Organic &
Biomolecular
Chemistry

Cite this: *Org. Biomol. Chem.*, 2011, **9**, 3130

www.rsc.org/obc

COMMUNICATION

Structure elucidation and biosynthesis of lysine-rich cyclic peptides in *Xenorhabdus nematophila*†

Sebastian W. Fuchs,[a] Anna Proschak,[a] Thorsten W. Jaskolla,[b] Michael Karas[b] and Helge B. Bode*[a]

Received 18th January 2011, Accepted 9th February 2011
DOI: 10.1039/c1ob05097d

Thirteen novel PAX (peptide-antimicrobial-*Xenorhabdus*) peptides were identified in *Xenorhabdus nematophila* HGB081. Their structures including the absolute configuration were elucidated using a combination of labeling experiments, detailed MS/MS experiments, the advanced Marfey's method, and a detailed analysis of the biosynthesis gene cluster, which was identified as well.

Entomopathogenic bacteria of the genus *Xenorhabdus* are well known for their capability to kill insect larvae.[1] This pathogenicity requires symbiotic interaction with nematodes of the genus *Steinernema*, which undergo a complex life cycle in the insect hemocoel after infection.[2] During this life cycle the nematode releases *Xenorhabdus* bacteria into the hemocoel, which then kill the insect larvae by bacterial pathogenicity factors such as insecticidal proteins.[3] To protect the insect corpse from food competitors, *Xenorhabdus* synthesizes an array of secondary metabolites.[4,5] Examples are antibacterial indole derivatives as well as xenorhabdins and xenocoumacins, the latter two exhibit antibacterial and antifungal activity.[4] Recently, Gualtieri and co-workers identified a group of five lysine-rich cyclolipopeptides (PAX-peptides) from *X. nematophila* F1/1, exhibiting strong antifungal activity against the opportunistic human pathogen *Fusarium oxysporum* as well as several plant pathogenic fungi.[5] After analysis of culture supernatants from *X. nematophila* HGB081, we were able to identify thirteen PAX-peptides. The structures of these derivatives were solved following detailed MS/MS experiments, feeding experiments as well as chemical derivatization followed by MS analysis. Additionally, the biosynthesis gene cluster for the production of the PAX peptides was identified confirming the stereochemical data as well as allowing the prediction of the full stereochemistry of these unusual peptides.

Internally calibrated high resolution (HR) matrix-assisted laser desorption/ionization mass spectra (MALDI-MS) of the culture supernatant from *X. nematophila* HGB081 allowed for detection of the PAX peptides characterized by Gualtieri and co-workers within an error of <1.8 ppm by means of a MALDI-LTQ Orbitrap XL (Fig. 1 and 2, compounds **2**, **3**, **5**, and **7**; Table S1†)[5]

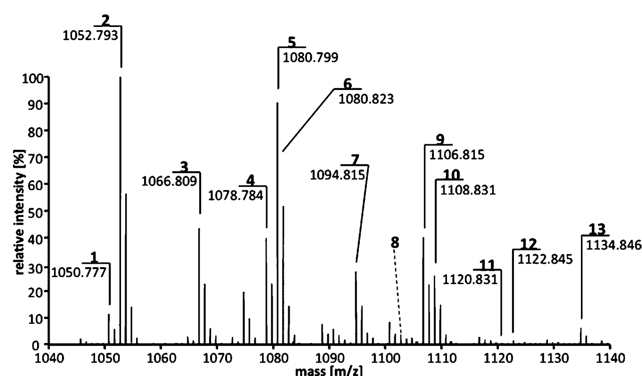


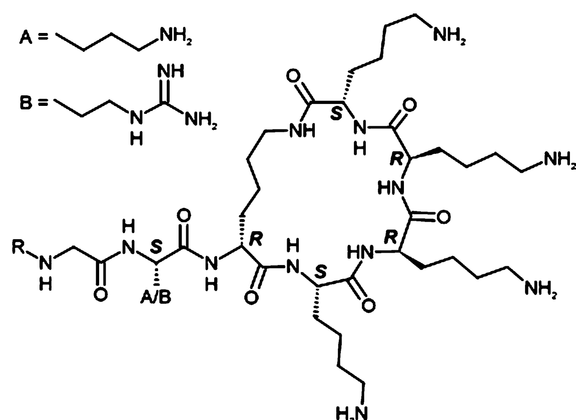
Fig. 1 MALDI-Orbitrap mass spectrum of a culture supernatant from *X. nematophila* HGB081. Monoisotopic *m*/*z*-values of putative PAX-derivatives were labeled with their compound number (Table S1†).

These PAX peptides were accompanied by additional unknown signals. By converting their respective accurate mass-over-charge (*m*/*z*) ratios into chemical compositions and taking into account the maximum measured error of 1.8 ppm, these signals could be assigned to additional PAX peptide derivatives exhibiting variations in the fatty acid side chains (Table S1†). Tandem mass spectrometry experiments of compounds **5** (Fig. S1†), **9** (Fig. S2†) as well as of **2**, **3**, **10**, and **13** (Table S2†) confirmed the proposed primary structures of the novel PAX-derivatives. Additionally, the structures of the novel PAX-derivatives were confirmed by feeding experiments and consecutive HR-mass spectrometry. After cultivation of *X. nematophila* HGB081 in [U-^{13}C] medium following the addition of the respective non-labeled (^{12}C) amino acids, we observed the complete incorporation of five [U-^{12}C]lysine moieties into compounds **5** and **6**, (Fig. 3; Table S1†) whereas the incorporation of the sixth and fifth [U-^{12}C]lysine, respectively, could not be detected for compounds **2**, **10**, and **13**. As expected, the incorporation of one [U-^{12}C]arginine was observed for compounds **5**, **9**, **10**, and **13** (Fig. 3; Table S1†). Since production in the [U-^{13}C] medium was lower than in LB-medium, not all described PAX-derivates (Table S1†) could be observed in this experiment.

Even though these results indicate that there are no differences in the primary structures of the PAX peptides produced by *X. nematophila* strains HGB081 and F1/1, the MS/MS-fragmentation

[a]*Molekulare Biotechnologie, Institut für Molekulare Biowissenschaften, Goethe Universität Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany. E-mail: h.bode@bio.uni-frankfurt.de; Fax: +49 69 798 29527; Tel: +49 69 798 29557*
[b]*Cluster of Excellence Macromolecular Complexes, Institute of Pharmaceutical Chemistry, Goethe-University Frankfurt, 60438 Frankfurt, Germany*
† Electronic supplementary information (ESI) available: General methods, HPLC chromatograms, and MS-data. See DOI: 10.1039/c1ob05097d

This journal is © The Royal Society of Chemistry 2011



No.	Name	A/B	R
1	**PAX5**	**A**	**(3*R*,7*Z*)-3-hydroxytetradec-7-enoyl**
2	**PAX1'**	**A**	**(3*R*)-3-hydroxytetradecanoyl**
3	**PAX3'**	**A**	**(3*R*)-3-hydroxypentadecanoyl**
4	**PAX6**	**B**	**(3*R*,7*Z*)-3-hydroxytetradec-7-enoyl**
5	**PAX2'**	**B**	**(3*R*)-3-hydroxytetradecanoyl**
6	**PAX7**	**A**	**(3*R*)-3-hydroxyhexadecanoyl**
7	**PAX4'**	**B**	**(3*R*)-3-hydroxypentadecanoyl**
8	**PAX8**	**A**	**(3*R*)-3-hydroxyoctadecanoyl**
9	**PAX9**	**B**	**(3*R*,9*Z*)-3-hydroxyhexadec-9-enoyl**
10	**PAX10**	**B**	**(3*R*)-3-hydroxyhexadecanoyl**
11	**PAX11**	**B**	**(3*R*,10*Z*)-3-hydroxyheptadec-10-enoyl**
12	**PAX12**	**B**	**(3*R*)-3-hydroxyheptadecanoyl**
13	**PAX13**	**B**	**(3*R*,11*Z*)-3-hydroxyoctadec-11-enoyl**

Fig. 2 Comprehensive depiction of all identified PAX-peptides from *X. nematophila* HGB081. Compounds with apostrophes differ from the known PAX peptides by the fatty acid moiety.

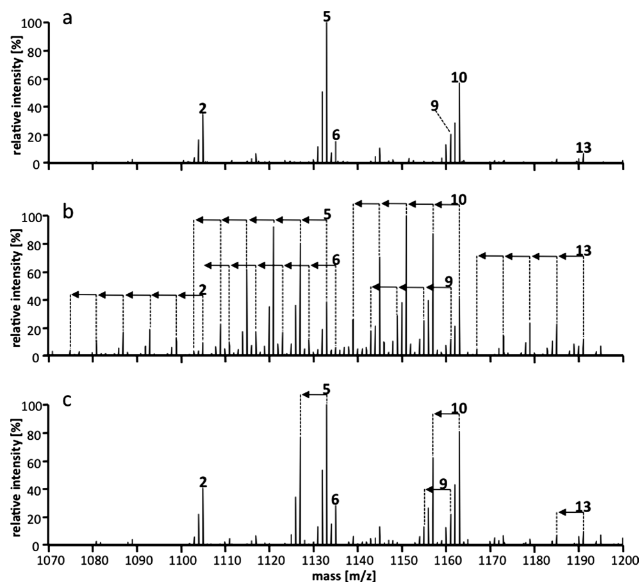


Fig. 3 MALDI-Orbitrap mass spectra of culture supernatants from *X. nematophila* HGB081 cultivated in [U-^{13}C] medium (a) and supplemented with L-[U-^{12}C]lysine (b) or L-[U-^{12}C]arginine (c).

patterns as well as the feeding experiments could not reveal the position of the lysine-residue participating in the cyclic peptide bond formation with the C-terminus, which has been shown to be Lys^3 for strain F1/1. To clarify this point, we acetylated the lysine residues in PAX peptides from strain HGB081 with acetic acid anhydride[6] to label those lysine side chains that are not involved in the cyclization (Fig. S3†, Table S1†). Analysis of the MS/MS-ions of the fully acetylated PAX peptides **5** and **9** (Fig. S4–S5†) unambiguously revealed the cyclization between the ε-amino group of Lys^3 and the C-terminus, since all but the originally cyclized side chain of Lys^3 were acetylated. Thus the peptide backbones of PAX peptides from strain HGB081 and F1/1 are identical.

NMR-experiments conducted by Gualtieri and co-workers[5] revealed that the fatty acids in PAX peptides from *X. nematophila* F1/1 are iso-branched. To test whether this also applies to PAX peptides from *X. nematophila* HGB081, we fed L-[5,5,5-$^2H_3$]leucine and DL-[2,3,4,4,4,5,5,5-$^2H_8$]valine as precursors of iso-branched fatty acids.[7] While incorporation of labeled valine or leucine was observed for selected secondary metabolites (*e.g.* xenocoumacins[8] and other peptides), no such incorporations were observed for any PAX peptides (data not shown), suggesting that all fatty acids in PAX peptides in strain HGB081 are straight chain (sc) fatty acids. This was confirmed by the incorporation of [U-$^2H_6$]propionic acid into fatty acids with an uneven number of carbon atoms (PAX3′ (**3**), PAX4′ (**7**), PAX11 (**11**), and PAX12 (**12**) (Fig. S6†). The position of the double bond was deduced from a fatty acid analysis of strain HGB081, which showed only ω7-desaturated fatty acids (Table S3†). Accordingly, we assume the fatty acid moiety of **11** to be (3*R*)-3-hydroxyheptadec-10-enoic acid, although this fatty acid has not been identified in the whole-cell fatty acid analysis probably due to its minute amounts. Based on these results, thirteen PAX peptides could be identified in strain HGB081 differing in the amino acid at the 2-position as well as the N-terminally linked fatty acids (Fig. 2).

However, information regarding the absolute configuration of the PAX peptides was still missing. In order to determine the stereochemistry of the 3-OH fatty acid, fatty acids from an enriched sample of PAX peptides (Fig. S7†) were converted into their methyl esters after peptide hydrolysis and 3-OH-groups were consecutively derivatized with (*R*)-(–)-α-methoxy-α-(trifluoromethyl)phenylacetyl chloride ((*R*)-(–)-MTPA-Cl) according to the method of Jenske and Vetter.[9] Comparison of the retention times with a (*R*)-(–)-MTPA-Cl-derivatized fatty acid methyl ester-standard containing racemic 14:0-3OH-fatty acid revealed the configuration of the 14:0-3OH fatty acids in PAX peptides (Fig. S8†, Table S4†). The derivatized 14:0-3OH fatty acid from the PAX peptide sample eluted after 23.7 min and racemic 14:0-3OH fatty acids from the standard eluted after 23.6 min and 23.7 min. According to Jenske and Vetter[9] as well as Ring and co-workers,[10] *S*-enantiomers always elute prior to the corresponding *R*-enantiomers, thus 14:0-3OH fatty acids from PAX-peptides **2** and **5** likely exhibit an *R*-configuration.

To determine the absolute configuration of the amino acids, an SPE-enriched mixture of PAX peptides was analyzed according to the advanced Marfey's method[11,12] resulting in the detection of L-arginine and D- and L-lysine (Table S5†, Fig. S9†). In order to differentiate between the different lysine residues, we tried to determine the absolute configuration of Lys^3 whose ε-amino group is involved in amide bond formation with the C-terminus. Reductive methylation of enriched PAX peptides by means of paraformaldehyde and $NaBH_3CN$ selectively dimethylated lysine residues with free ε-amines while the side chain of cyclized Lys^3 remained unmodified as confirmed by the formation of the

expected octamethyl and decamethyl derivatives for PAX peptides carrying five or six lysine residues, respectively (data not shown). After acid hydrolyzation of the methylated sample followed by Marfey's analysis the configuration of Lys[3] was assigned as D (Table S5†, Fig. S9†).[13]

In order to determine the configuration of Lys[4]–Lys[7], we wanted to analyze the biosynthesis gene cluster involved in the PAX peptide production. We have identified 17 biosynthesis gene clusters in the genome of *X. nematophila* HGB081, which might be involved in the biosynthesis of non-ribosomal peptides, polyketides and hybrids thereof (unpublished data). A detailed analysis of one of these biosynthesis gene clusters (Fig. 4) revealed that it might be involved in the biosynthesis of a lipoheptapeptide starting with a fatty acid followed by a glycine as expected for the PAX peptides. In order to prove its function, we constructed a plasmid insertion mutant into gene *xnc1_2781*.

Fig. 4 (a) Schematic illustration of the PAX-peptide biosynthesis gene cluster in *X. nematophila* HGB081. The depicted region comprises 26.3 kbp and encodes three individual NRPS-genes as well as a putative membrane transporter. (b) Proposed biosynthesis of PAX-peptide **5** in strain HGB081.

A comparative MALDI-time of flight analysis of strain HGB081 and the plasmid insertion mutant revealed that the production of PAX peptides was completely abolished in the case of the insertion mutant (Fig. S10†). This unambiguously identified the gene cluster responsible for PAX-biosynthesis (Fig. 4a), which exhibited three open reading frames encoding NRPSs (*paxABC*) as well as one open reading frame coding for a transporter (*paxT*). A detailed analysis of all three NRPS open reading frames revealed eight individual modules containing one C (condensation), A (adenylation), and T (thiolation) domain each.[14] The A-domain encoded by *paxA* was predicted to incorporate glycine (Table S6†). No specificity could be predicted for the remaining six A-domains but showed that A2 was slightly different to A3–A7 (Table S6†). An alignment of all C-domains in the PAX-biosynthesis gene cluster with known C-domains as well as bifunctional C/E (condensation/epimerisation) domains[15] revealed that the third C-domain of PaxB as well as the second and third C-domain of PaxC exhibit C/E-domain-character (Fig. S11†; Table S6†). C/E-domains have been shown to catalyze the condensation of amino acids with the non-ribosomal peptide intermediate as well as the epimerization of the amino acid residue incorporated by the previous module. This finding is in accordance with the results described above which revealed the absolute configuration of L-Arg[2] and D-Lys[3]. Thus we feel confident to assign the absolute configuration to the remaining amino acids according to the presence of C/E-domains in the different modules as D for Lys[5] and Lys[6] (Fig. 2 and 4).

In conclusion, here we have elucidated the structure of novel PAX peptides including their absolute configuration. We did this using crude mixtures or partially purified peptide fractions by applying a combination of MS analyses, labeling experiments, chemical derivatization, and analysis of the biosynthesis gene clusters and could demonstrate the power of this approach. Thus, our work has set the stage for the synthesis and mode of action studies of the PAX peptides and can be applied for similar compound classes avoiding time consuming and difficult initial isolation of these compounds.

Acknowledgements

The authors are grateful to the *Xenorhabdus* genome consortium for access to the genome sequence prior to publication and to Wolfram Lorenzen for help with the fatty acid analysis. Financial support for this project was provided in part by the DFG and the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 223328.

References

1 S. Forst, B. Dowds, N. E. Boemare and E. Stackebrandt, *Annu. Rev. Microbiol.*, 1997, **51**, 47–72.
2 E. E. Herbert and H. Goodrich-Blair, *Nat. Rev. Microbiol.*, 2007, **5**, 634–646.
3 S. E. Brown, A. T. Cao, P. Dobson, E. R. Hines, R. J. Akhurst and P. D. East, *Appl. Environ. Microbiol.*, 2006, **72**, 1653–1662.
4 H. B. Bode, *Curr. Opin. Chem. Biol.*, 2009, **13**, 224–230.
5 M. Gualtieri, A. Aumelas and J. O. Thaler, *J. Antibiot.*, 2009, **62**, 295–302.
6 D. Suckau, M. Mak and M. Przybylski, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 5630–5634.
7 R. L. Conner, K. E. Koo and J. R. Landrey, *Lipids*, 1974, **9**, 554–559.
8 D. Reimer, E. Luxenburger, A. O. Brachmann and H. B. Bode, *ChemBioChem*, 2009, **10**, 1997–2001.
9 R. Jenske and W. Vetter, *J. Chromatogr., A*, 2007, **1146**, 225–231.
10 M. W. Ring, G. Schwär and H. B. Bode, *ChemBioChem*, 2009, **10**, 2003–2010.
11 K. Fujii, Y. Ikai, H. Oka, M. Suzuki and K. Harada, *Anal. Chem.*, 1997, **69**, 5146–5151.
12 K. Fujii, T. Shimoya, Y. Ikai, H. Oka and K. Harada, *Tetrahedron Lett.*, 1998, **39**, 2579–2582.
13 K. Fujii, Y. Ikai, T. Mayumi, H. Oka, M. Suzuki and K. Harada, *Anal. Chem.*, 1997, **69**, 3346–3352.
14 http://nrps.igs.umaryland.edu/nrps/.
15 C. J. Balibar, F. H. Vaillancourt and C. T. Walsh, *Chem. Biol.*, 2005, **12**, 1189–1200.

This journal is © The Royal Society of Chemistry 2011