CHEMISTRY
A EUROPEAN JOURNAL

DOI: 10.1002/chem.201103479

Determination of the Absolute Configuration of Peptide Natural Products by Using Stable Isotope Labeling and Mass Spectrometry

Helge B. Bode,* Daniela Reimer, Sebastian W. Fuchs, Ferdinand Kirchner, Christina Dauth, Carsten Kegler, Wolfram Lorenzen, Alexander O. Brachmann, and Peter Grün[a]

Abstract: Structure elucidation of natural products including the absolute configuration is a complex task that involves different analytical methods like mass spectrometry, NMR spectroscopy, and chemical derivation, which are usually performed after the isolation of the compound of interest. Here, a combination of stable isotope labeling of *Photorhabdus* and *Xenorhabdus* strains and their transaminase mutants followed by detailed MS analysis enabled the structure elucidation of novel cyclopeptides named GameXPeptides including their absolute configuration in crude extracts without their actual isolation.

Keywords: configuration determination · isotopic labeling · mass spectrometry · peptides · structure elucidation

Introduction

An important and time-consuming step in natural product research is the structural elucidation of novel compounds. Usually a crude extract from the producer of a natural product is analyzed by HPLC/UV or HPLC/MS to clarify whether the respective peak in the chromatogram represents a compound that is worth its isolation. In that case, the compound of interest must be purified and its structure subsequently elucidated by means of NMR spectroscopic analysis.[1] Today, high-resolution mass spectrometry allows the determination of the mass of the respective compound to such accuracy that a sum formula can be predicted, which can be used for database searches for rapid identification of novel compounds. Whereas the possible compositions for protein-derived peptides or normal lipids would allow only a certain number of possible sum formulae due to their known building-block composition,[2,3] this is different for natural products, which can be built from very unusual precursors and thus increase the number of possible sum formulae. To address this problem and to speed up the identification and structure elucidation of natural products especially from microorganisms, one can apply an isotope-labeling strategy followed by mass spectrometry. This approach allows the reliable determination of the correct sum formula and has been used extensively for metabolomics in plants and bacteria.[4–8] Here we describe the application and further development of this simple and robust approach for the structure elucidation of natural products from entomopathogenic bacteria, applied to four novel cyclopeptides. The approach allows for differentiation between isobar building blocks such as leucine or isoleucine, as well as the determination of the absolute configuration in the case of amino acids.

[a] Prof. Dr. H. B. Bode, D. Reimer, Dipl.-Biol. S. W. Fuchs, F. Kirchner, Dipl.-Chem. C. Dauth, Dr. C. Kegler, Dipl.-Pharm. W. Lorenzen, Dr. A. O. Brachmann, P. Grün
Institut für Molekulare Biowissenschaften
Molekulare Biotechnologie, Goethe Universität Frankfurt
Max-von-Laue-Strasse 9, 60436 Frankfurt am Main (Germany)
Fax: (+49) 69-798-29527
E-mail: h.bode@bio.uni-frankfurt.de

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201103479.

Results and Discussion

Sum formula determination: Entomopathogenic bacteria of the genus *Xenorhabdus* and *Photorhabdus* were analyzed because these bacteria are usually multiproducers of several different natural products simultaneously.[9] Additionally, highly similar compounds have been observed in different strains, and thus a method was needed to differentiate these compounds. When *Photorhabdus luminescens* strain TT01 was grown in standard growth medium, the sum formulas of all peaks that appeared in an HPLC/MS analysis (Figure 1a) could be predicted from high-resolution MS analysis, as exemplified by compound **1** with *m*/*z* 586.39545 (Figure 1b). However, because usually several sum formulae are chemically and biologically possible, it is often difficult to find the correct one. It was previously shown that the number of possible chemical formulae can be significantly reduced by taking into account the isotope pattern of ultrahigh-resolution MS data.[10] Similarly, growing bacteria in culture media fully labeled with ^{13}C drastically reduces the number of possible sum formulae for lipids[11] and therefore we have applied this approach to *Photorhabdus luminescens*. When comparing HPLC/MS analyses of strain TT01 grown in standard growth medium (natural abundance of all isotopes),

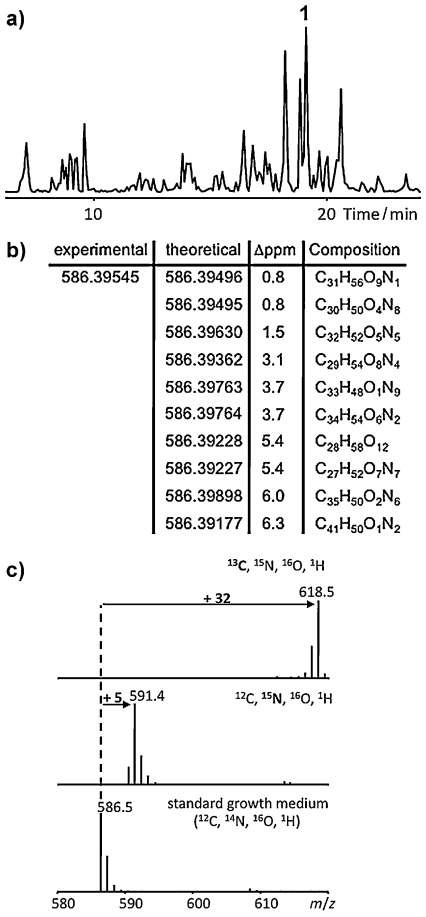


Figure 1. a) HPLC/MS analysis (base peak chromatogram) of *P. luminescens* TT01 highlighting compound **1** (arrow). b) Sum formula prediction for **1** using HRESI-MS data. c) Determination of the number of carbon and nitrogen atoms for **1** as determined from growth of strain TT01 in standard growth medium, or medium fully labeled with ^{15}N or ^{13}C.

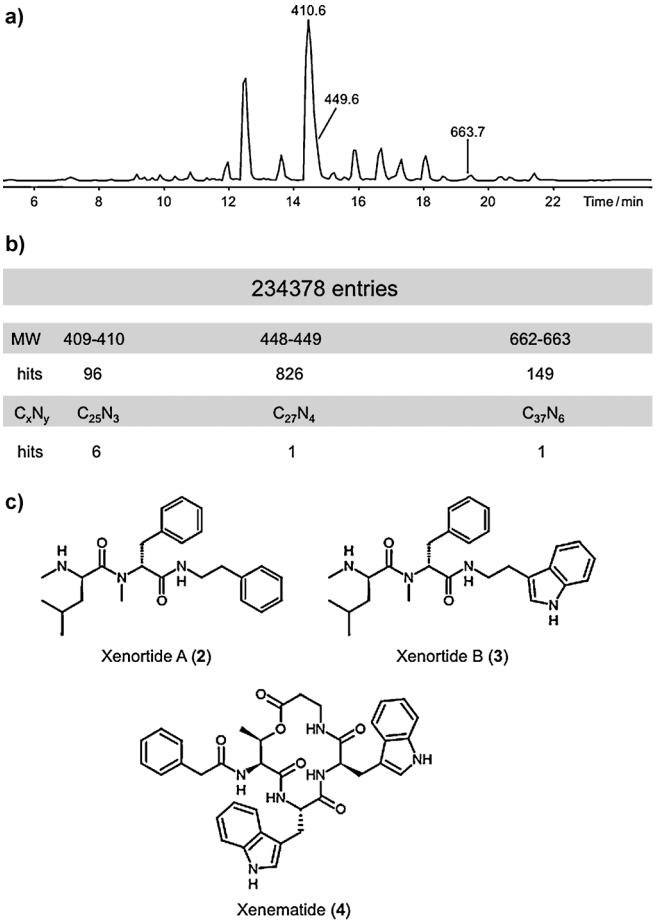


Figure 2. a) HPLC/MS analysis of *X. nematophila* HGB081 indicating the positions of compounds **2**–**4**. b) Identification of **2**–**4** in the Dictionary of Natural Products (http://dnp.chemnetbase.com/dictionary-search.do?method=view&id=2265390&struct=start&props=&&si= last accessed on November 3, 2011) based on the mass range and the number of carbon and nitrogen atoms identified from labeling experiments. c) Structures of xenortide A (**2**), xenortide B (**3**), and xenematide (**4**).[7]

^{15}N or ^{13}C medium allowed the rapid identification of the correct sum formula as the number of nitrogen and carbon atoms can be easily determined from the mass shifts of **1** in the different growth media (Figure 1c). Despite its simplicity, this approach is also cheap as culture volumes as little as 1 mL can be used (media costs are around 1 €mL^{-1}). Moreover, the sum formulae for all compounds produced under the selected growth conditions (even produced in minute amounts) can be determined in parallel (Table S3 in the Supporting Information) as has also been described in metabolome experiments in different organisms.[4–8] Indeed, the approach is so powerful that it works even without HRMS data. Analysis of *Xenorhabdus nematophila* HGB081 led to the identification of compounds **2**–**4** with *m*/*z* 410.6, 449.6, and 663.7. A database search in the Dictionary of Natural Products with a total of 234378 entries using the number of carbon and nitrogen atoms ($C_{25}N_3$, $C_{27}N_4$, and $C_{37}N_6$) as determined by labeling experiments and the molecular-weight range (409–410, 448–449, and 662–663) resulted in only six, one, and one hit, respectively, and thus led to the identification of xenortide A (**2**) and xenortide B (**3**), and xenematide (**4**), which were already known from another *Xenorhabdus* strain (Figure 2).[12]

Building-block determination: Similarly, this labeling approach can be used to identify building blocks of natural products such as amino acids. Here, labeled precursors (^{13}C, ^2H, ^{15}N) are usually added to a producing culture and their incorporation would confirm the involvement of the precursor in the natural product biosynthesis. Again, MS-based detection allows for the analysis of compounds produced in minute amounts. However, because not all possible precursors are available for a reasonable price, one can also add a [$^{12}C/^{14}N$] precursor with natural abundance to a culture grown in a fully labeled ^{13}C or ^{15}N medium as described above. Precursor incorporation in such an 'inverse' labeling experiment is easily visible by a shift to lower masses. In the case of compound **1**, its composition was determined to be cyclo(FLLLV) or a positional isomer thereof, as shown by inverse labeling experiments with valine (Val), leucine

(Leu), and phenylalanine (Phe) in ^{13}C medium (Figure 3). The connectivity of these building blocks can be determined easily from MS-MS experiments as shown for **1** (Figure S1 in the Supporting Information), which was demonstrated to

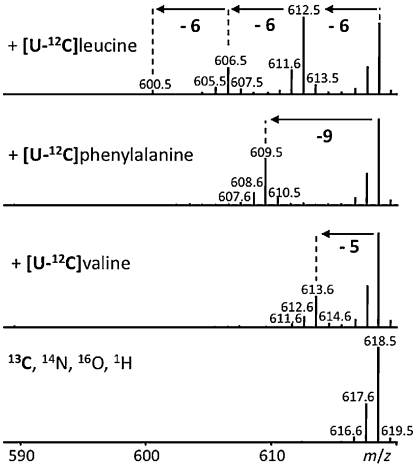


Figure 3. Determination of building blocks of **1** derived from growth of strain TT01 in fully labeled ^{13}C medium without (bottom) or with the addition of different L-^{12}C amino acids (natural abundance).

be cyclo(VLFLL) and named GameXPeptide A (**1**). This approach additionally facilitates the differentiation of isobar building blocks as leucine, isoleucine, and *N*-methyl valine, which are often found in peptides from entomopathogenic bacteria. To demonstrate this, the method was applied to entolysin A (**5**) from *Pseudomonas entomophila*, for which previous structure-elucidation efforts could not differentiate between leucine and isoleucine.[13] From labeling experiments with deuterated [$^2H_9$]leucine followed by tandem MS analysis, the correct peptide sequence was determined (Figure 4). This could have been done for the entolysin isotopologue that exhibited the incorporation of all four expected deuterated leucine residues. However, since the fully labeled isotopologue is usually present in minute amounts, the positional analysis was performed with the isotopologue carrying only one labeled leucine. If the random incorporation of one [$^2H_9$]leucine into one of the four possible positions of each fragmented molecule is considered, the ratio of individual sequence ions (in this case, b ions) with and without the incorporated [$^2H_9$]leucine residue would change in a predictable manner. As one isoleucine (Ile) and four Leu residues are present in **5** as deduced from the building-block analysis (data not shown), the following expectations for ratios of labeled and unlabeled b ion isotopologues can

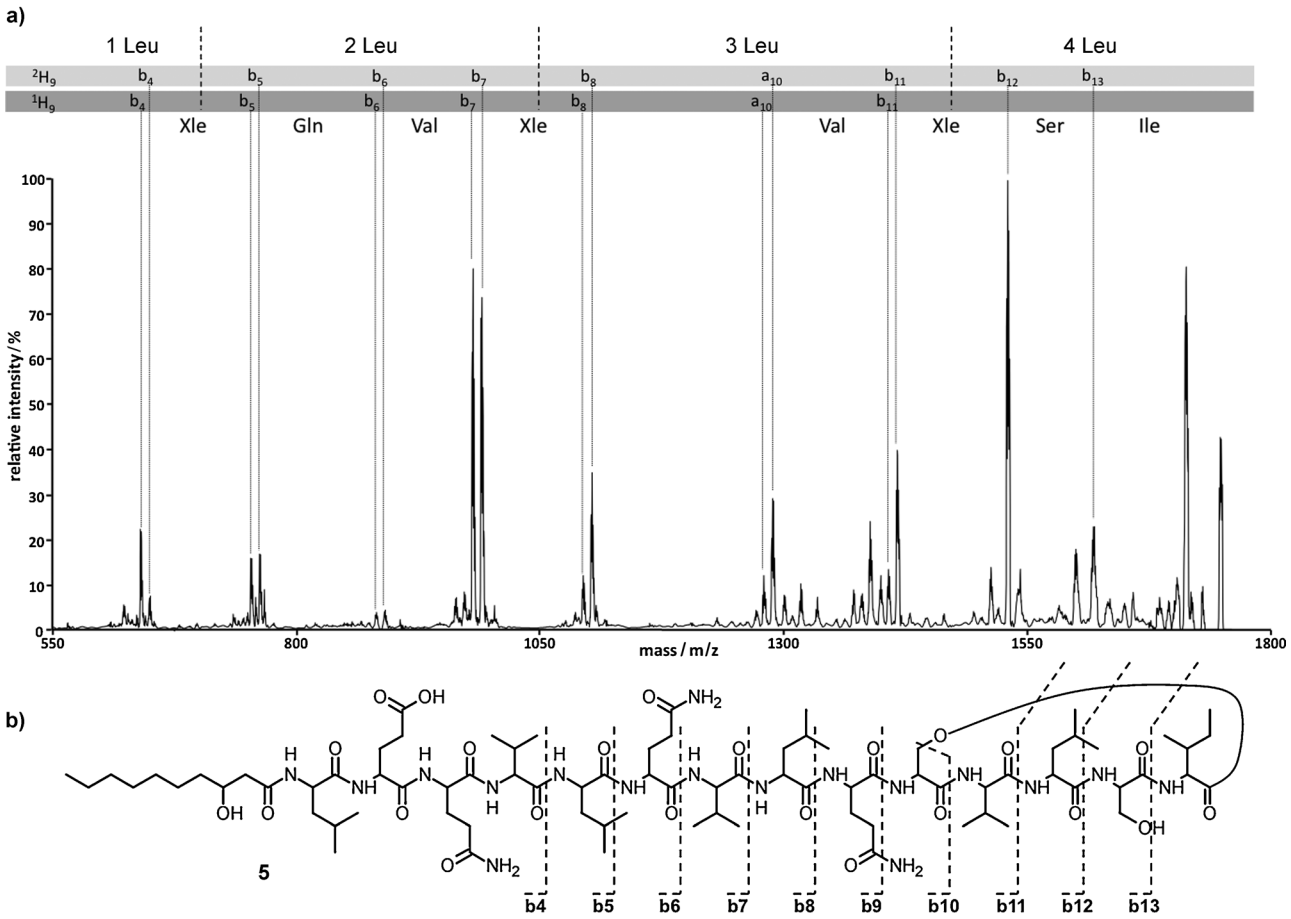


Figure 4. Structure of entolysin A (**5**) and determination of the position of leucine and isoleucin residues in **5**. Depicted is an MS^2 resulting from **5** carrying one [$^2H_9$]leucine; b ions are labeled with dotted lines highlighting deuterated (light gray) and nondeuterated (dark gray) isotopologues.

be made for **5** labeled with one [$^2H_9$]leucine: Every b ion that results from fragmentation at the N-terminal position of Leu should exhibit an altered ratio of the b ion isotopologues ([$^2H_9$]/[$^1H_9$]) in comparison to the b_{+1} ion isotopologues, which still contain the Leu, in which the relative amount of the unlabeled b ion ([$^1H_9$]) would increase. Therefore b ions that harbor 3 of 4, 2 of 4, and 1 of 4 Leu should exhibit ratios of unlabeled to labeled isotopologues [$^1H_9$]/[$^2H_9$] of 1:3 (b_{11}/b_8), 1:1 (b_7/b_5), and 3:1 (b_4), respectively, as was indeed the case (Figure 2). The b_{13} ion could exclusively be detected as its labeled isotopologue, thereby indicating that the C-terminal amino acid residue of entolysin is isoleucine (Figure 4).

Determination of amino acid configuration: The absolute configuration of the amino acid building blocks still needs to be determined. Peptides like **1**–**5** are usually derived from nonribosomal peptide synthetases (NRPS).[14] These multifunctional giant multienzyme complexes consist of single domains for amino acid activation: adenylation (A domain), condensation (C domain) of two amino acids as peptidyl carrier protein, or thiotemplate (T domains) for covalent binding of peptide intermediates. All domains necessary for the incorporation of a single amino acid are grouped into modules and thus the number of modules usually reflects the number of amino acids in the final natural product and vice versa. In most cases in which D-amino acids are present in the natural product, epimerization (E) domains or mixed C/E domains are present at the corresponding position in the NRPS. These E domains racemize the enzyme-bound L-amino acid or the bound L-peptide intermediate, and subsequently only the D derivative is further processed. During this process, the amino acid is converted to the corresponding enolate anion and vice versa.[15] Thus, by using an amino acid with a ^2H label at the α position, the presence of an E domain can be probed indirectly: Is the label found in the final product? Then the amino acid at this position must be L due to a missing or nonfunctional E domain. Is the label lost? Then it has been exchanged against ^1H from the culture medium by means of E-domain-catalyzed racemization, and the amino acid at this position must be D (Figure 5a) if no other factor contributes to the peptide configuration.

Unfortunately, in wild-type strains, this approach cannot be used because of the catalytic activity of transaminases in the cells, which convert amino acids into the corresponding 2-keto carboxylic acids and vice versa and that are involved in the last step in amino acid biosynthesis and the first step

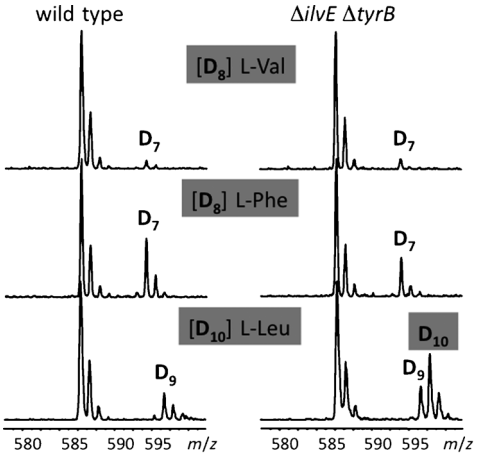


Figure 6. Labeling of GameXPeptide A (**1**) with [$^2H_8$]valine, [$^2H_{10}$]leucine, or [$^2H_8$]phenylalanine in wild type and a Δ*ilvE*Δ*tyrB* mutant of *P. luminescens*.

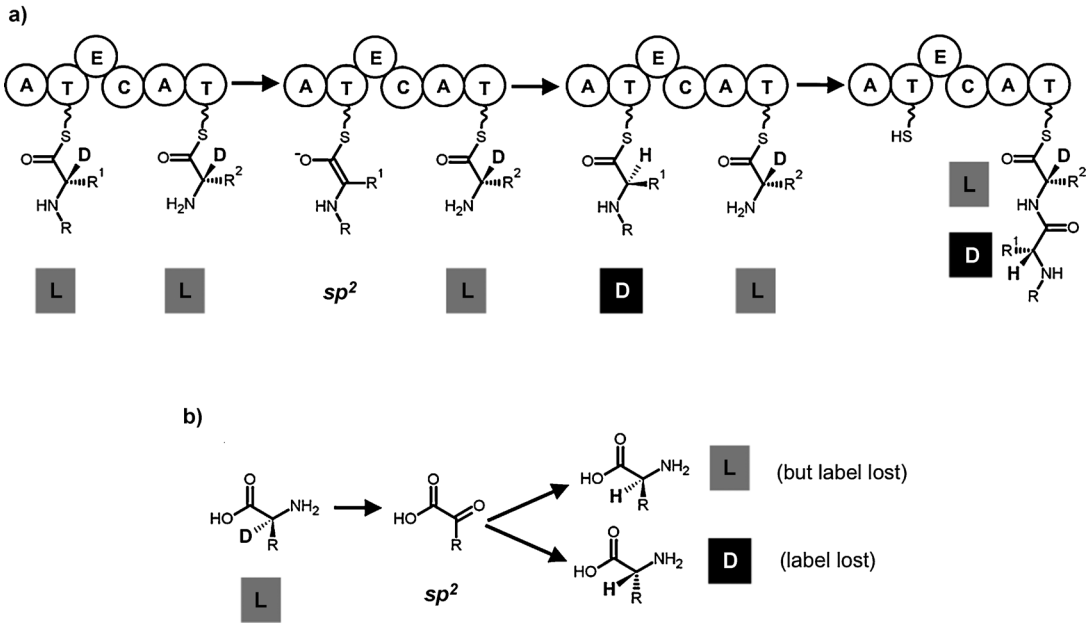


Figure 5. a) Epimerization-domain-catalyzed conversion of L- into D-amino acids based on ^2H-labeled amino acids. Adenylation (A), peptidyl carrier protein (PCP), epimerization (E), and condensation (C) domains are indicated. b) Transaminase-catalyzed loss of label from ^2H-labeled amino acids.

in amino acid degradation[16] (Figure 5b). In Gram-negative bacteria, these transaminases are encoded by *ilvE*, *tyrB*, and *aspC*, which show specificities to different sets of amino acids. These genes are highly conserved and could be identified easily in the genome of *P. luminescens*. When fully labeled [$^2H_8$]valine, [$^2H_{10}$]leucine, or [$^2H_8$]phenylalanine were added to *P. luminescens* wild-type cultures, the percentage of the fully labeled amino acids decreased to 34, 20, and 11% in favor of [$^2H_7$]valine, [$^2H_9$]leucine, or [$^2H_7$]phenylalanine, respectively (Figures S2 and S3 in the Supporting Information). As expected, a deletion of *ilvE* resulted in 95% preservation of [$^2H_{10}$]leucine after 24 h and a deletion of *tyrB* in full preservation of [$^2H_8$]phenylalanine. Unfortunately, the loss of label from deuterated valine could not be decreased further in Δ*ilvE*, Δ*tyrB*, or Δ*ilvE*Δ*tyrB* mutants, and we have not been able to construct an *ilvE*/*tyrB*/*aspC* mutant.

When we tested the incorporation of [$^2H_8$]valine and [$^2H_8$]phenylalanine into compound **1** in Δ*ilvE* or Δ*ilvE*Δ*tyrB* mutants, we could only observe isotopes of **1** that showed the loss of one deuterium for valine and phenylalanine (*m*/*z* 593.4) as expected for the incorporation of [$^2H_7$]valine or [$^2H_7$]phenylalanine, respectively. In the case of leucine, we could observe signals from isotopes at *m*/*z* 595.4 and 596.4 in a ratio of 1:2, thereby indicating the incorporation of [$^2H_9$]leucine and [$^2H_{10}$]leucine (Figure 6). Thus, **1** is composed out of D-Val, D-Phe, D-Leu, and two L-Leu, and therefore the configuration of all amino acids has been identified. Subsequently, MS^3 fragmentation experiments of the isotope that was carrying one [$^2H_9$]leucine and one [$^2H_{10}$]leucine (*m*/*z* 605.4) to determine the L- and D-leucine positions were performed. Fragmentation of *m*/*z* 458.4 (loss of phenylalanine; Figure 7a) showed a [$^2H_9$]leucine-derived fragment, whereas for *m*/*z* 506.5 (loss of valine; Figure 7b), *m*/*z* 492.4 (loss of leucine; Figure 7c), and *m*/*z* 483.4 (loss of [$^2H_9$]leucine; Figure 7d), only [$^2H_{10}$]leucine-derived fragments were observed. Fragmentation of *m*/*z* 482.4 (loss of [$^2H_{10}$]leucine; Figure 7e) showed the expected loss of valine. For comparison, unlabeled **1** was also fragmented similarly (Figure 7f–h). Thus, from the different connections of the five different building blocks as determined in Figure 7a–e and summarized in Figure 7i, the absolute configuration of **1** was determined to be cyclo(vLflL). We confirmed the structure of **1** by solid-phase synthesis and could show that synthetic **1** has indeed the same retention time (Figure S4 in the Supporting Information) and identical NMR spectroscopic data (Table S4 in the Supporting Information) as the natural compound. Furthermore, a small amount of **1** was isolated and subjected to Marfey's advanced analysis, which also indicated the correct configura-

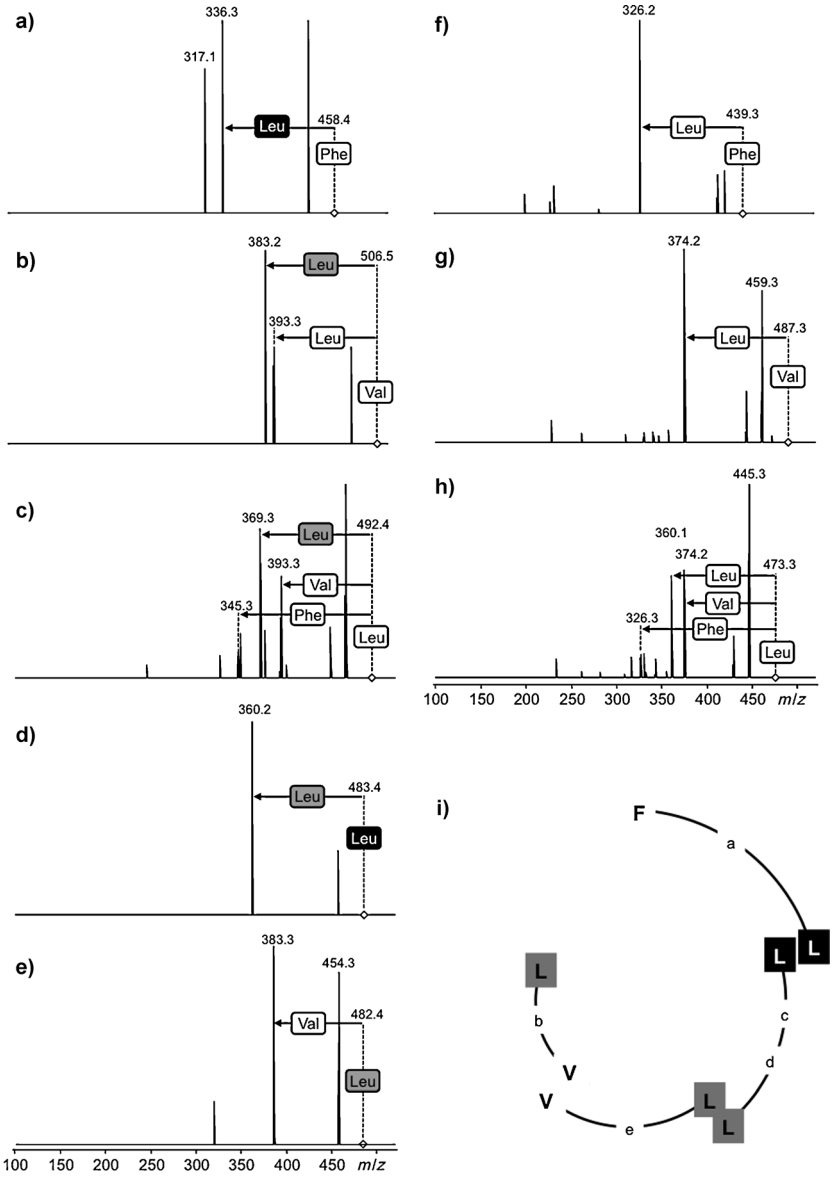
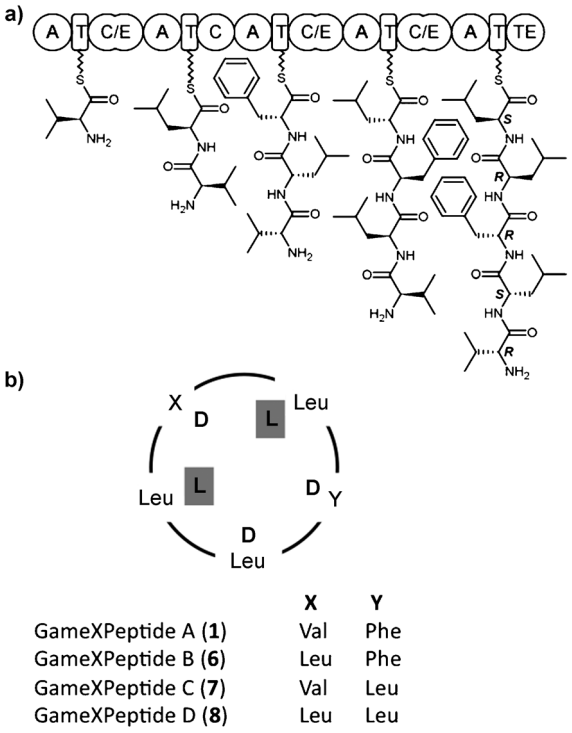


Figure 7. MS^3 spectra of **1** isolated from *P. luminescens* Δ*ilvE*Δ*tyrB* a–e) labeled with [$^2H_{10}$]leucine and [$^2H_9$]leucine (*m*/*z* 605.5) and f–h) nonlabeled **1** as control. The parent MS^2 ions are labeled with a diamond. Amino acid labels in white boxes refer to unlabeled amino acids; [$^2H_{10}$]leucine is shown in gray boxes and [$^2H_9$]leucine in black boxes. i) The assignment of the stereochemistry as concluded from the spectra in (a)–(e).

tion of the amino acids (Figure S5 in the Supporting Information). In addition, we identified the gene encoding of the GameXPeptide NRP*S* *gxpS* by plasmid integration and the expected loss of production of **1** when compared to the wild type (Figure S4 in the Supporting Information). A detailed analysis of GxpS allowed the prediction of all required domains for GameXPpetide biosynthesis including three C/E domains at the positions determined by the labeling experiments (Figure 8a, Figure S6 in the Supporting Information).



	X	Y
GameXPeptide A (**1**)	Val	Phe
GameXPeptide B (**6**)	Leu	Phe
GameXPeptide C (**7**)	Val	Leu
GameXPeptide D (**8**)	Leu	Leu

Figure 8. a) Domain organization of GxpS responsible for the biosynthesis of **1** (enzyme-bound intermediates are shown) and **6**–**8**. b) Structures of GameXPeptides A–D (**1** and **6**–**8**).

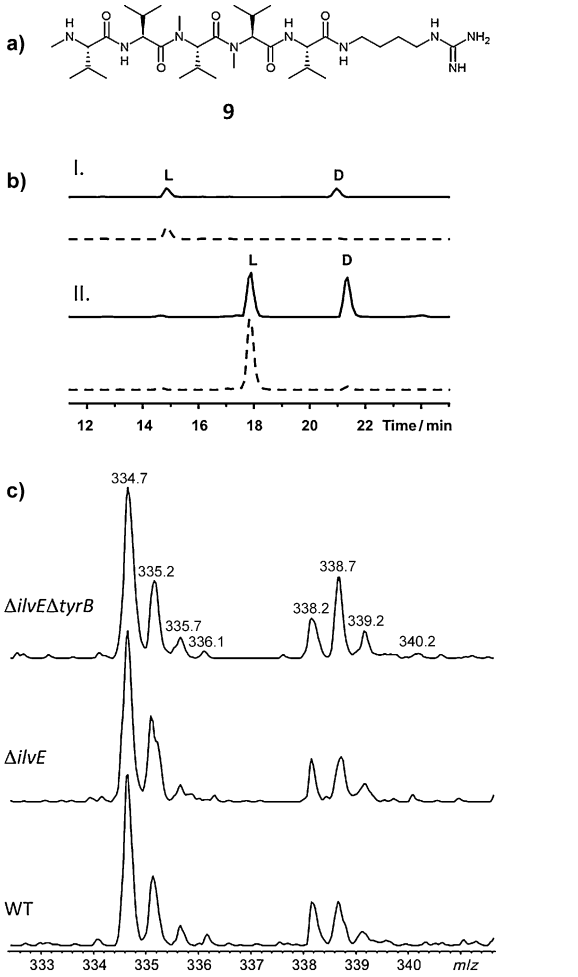


Figure 9. a) Structure results from Marfey's advanced analysis showing the D,L-FDLA derivatives (continuous lines) and the L-FDLA derivatives (dashed lines) of *N*-methylvaline (*m*/*z* 426.2, **I**), and valine (*m*/*z* 412.2, **II**) in b) the positive mode, and results from labeling experiments with c) [$^2H_8$]valine of mevalagmapeptide (**9**) produced by *P. luminescence* TT01. Mevalagmapeptide **9** (*m*/*z* 334.7, $C_{33}H_{67}O_5N_9$) is detected as a double-charged ion, thus incorporation of [$^2H_7$]valine and [$^2H_8$]valine led to *m*/*z* 338.2 (+3.5) and 338.7 (+4.0), respectively.

A detailed analysis of the *P. luminescens* extract revealed the presence of three additional derivatives of **1** named GameXPeptide B–D (**6**–**8**, Figure 8b), the structure of which was elucidated similarly (Table S3 in the Supporting Information) and which all were lost in the *gxpS* mutant. Their structural variability results from the exchange of Val[1] and/or Phe[3] against Leu.

As the presence of [$^2H_7$]valine in **1** could not fully differentiate between the incorporation of [$^2H_8$]valine followed by epimerization-domain-catalyzed loss of one deuterium and the incorporation of [$^2H_7$]valine that resulted from residual transamination activity in the bacterial culture, we analyzed another novel peptide named mevalagmapeptide (**9**) from strain TT01 (Figure 9a) that is composed of L-valine exclusively as determined by Marfey's advanced analysis (Figure 9b) after the isolation of **9**. Here, the presence of [$^2H_7$]valine but also of [$^2H_8$]valine was detected almost in a 1:1 ratio in the wild type (Figure 9c, bottom line), which is in contrast to the exclusive [$^2H_7$]valine-derived isotopomer observed for **1** (Figure 6), thus confirming the L configuration and indicating that even a presence of 66% of fully deuterated valine (see Figure S3 in the Supporting Information) is sufficient for a stereochemical analysis. Similarly, the configuration of leucine in **2**, **3**, and xenocoumacin-1,[17] and phenylalanine in **3** was determined in transaminase mutants of *X. nematophila* HGB081. Comparison of the labeling results from the *ilvE* or *tyrB* mutants with the wild type allowed the differentiation between D- and L-amino acids and confirmed that all analyzed amino acids are L in these compounds (Figures S7 and S8 in the Supporting Information) as described previously.[12,18]

As the three transaminases are highly conserved at the protein but also at the DNA level in *Xenorhabdus* and *Photorhabdus* and other bacteria (data not shown), their corresponding genes can probably be disrupted or deleted even

by means of homologous recombination based on gene sequences of closely related strains, thus enabling the determination of absolute configuration in similar peptides also in strains in which no genome sequence is available.

Conclusion

In summary, we have shown that a combination of labeling experiments of wild-type and transaminase mutants with MS analysis allows 1) the correct determination of sum formulas; 2) the rapid and reliable identification of building blocks, and especially amino acids; and 3) the determination of the absolute configurations of these amino acids. Our approach is especially useful for peptides because they can often be synthesized more easily than they are isolated. In particular, when several different derivatives are produced by microorganisms, with some of them present only in trace amounts, our approach might enable the structure elucidation of the derivatives, which can then be synthesized and subsequently tested for their bioactivity following their synthesis.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) and the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 223328. Work in the author's laboratory is additionally supported by the Bundesministerium für Bildung und Forschung (BMBF) and the Frankfurt Initiative for Microbial Sciences (FIMS).

[1] S. Donadio, S. Maffioli, P. Monciardini, M. Sosio, D. Jabes, *J. Antibiot.* **2010**, *63*, 423–430.
[2] V. An Thieu, D. Kirsch, T. Flad, C. Müller, B. Spengler, *Angew. Chem.* **2006**, *118*, 918; *Angew. Chem. Int. Ed.* **2006**, *45*, 904.
[3] B. Spengler, *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 703–714.
[4] P. Giavalisco, K. Kohl, J. Hummel, B. Seiwert, L. Willmitzer, *Anal. Chem.* **2009**, *81*, 6546–6551.
[5] P. Giavalisco, Y. Li, A. Matthes, A. Eckhardt, H. M. Hubberten, H. Hesse, S. Segu, J. Hummel, K. Kohl, L. Willmitzer, *Plant J.* **2011**, *68*, 364–376.
[6] P. Giavalisco, J. Hummel, J. Lisec, A. C. Inostroza, G. Catchpole, L. Willmitzer, *Anal. Chem.* **2008**, *80*, 9417–9425.
[7] L. Feldberg, I. Venger, S. Malitsky, I. Rogachev, A. Aharoni, *Anal. Chem.* **2009**, *81*, 9257–9266.
[8] R. Baran, B. P. Bowen, N. J. Bouskill, E. L. Brodie, S. M. Yannone, T. R. Northen, *Anal. Chem.* **2010**, *82*, 9034–9042.
[9] H. B. Bode, *Curr. Opin. Chem. Biol.* **2009**, *13*, 224–230.
[10] N. Stoll, E. Schmidt, K. Thurow, *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1692–1699.
[11] R. P. Rodgers, E. N. Blumer, C. L. Hendrickson, A. G. Marshall, *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 835–840.
[12] G. Lang, T. Kalvelage, A. Peters, J. Wiese, J. F. Imhoff, *J. Nat. Prod.* **2008**, *71*, 1074–1077.
[13] I. Vallet-Gely, A. Novikov, L. Augusto, P. Liehl, G. Bolbach, M. Pechy-Tarr, P. Cosson, C. Keel, M. Caroff, B. Lemaitre, *Appl. Environ. Microbiol.* **2010**, *76*, 910–921.
[14] R. Finking, M. A. Marahiel, *Annu. Rev. Microbiol.* **2004**, *58*, 453–488.
[15] T. Stachelhaus, C. T. Walsh, *Biochemistry* **2000**, *39*, 5775–5787.
[16] G. Michal, *Biochemical Pathways*, Spektrum Akad. Verlag, Heidelberg **1999**.
[17] D. Reimer, E. Luxenburger, A. O. Brachmann, H. B. Bode, *ChemBioChem* **2009**, *10*, 1997–2001.
[18] D. Reimer, K. M. Pos, M. Thines, P. Grün, H. B. Bode, *Nat. Chem. Biol.* **2011**, *7*, 888–890.

Received: November 4, 2011
Published online: January 20, 2012

