synthesized standard. This leads to the conclusion that not only the sequence of amino acids is identical in the synthetic and natural peptide, but also the chirality.

---

**Figure 2.** Correlation between the expression level of *HcpA* and metabolite formation during growth. A: Time dependent expression level of *hcpA* as determined by quantitative RT-PCR during growth of *P. chrysogenum* as monitored by biomass formation. B: Internal standard corrected concentration of the cyclic peptide 1 and its degradation products 11–13 present in the growth media. The concentration of peptides was determined by HPLC-UV-MS.

doi:10.1371/journal.pone.0098212.g002
Figure 3. Hypothetical model for the biosynthesis of compound 1 in *P. chrysogenum* and *A. niger* by the HcpA NRPS. A: Binding of the monomers on the carrier protein domains. Both isolated condensation and thiolation domains $C_4$ and $T_4$, missing a preceding adenylation domain, are assumed to correspond to the adenylation domain $A_4$, located upstream. B: Proposed assembly of compound 1 including condensation domains presumed for catalyzing the formation of the peptide bond.

doi:10.1371/journal.pone.0098212.g003
of the individual amino acids. Therefore, the cyclic tetrapeptide 2 has the same stereochemistry as observed for 1 [19] with the first amino acid of an aliphatic and aromatic pair in D- and the second amino acid in L-form. Due to the low concentration of cyclic tetrapeptides 3–10, reliable stereochemical information could not be obtained. However, as these peptides originate from the same NRPS as 1–2 and share the same sequence of aliphatic and aromatic amino acids, identical stereochemistry is expected.

The second category of identified compounds consists of 18 linear tetrapeptides which are comprised of the same five amino acids as their cyclic analogues (Table S3 and S7). Similar to the cyclic peptides, each linear tetrapeptide contains two aliphatic and two aromatic amino acids in different arrangements yielding various isomeric structures. Due to similar chromatographic properties and an identical mass-over-charge ratio, these isomers are represented as a group in data obtained from metabolic profiling (Figure S6A and B). Their structure elucidation is challenging as minor fragments can be attributed to a fragmentation of low abundant linear tetrapeptides as well as to possible sequence scrambling of major linear tetrapeptides which was reported for similar linear peptides [29]. To separate and sequence these isomeric tetrapeptides and to prevent possible sequence scrambling, their N-terminus was derivatized using 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) (Figure S6C and D) [30]. De novo peptide sequencing was performed by following the consecutive amino acid losses of various b-ions using multiple-stage fragmentation mass spectrometry (Figure S6E and Table S7). Similar to the cyclic peptides, each linear tetrapeptide incorporates two consecutive aromatic and/or two consecutive aliphatic amino acids.

In conclusion, several linear and cyclic tetrapeptides with similar structural features were found to be present in the host but absent in the deletion strain which shows that they originate from one single NRPS, namely HcpA.

Expression of the hcpA gene and secondary metabolite production
To examine the expression of the hcpA gene, the host strain was grown for up to 216 hours in SMP medium. Samples were collected for total mRNA extraction and extracellular metabolites analysis during growth. Metabolite concentrations were determined by HPLC-UV-MS, while transcript levels were determined by quantitative PCR using γ-actin as a reference gene. A high expression profile of the hcpA gene was observed between 120 and 168 hours of growth, which was paralleled by a 12 times increase in the concentration of cyclic tetrapeptide 1 (Figure 2A and B) in the medium. In general, the concentration of cyclic tetrapeptides was exceptionally high around 168 hours of growth, while the concentration of the linear tetrapeptides increased with time. These data suggest that the linear tetrapeptides are derived from the cyclic tetrapeptides by degradation.

Discussion
Here we have demonstrated that the NRPS HcpA is responsible for the biosynthesis of cyclic hydrophobic tetrapeptides. Bioinformatics analysis of HcpA shows an unusual domain architecture in comparison to classical linear NRPSs with no products associated so far. However, through the deletion of the hcpA gene and comparative metabolite profiling, ten cyclic tetrapeptides were identified, including the previously described secondary metabolite fungisporin 1 [19]. According to the non-ribosomal code in combination with discovered cyclic products, the first NRPS module is specific for phenylalanine. Due to the adjacent E domain, responsible for epimerization of the activated amino acid,
a D-configuration is expected as observed for product 1 (Figure 1). The next two modules of the NRPS contain an unusual architecture in which PCP2 and C3 are flanked by two neighboring adenylation domains A2 and A4. As adenylation domain A4 shows high specificity towards valine and adenylation domain A2 shows high homology towards A1, both adenylation domains activate phenylalanine as supported by the structure of 1. Due to the lack of an adjacent PCP domain for A4 and a missing C domain for A2, which are necessary for loading and condensation of the substrate, no module activity is expected according to the classical C-A-PCP geometry (Figure 3A). Surprisingly, both products 1 and 2 show the incorporation of L-phenylalanine at the second position. As module 2 is predicted to be the only module capable to catalyze the incorporation of L-phenylalanine, A2 indeed must be active. Consequently, it seems likely that A4 is skipped, leaving the N-terminal C2 domain of module 2 to catalyze the condensation of the first amino acid from module 1 to the second amino acid from module 2, as observed for the products 1 and 2. The third module of HcpA contains the domains C3, A3, PCP3, and E arranged in a common linear order. The A3 domain is predicted to activate valine, which agrees with the peptide sequence of 1 and 2, as D-valine is their third amino acid. The fourth module of the NRPS is an incomplete module consisting of C4 and PCP4. Due to a missing preceding A domain no activity is expected. However, the chemical structures of both cyclic tetrapeptides 1 and 2 show the incorporation of L-valine as fourth amino acid in their peptide sequence. As A3 is the only domain predicted to be specific to valine without an adjacent epimerization domain, it is very likely that this domain is a ‘trans-acting’ A domain that interacts with C4 and PCP4 to add the last amino acid to the tetrapeptide. A similar architectural flexibility has been observed in the biosynthesis of versiniabactin, in which one A domain, located in HMWP2, loads three PCPs located on different modules [31–33]. As the linear domain organization of HcpA does not reflect a linear assembly of substrate incorporation into the final product, non-linear interactions are deduced. Although A1, C1, and PCP1 are not in a consecutive sequence on a genomic level, they might still be closely arranged in the final three-dimensional enzymatic structure. Structural characterization would be necessary to determine spatial proximity. Finally, after the incorporation of L-valine into the peptide chain, the C and PCP domain of the last module catalyze the cyclization of the peptide leading to the final cyclic structure, as previously observed in other NRPS systems [34,35]. It should be stressed that non-linear NRPS organizations are a very heterogenous group of NRPS systems which deviate from the colinearity rule thus showing various unusual mechanisms [6]. In an alternative model A4 might be non-functional, leaving A3 loading two PCPs, namely PCP3 and PCP4, similar to the cysteine-specific A domain of HMWP2 in versiniabactin biosynthesis [3]. However, more detailed biochemical studies are required to fully understand the interplay between these enzymes and to confirm the exact biosynthetic mechanisms involved.

Next to the production of 1 and 2, eight additional lower abundant cyclic tetrapeptides were identified to be present in the host strain and absent in the deletion strain (Figure 1). They show a similar peptide sequence as 1, containing two aromatic amino acids followed by two aliphatic amino acids. Although stereochemical information is only available for the compounds 1 and 2, it can be assumed that each of the cyclic tetrapeptides contains an aliphatic and aromatic amino acid in the D configuration, more specifically at the first and third position. These assumptions in combination with the stereochemical structure of 1 and 2 lead to the conclusion, that each adenylation domain of HcpA shows specificity towards more than one precursor amino acid with A1 being specific towards phenylalanine and tyrosine and A2 being specific towards phenylalanine and to a lesser extent to tryptophan (Table S8) reminiscent of microheterogeneity. Together with the two aliphatic amino acid selecting adenylation domains A3 and A4, which preferably activate valine before isoleucine, 16 cyclic tetrapeptide combinations are theoretically possible. However, only ten of these were detected in the fermentation broth of P. chrysogenum confirming a different degree of specificity towards their precursors. Based on a similar chemical scaffold of identified
Hydrophobic Cyclic Tetrapeptide Biosynthesis

A. niger

Multiple-stage fragmentation for de novo sequencing of the linear isomers 11 (Phe-Phe-Val-Val). As this change is attributed to the deletion of the hcpA gene, the hydrophobic cyclic peptides 1–10 need to be involved. Possibly, these molecules function analogous to hydrophobins in altering the surface properties and influencing aerial growth. The exact function is, however, still unclear.

In addition to the cyclic tetrapeptides several highly abundant linear tetrapeptides could be observed in the cultural broth of the host strain that were absent in the deletion strain. To each of the cyclic tetrapeptides, several linear tetrapeptides with the same sequence were present. For instance, for the cyclic tetrapeptide 1 with the sequence cyclo-(Phe-Phe-Val-Val), three linear tetrapeptides with the sequences Phe-Val-Val-Phe, Val-Phe-Val-Phe and Phe-Val-Phe-Val could be found at different ratios. Their concentration increased over time in the media while their cyclic counterpart decreased after 168 hours (Figure 2B). This leads to the conclusion that the linear peptides originate from the degradation of their cyclic counterparts by hydrolysis of their peptide bonds, which was observed exclusively between two aromatic, two aliphatic or an aliphatic followed by an aromatic amino acid (Table S3 and S7). Linear tetrapeptides with a N-terminal aliphatic amino acid and a C-terminal aromatic amino acid were not detected, leading to the conclusion that cleavage of this bond is not favorable. As cyclic tetrapeptides are relatively stable towards chemical and thermal degradation, enzymatic hydrolysis might be most probable.

P. chrysogenum contains a second NRPS that could potentially be involved in tetrapeptide formation, i.e., Pc13g14330. This protein has a linear organization and thus differs from the HcpA protein. Pc13g14330 is hardly expressed under batch culture conditions as is expected for the hydrophobic cyclic peptides presented here [36]. In addition, the colonies of the Δhcf1 strain lost the ability to produce a wrinkled surface leading to a rather smooth appearance (Figure 4). As this change is attributed to the deletion of the hcf1 gene, the hydrophobic cyclic peptides 1–10 need to be involved. It is hypothesized that these molecules function analogous to hydrophobins in altering the surface properties and influencing aerial growth.

Conclusions

Cyclic tetrapeptides are a structurally interesting group of peptides produced by fungi that have attracted much interest because of their cardiac ion-channel blocking properties. The complete biosynthetic mechanism of these metabolites has been deduced by gene deletion experiments in P. chrysogenum in combination with comparative metabolite profiling and consecutive structure elucidation. By this analysis, trans-aminoaacetylation for chain elongation in NRPS has been found. Furthermore, a distinct microheterogeneity of each adenylation domain towards different amino acid building blocks resulted in a range of cyclic tetrapeptides as produced by a single NRPS synthetase.

Supporting Information

Figure S1 Plasmid used for the deletion of the hcpA gene in P. chrysogenum.

Figure S2 Multiple-stage fragmentation for de novo sequencing of compound 2 based on sequential amino acid losses. A: MS² fragmentation spectra of the cyclic tetrapeptide 2 cyclo-(d-Tyr-l-Phe-d-val-l-Val). Due to ring opening of the cyclic peptide in the mass spectrometer at different positions, three different amino acid losses occurred, yielding different b⁺ ions. B-D: MS³ fragmentation spectra obtained by further fragmenting b⁺ ions from MS² showing b₋ ions used for peptide sequencing. The cyclic tetrapeptides 1 and 3–10 were identified accordingly.

Figure S3 HSQC spectrum of isolated mixture of cyclic tetrapeptides in DMSO. The spectra were used for the identification of isoleucine present in the minor abundant products 5 and 9. Signals corresponding to isoleucine are indicated with circles. Conducted TOCSY, COSY and HMBD experiments further confirm this conclusion (data not shown). ¹H and ¹³C chemical shifts are shown in Table S6.

Figure S4 ¹H-NMR spectra of mixtures of cyclic tetrapeptides. A: ¹H-NMR spectrum of a precipitate of various cyclic tetrapeptides containing primarily peptide 1 and 2 (bottom). Synthetic compound 2 spiked to the precipitated mix of various cyclic tetrapeptides in DMSO at 340 K (top). Signals corresponding to 2 were increased as compared to the impurities whereas additional signals did not appear. B: Zoomed regions of ¹H-NMR spectrum of natural precipitate (bottom) and precipitate spiked with compound 2 (top). Signals which increased after spiking are indicated (*).

Figure S5 Superimposed HMBD spectra of synthetic compound 2 (red) and an isolated mixture of cyclic tetrapeptides containing naturally produced compound 2 (black). Correlations between NH and CO are shown which indicate identical shifts for both samples. Assignments can be found in Table S4.

Figure S6 Sequencing of the linear isomers 11 (Phe-Phe-Val-Val), 12 (Val-Phe-Phe-Val) and 13 (Phe-Val-Phe). A: Extracted ion chromatogram (EIC) of the linear tetrapeptides 11, 12 and 13 using the profiling method. No chromatographic separation could be achieved. B: MS² fragmentation spectrum of unseparated linear tetrapeptides 11–13. Although the spectrum is dominated by fragments originating from 11, several lower abundant fragments can be found originating from a fragmentation of 12 and 13 or possible sequence scrambling of 11. C: Sequence of linear tetrapeptides after N-terminal AQC derivatization to achieve better chromatographic separation and to
prevent sequence scrambling. D: Normalized total ion chromato-
gram (TIC) of AQC derivatized linear peptides 11–13 after first C-
terminal amino acid loss showing chromatographic separation and
allowing peptide sequencing. E: Individual MS3 fragmentation
spectra of chromatographically separated derivatized peptides 11, 12
and 13 showing $b_2$ and $b_3$ ions used for peptide sequencing.
The linear peptides 14–28 were identified accordingly.

Table S1 Primers designed for making the hcpA deletion construct in P. chrysogenum.

Table S2 Primers designed for the expression analysis of the hcpA and actin gene.

Table S3 Retention time, formula and acquired m/z of cyclic and linear tetrapeptides obtained from metabolic profiling. Isomers have the same retention time as they could not be chromatographically separated during profiling.

Table S4 ¹H and ¹³C-NMR chemical shifts of synthetically produced compound 2 with sequence cyclo-(d-Tyr-
Phe-d-Val-l-Val) in DMSO at 340 K. As synthetically and naturally produced 2 show identical NMR spectra, only chemical shifts for the synthetically produced compound are shown. $\delta_{\text{DMSO}}$ (¹H/¹³C) = (2.55/40.50), (δ in ppm).

Table S5 ¹H and ¹³C chemical shifts of naturally produced compound 1 with sequence cyclo-(d-Phe-l-Phe-
D-Val-l-Val) present in a mix of various cyclic tetrapeptides in DMSO acquired at 340 K. Signals overlapping with the highest abundant compound 2 are indicated (*). $\delta_{\text{DMSO}}$ (¹H/¹³C) = (2.55/40.50), (δ in ppm).

Table S6 ¹H and ¹³C chemical shifts of isoleucine in compound 5 with sequence cyclo-(Phe-Phe-Val-Ile) and compound 9 with sequence cyclo-(Phe-Phe-Val-Ile) present in a extracted mix of cyclic tetrapeptides. NMR signals corresponding to C = O and NH as well as remaining amino acid signals are not observed (n.o.) due to overlap with the main constituents. $\delta_{\text{DMSO}}$ (¹H/¹³C) = (2.55/40.50).

Table S7 Sequencing data leading to the identification of linear tetrapeptides after AQC derivatization. Multi-
ple-stage fragmentation was used for determination of b-ions for de-novo peptide sequencing.

Table S8 Identified and proposed cyclic and linear tetrapeptides in respect to the adenylation domain specificity in the NRPS HcpA.

Acknowledgments

We thank Oleksandr Salo for processing of the hcpA gene sequence.

Author Contributions

Conceived and designed the experiments: HA MR PL NvP RB RV AD. Performed the experiments: HA MR PL. Analyzed the data: HA MR PL. RvIH OS MN TH NvP RB RV AD. Contributed reagents/materials/analysis tools: RvIH OS NvP. Wrote the paper: HA MR RV AD.
27. Noga MJ, Dane A, Shi S, Attali A, van Aken H, et al. (2012) Metabolomics of cerebrospinal fluid reveals changes in the central nervous system metabolism in a rat model of multiple sclerosis. Metabolomics 8: 253–263.

28. Rottig M, Medema MH, Blin K, Weber T, Rausch C, et al. (2011) NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res 39: W362–367.

29. Bleiholder C, Osburn S, Williams TD, Suhai S, Van Stipdonk M, et al. (2008) Sequence-scrambling fragmentation pathways of protonated peptides. J Am Chem Soc 130: 17774–17789.

30. Cohen SA, Michaud DP (1993) Synthesis of a fluorescent derivatizing reagent, 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate, and its application for the analysis of hydrolysate amino acids via high-performance liquid chromatography. Anal Biochem 211: 279–287.

31. Gehring AM, DeMoll E, Fetherston JD, Mori I, Mayhew GF, et al. (1998) Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by Yersinia pestis. Chem Biol 5: 573–586.

32. Gehring AM, Mori I, Perry RD, Walsh CT (1998) The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of Yersinia pestis. Biochemistry 37: 11637–11650.

33. Suo Z, Tseng CC, Walsh CT (2001) Purification, priming, and catalytic acylation of carrier protein domains in the polyketide synthase and nonribosomal peptidyl synthetase modules of the HMWP1 subunit of yersiniabactin synthetase. Proc Natl Acad Sci U S A 98: 99–104.

34. Gao X, Haynes SW, Ames BD, Wang P, Vien LP, et al. (2012) Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain. Nat Chem Biol 8: 823–830.

35. Keating TA, Elmann DE, Kohli RM, Marshall CG, Trauger JW, et al. (2001) Chain termination steps in nonribosomal peptide synthetase assembly lines: directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis. Chembiochem 2: 99–107.

36. Perez-Victoria I, Martin J, Gonzalez-Mendez V, de Pedro N, El Aouad N, et al. (2012) Isolation and structural elucidation of cyclic tetrapeptides from Onychochaola sclerotica. J Nat Prod 75: 1210–1214.
Supplementary material

A non-canonical NRPS is involved in the synthesis of Fungisporin and related hydrophobic cyclic tetrapeptides in *Penicillium chrysogenum*

Hazrat Ali$^{1,2,§}$, Marco I. Ries$^{3,§}$, Peter P. Lankhorst$^4$, Rob A.M. van der Hoeven$^4$, Olaf L. Schouten$^4$, Marek Noga$^{3,5}$, Thomas Hankemeier$^{3,5}$, Noël N.M.E. van Peij$^5$, Roel A.L. Bovenberg$^{4,6}$, Rob J. Vreeken$^{3,5}$, Arnold J.M. Driessen$^{1,2,*}$

$^1$Molecular Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 7, 9747AG Groningen, The Netherlands
$^2$Kluyver Centre for Genomics of Industrial Fermentations, Julianalaan 67, 2628BC Delft, The Netherlands
$^3$Division of Analytical Biosciences, Leiden Academic Centre for Drug Research, Leiden University, Einsteinweg 55, 2333CC Leiden, The Netherlands
$^4$DSM Biotechnology Center, Alexander Fleminglaan 1, 2613 AX Delft, The Netherlands
$^5$Netherlands Metabolomics Centre, Leiden University, Einsteinweg 55, 2333CC Leiden, The Netherlands
$^6$Synthetic Biology and Cell Engineering, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 7, 9747AG Groningen, The Netherlands

*Correspondence: a.j.m.driessen@rug.nl
§These authors contributed equally to this work
Supplementary Figure 1.
Plasmid used for the deletion of the hcpA gene in P. chrysogenum
Supplementary Figure 2. Multiple-stage fragmentation for de novo sequencing of compound 2 based on sequential amino acid losses

A: MS\(^2\) fragmentation spectra of the cyclic tetrapeptide 2 cyclo-(d-Tyr-I-Phe-d-Val-l-Val). Due to ring opening of the cyclic peptide in the mass spectrometer at different positions, three different amino acid losses occurred, yielding different b\(_3\)-ions.

B-D: MS\(^3\) fragmentation spectra obtained by further fragmenting b\(_3\)-ions from MS\(^2\) showing b\(_2\)-ions used for peptide sequencing.

The cyclic tetrapeptides 1 and 3-10 were identified accordingly.
Supplementary Figure 3.

HSQC spectrum of isolated mixture of cyclic tetrapeptides in DMSO used for the identification of isoleucine present in the minor abundant products 5 and 9. Signals corresponding to isoleucine are indicated with circles. Conducted TOCSY, COSY and HMBC experiments further confirm this conclusion (data not shown). $^1$H and $^{13}$C chemical shifts are shown in Supplementary Table 6.
Supplementary Figure 4.

A: $^1$H-NMR spectrum of a precipitate of various cyclic tetrapeptides containing primarily peptide 1 and 2 (bottom). Synthetic compound 2 spiked to the precipitated mix of various cyclic tetrapeptides in DMSO at 340 K (top). Signals corresponding to 2 were increased as compared to the impurities whereas additional signals did not appear.

B: Zoomed regions of $^1$H-NMR spectrum of natural precipitate (bottom) and precipitate spiked with compound 2 (top). Signals which increased after spiking are indicated (*).
Supplementary Figure 5.

Superimposed HMBC spectra of synthetic compound 2 (red) and an isolated mixture of cyclic tetrapeptides containing naturally produced compound 2 (black). Correlations between NH and CO are shown which indicate identical shifts for both samples. Assignments can be found in Supplementary Table 4.
Supplementary Figure 6. Sequencing of the linear isomers 11 (Phe-Phe-Val-Val), 12 (Val-Phe-Phe-Val) and 13 (Phe-Val-Val-Phe).

A: Extracted ion chromatogram (EIC) of the linear tetrapeptides 11, 12 and 13 using the profiling method. No chromatographic separation could be achieved.

B: MS\(^2\) fragmentation spectrum of unseparated linear tetrapeptides 11-13. Although the spectrum is dominated by fragments originating from 11, several lower abundant fragments can be found originating from a fragmentation of 12 and 13 or possible sequence scrambling of 11.

C: Sequence of linear tetrapeptides after N-terminal AQC derivatization to achieve better chromatographic separation and to prevent sequence scrambling.

D: Normalized total ion chromatogram (TIC) of AQC derivatized linear peptides 11-13 after first C-terminal amino acid loss showing chromatographic separation and allowing peptide sequencing.

E: Individual MS\(^3\) fragmentation spectra of chromatographically separated derivatized peptides 11, 12 and 13 showing b\(_2\) and b\(_1\) ions used for peptide sequencing.

The linear peptides 14 – 28 were identified accordingly.
**Supplementary Table 1.**

Primers designed for making the hcpA deletion construct in *P. chrysogenum*.

| No. | Primer sequence (5’ – 3’) |
|-----|--------------------------|
| attB4F | GGGGACAACCTTTGTATAGAAAAAGTTGGGTACCACAACGCGAATGCGGGGC |
| attB1R | GGGGACTGCTTTTTTGTACAAAACTTGGTGTGTTCAAGTGCAGC |
| attB2F | GGGGACAGCTTTCTTGTACAAAAGTGCCCATGATCAAGGTGCGATCAGGCAATCAGGCG |
| attB3R | GGGGACAACCTTTGTATAATAAAGTTGGGTACCACAATTCGAGCTCGACATGGGCC |

**Supplementary Table 2.**

Primers designed for the expression analysis of the hcpA and actin gene.

| Target | Primer sequence (5’- 3’) |
|--------|--------------------------|
|        | Forward                  | Reverse                  |
| hcpA   | GCTCGTGCCAGGCAGGTCTGCG   | CGACTGGTGACCCTGTTGCACC  |
| actin  | CTGGCGGTATCCACGTCACC     | AGGCCAGAATGGATCCACCG    |

| peptide sequence | RT min | Formula | acquired mass [M+H]+ m/z | accuracy ppm |
|------------------|--------|---------|-------------------------|--------------|
|    1   | cyclo-(FFVV) | 27.33 | C28H36N4O4 | 493.2806 | -0.673   |
|    2   | cyclo-(YFVV) | 26.63 | C28H36N4O5 | 509.2757 | -0.288   |
|    3   | cyclo-(YWVV) | 26.59 | C30H37N5O5 | 548.2868 | 0.099    |
|    4   | cyclo-(FWVV) | 27.21 | C30H37N5O4 | 532.2916 | -0.434   |
|    5   | cyclo-(FVI)  | 27.56 | C29H38N4O4 | 507.2962 | -0.753   |
|    6   | cyclo-(FFIV) | 27.56 | C29H38N4O4 | 507.2962 | -0.753   |
|    7   | cyclo-(YWVI) | 26.86 | C31H39N5O5 | 562.3017 | -1.237   |
|   | Peptide     | RT (min) | Formula       | m/z     | Delta m/z |
|---|-------------|----------|---------------|---------|-----------|
| 8 | cyclo-(YWIV)| 26.86    | C31H39N5O5    | 562.3017| -1.237    |
| 9 | cyclo-(YFVI)| 26.90    | C29H38N4O5    | 523.2914| -0.185    |
|10 | cyclo-(YFIV)| 26.90    | C29H38N4O5    | 523.2914| -0.185    |
|11 | FVVF        | 23.34    | C28H38N4O5    | 511.2912| -0.580    |
|12 | VFFV        | 23.34    | C28H38N4O5    | 511.2912| -0.580    |
|13 | FFVV        | 23.34    | C28H38N4O5    | 511.2912| -0.580    |
|14 | YFVV        | 20.06    | C28H38N4O6    | 527.2864| -0.022    |
|15 | VYFV        | 20.06    | C28H38N4O6    | 527.2864| -0.022    |
|16 | FYVV        | 20.06    | C28H38N4O6    | 527.2864| -0.022    |
|17 | YWVV        | 20.40    | C30H39N5O6    | 566.2972| -0.195    |
|18 | VYVV        | 20.40    | C30H39N5O6    | 566.2972| -0.195    |
|19 | WVYY        | 20.40    | C30H39N5O6    | 566.2972| -0.195    |
|20 | VFWV        | 23.57    | C30H39N5O5    | 550.3023| -0.174    |
|21 | FWVV        | 23.57    | C30H39N5O5    | 550.3023| -0.174    |
|22 | WVVF        | 23.57    | C30H39N5O5    | 550.3023| -0.174    |
|23 | FVIF        | 24.66    | C29H40N4O5    | 525.3077|  1.053    |
|24 | FIVF        | 24.66    | C29H40N4O5    | 525.3077|  1.053    |
|25 | FVIV        | 21.43    | C29H40O6N4    | 541.3023|  0.441    |
|26 | IFVY        | 21.43    | C29H40O6N4    | 541.3023|  0.441    |
|27 | FIVY        | 21.43    | C29H40O6N4    | 541.3023|  0.441    |
|28 | VYFI        | 21.43    | C29H40O6N4    | 541.3023|  0.441    |

**Supplementary Table 3.**

Retention time, formula and acquired m/z of cyclic and linear tetrapeptides obtained from metabolic profiling. Isomers have the same retention time as they could not be chromatographically separated during profiling.
### Supplementary Table 4.

$^1$H and $^{13}$C-NMR chemical shifts of synthetically produced compound 2 with sequence *cyclo-*(d-Tyr-l-Phe-d-Val-l-Val) in DMSO at 340 K. As synthetically and naturally produced 2 show identical NMR spectra, only chemical shifts for the synthetically produced compound are shown. $\delta_{\text{DMSO}}$ ($^1$H/$^{13}$C) = (2.55/40.50), ($\delta$ in ppm).

| Amino Acid | Position | $\delta$ ($^1$H) | Position | $\delta$($^{13}$C) |
|------------|----------|-----------------|----------|------------------|
| $d$-Tyr    | $\alpha$ | 4.53            | $\alpha$ | 54.10            |
|            | $\beta_1$| 2.93            | $\beta$  | 34.59            |
|            | $\beta_2$| 2.69            | $\gamma$ | 128.43           |
|            | $\delta$ | 6.99            | $\delta$ | 130.32           |
|            | $\epsilon$| 6.65           | $\epsilon$ | 115.63         |
|            | NH       | 7.82            | $\zeta$  | 156.35           |
|            | OH       | 8.99            | C=O      | 173.35           |
| l-Phe      | $\alpha$ | 4.62            | $\alpha$ | 53.97            |
|            | $\beta_1$| 3.05            | $\beta$  | 35.47            |
|            | $\beta_2$| 2.80            | $\gamma$ | 138.48           |
|            | $\delta$ | 7.20            | $\delta$ | 129.41           |
|            | $\epsilon$| 7.21           | $\epsilon$ | 128.68         |
|            | $\zeta$  | 7.26            | $\zeta$  | 126.72           |
|            | NH       | 7.80            | C=O      | 173.56           |
| $d$-Val    | $\alpha$ | 4.00            | $\alpha$ | 59.90            |
|            | $\beta$  | 2.00            | $\beta$  | 27.66            |
|            | $\gamma_1$| 0.87           | $\gamma_1$ | 19.76         |
|            | $\gamma_2$| 0.82           | $\gamma_2$ | 19.01         |
|            | NH       | 7.58            | C=O      | 173.81           |
| l-Val      | $\alpha$ | 3.98            | $\alpha$ | 59.71            |
|            | $\beta$  | 2.00            | $\beta$  | 27.51            |
|            | $\gamma_1$| 0.90           | $\gamma_1$ | 19.73        |
|            | $\gamma_2$| 0.80           | $\gamma_2$ | 19.07         |
|            | NH       | 7.62            | C=O      | 173.32           |
| amino acid | position | δ (1H) | Position | δ(13C) |
|------------|----------|--------|----------|--------|
| d-Phe      | α        | 4.62*  | α        | 53.69  |
|            | β1       | 3.05*  | β        | 35.29  |
|            | β2       | 2.80*  | γ        | 138.40 |
|            | δ        | 7.20*  | δ        | 129.43 |
|            | ε        | 7.21*  | ε        | 128.66 |
|            | ζ        | 7.26*  | ζ        | 126.74 |
|            | NH       | 7.88   | C=O      | 173.16 |
| l-Phe      | α        | 4.62*  | α        | 54.03  |
|            | β1       | 3.05*  | β        | 35.56  |
|            | β2       | 2.80*  | γ        | 138.44 |
|            | δ        | 7.20*  | δ        | 129.42 |
|            | ε        | 7.21*  | ε        | 128.68*|
|            | ζ        | 7.26*  | ζ        | 126.72*|
|            | NH       | 7.78   | C=O      | 173.56*|
| d-Val      | α        | 4.00*  | α        | 59.88  |
|            | β        | 2.00*  | β        | 27.61  |
|            | γ1       | 0.86   | γ1       | 19.77  |
|            | γ2       | 0.81   | γ2       | 19.01* |
|            | NH       | 7.58   | C=O      | 173.79 |
| l-Val      | α        | 3.98*  | α        | 59.73  |
|            | β        | 2.00*  | β        | 27.54  |
|            | γ1       | 0.89   | γ1       | 19.67  |
|            | γ2       | 0.78   | γ2       | 19.06  |
|            | NH       | 7.58   | C=O      | 173.36 |

**Supplementary Table 5.**

1H and 13C chemical shifts of naturally produced compound 1 with sequence cyclo-(d-Phe-l-Phe-d-Val-l-Val) present in a mix of various cyclic tetrapeptides in DMSO acquired at 340 K. Signals overlapping with the highest abundant compound 2 are indicated (*). δ_DMSO (1H/13C) = (2.55/40.50), (δ in ppm).
## Supplementary Table 6.

$^1$H and $^{13}$C chemical shifts of isoleucine in compound 5 with sequence cyclo-(Phe-Phe-Val-Ile) and compound 9 with sequence cyclo-(Phe-Phe-Val-Ile) present in an extracted mix of cyclic tetrapeptides. NMR signals corresponding to C=O and NH as well as remaining amino acid signals are not observed (n.o.) due to overlap with the main constituents. $\delta_{\text{DMSO}}$ ($^1$H/$^{13}$C) = (2.55/40.50).
| linear peptide | underivatized sequence | RT (min) | formula derivatized peptide | [M+H]^+ | b₂ ion in MS² | b₂ ion in MS³ derived from b₃ ion | b₁ ion in MS³ derived from b₃ ion |
|----------------|------------------------|----------|----------------------------|---------|-------------|-------------------------------|-------------------------------|
| 11 FVVF        | 7.50                   | C38H44N6O6 | 681 | 516 (-F) | 417 (-V) | 318 (-VV)                     |                               |
| 12 VFFV        | 7.71                   | C38H44N6O6 | 681 | 564 (-V) | 417 (-F) | 270 (-FF)                     |                               |
| 13 FFVV        | 7.98                   | C38H44N6O6 | 681 | 564 (-V) | 465 (-V) | 318 (-FV)                     |                               |
| 14 YFVV        | 4.46                   | C38H44N6O7 | 697 | 580 (-V) | 481 (-V) | 334 (-FV)                     |                               |
| 15 VYFV        | 3.28                   | C38H44N6O7 | 697 | 580 (-V) | 433 (-F) | 270 (-YF)                     |                               |
| 16 FVYV        | 3.85                   | C38H44N6O7 | 697 | 516 (-Y) | 417 (-V) | 318 (-VV)                     |                               |
| 17 YWVV        | 4.20                   | C40H45N7O7 | 736 | 619 (-V) | 520 (-V) | 334 (-WV)                     |                               |
| 18 VYWV        | 2.55                   | C40H45N7O7 | 736 | 619 (-V) | 433 (-W) | 270 (-YW)                     |                               |
| 19 WVYV        | 3.64                   | C40H45N7O7 | 736 | 555 (-Y) | 456 (-V) | 357 (-VV)                     |                               |
| 20VFVV         | 7.01                   | C40H45N7O6 | 720 | 603 (-V) | 417 (-W) | 270 (-FW)                     |                               |
| 21 FWVV        | 7.78                   | C40H45N7O6 | 720 | 603 (-V) | 504 (-V) | 318 (-WV)                     |                               |
| 22 WVVF        | 7.41                   | C40H45N7O6 | 720 | 555 (-F) | 546 (-V) | 357 (-VV)                     |                               |
| 23 FVIF        | 8.24                   | C39H46N6O6 | 695 | 530 (-F) | 417 (-I) | 318 (-VI)                     |                               |
| 24 FIYV        | 8.38                   | C39H46N6O6 | 695 | 530 (-F) | 431 (-V) | 318 (-IV)                     |                               |
| 25 FVIY        | 5.58                   | C39H46N6O7 | 711 | 530 (-Y) | 417 (-I) | 318 (-VI)                     |                               |
| 26 IYFV        | 4.62                   | C39H46N6O7 | 711 | 594 (-V) | 447 (-F) | 284 (-FY)                     |                               |
| 27 FIVY        | 6.14                   | C39H46N6O7 | 711 | 530 (-Y) | 431 (-V) | 318 (-IV)                     |                               |
| 28 VYFI        | 5.52                   | C39H46N6O7 | 711 | 580 (-I) | 433 (-F) | 270 (-YF)                     |                               |

**Supplementary Table 7.**

Sequencing data leading to the identification of linear tetrapeptides after AQC derivatization. Multiple-stage fragmentation was used for determination of b-ions for de-novo peptide sequencing.
| cyclic peptide | specificity of adenylation domain | corresponding linear peptide |
|---------------|----------------------------------|-------------------------------|
|               | A₁   | A₂   | A₃   | A₄   |                           |
| 1             | Phe  | Phe  | Val  | Val  | 11-13                       |
| 2             | Tyr  | Phe  | Val  | Val  | 14-16                       |
| 3             | Tyr  | Trp  | Val  | Val  | 17-19                       |
| 4             | Phe  | Trp  | Val  | Val  | 20-22                       |
| 5             | Phe  | Phe  | Val  | Ile  | 23                          |
| 6             | Phe  | Phe  | Ile  | Val  | 24                          |
| 7             | Tyr  | Trp  | Val  | Ile  | *not observed*              |
| 8             | Tyr  | Trp  | Ile  | Val  | *not observed*              |
| 9             | Tyr  | Phe  | Val  | Ile  | 25-26                       |
| 10            | Tyr  | Phe  | Ile  | Val  | 27-28                       |
| *proposed*    | Phe  | Trp  | Val  | Ile  | *not observed*              |
| *proposed*    | Phe  | Trp  | Ile  | Val  | *not observed*              |
| *proposed*    | Phe  | Phe  | Ile  | Ile  | *not observed*              |
| *proposed*    | Phe  | Trp  | Ile  | Ile  | *not observed*              |
| *proposed*    | Tyr  | Phe  | Ile  | Ile  | *not observed*              |
| *proposed*    | Tyr  | Trp  | Ile  | Ile  | *not observed*              |

**Supplementary Table 8.**

Identified and proposed cyclic and linear tetrapeptides in respect to the adenylation domain specificity in the NRPS HcpA.