Natural Products

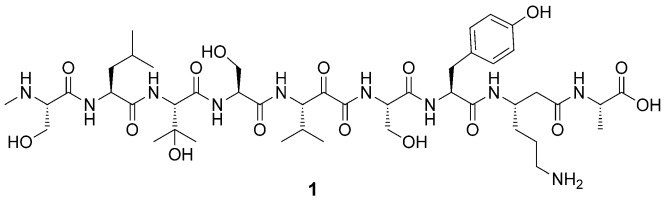
DOI: 10.1002/anie.201106305

Myxoprincomide: A Natural Product from *Myxococcus xanthus* Discovered by Comprehensive Analysis of the Secondary Metabolome**

*Niña Socorro Cortina, Daniel Krug, Alberto Plaza, Ole Revermann, and Rolf Müller**

Bacteria are important sources of therapeutically relevant natural products.[1] Early established bacterial producers include the streptomycetes, pseudomonads, and bacilli, but the list has recently expanded to include further sources, such as the myxobacteria.[2–5] Access to whole genome sequences has indicated that among bacteria producing natural products through the action of polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) the number of known compounds is remarkably lower than the genetic capacity of the organism for secondary-metabolite biosynthesis.[1,6] Thus, identification of new compound classes and their assignment to biosynthetic gene clusters is a crucial step in the discovery of novel natural products.

Here, we report the discovery and structural elucidation of myxoprincomide (**1**), a novel NRPS/PKS natural product from *Myxococcus xanthus* DK1622, by combining methods of targeted mutagenesis, liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS), and statistical data evaluation (Figure 1). Myxoprincomide (**1**), a linear peptide bearing some unusual residues, is produced by the NRPS/PKS biosynthetic machinery encoded by the *mxp* (MXAN_3779) gene locus. Moreover, by correlating additional very-low-abundance natural products to biosynthetic pathways in DK1622 we demonstrate how a comprehensive "metabolome-mining" approach can complement genome-mining strategies in the discovery of secondary metabolites.



1

The myxobacterial strain DK1622, considered a model organism for the study of myxobacterial social motility and multicellular differentiation, was not recognized as a producer of secondary metabolites until its genome was sequenced.[7] However, interest in resolving its secondary metabolome has increased since bioinformatic analysis revealed it to contain 18 biosynthetic gene clusters encoding NRPS, PKS, and NRPS/PKS hybrid systems. To date only five compound classes derived from NRPS and PKS biosynthetic machineries have been characterized and correlated to their gene clusters in DK1622 (see Figure 1 in the Supporting Information).[8–13] Intriguingly, evaluation at the transcriptomic and proteomic levels asserts that the remaining 13 unassigned pathways are active under standard conditions for the cultivation of DK1622.[14,15] We reasoned that the low abundance of the corresponding compounds may have previously prevented their detection, and thus set out to find these compounds by utilizing advanced analytical methods.

The available DK1622 genome sequence facilitated the construction of a targeted mutant library including knockouts of every secondary-metabolite biosynthetic gene cluster (Table 1 in the Supporting Information).[16] As the process of finding the possibly subtle differences between wild-type and mutant secondary-metabolite profiles by manual comparison of LC-MS data is frequently tedious, error-prone, and low in sensitivity, we sought to implement statistical tools in order to expedite our LC-MS data evaluation. In preparation for the comprehensive statistical analysis, mutant and wild-type strains were grown in small-scale fermentation in quadruplicate, replicate extracts were analyzed by LC-HRMS, and data were pretreated by using a compound-finding algorithm, resulting in the definition of >1000 molecular features per sample.[17–19] In order to identify molecular features specifically missing in culture extracts from DK1622 mutant strains, we applied principal-component analysis (PCA) to the preprocessed LC-MS datasets

[*] M. Sc. N. S. Cortina,[+] Dr. D. Krug,[+] Dr. A. Plaza, Dipl.-Chem. O. Revermann, Prof. Dr. R. Müller
Abteilung Mikrobielle Naturstoffe, Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS), Helmholtz-Zentrum für Infektionsforschung
and
Institut für Pharmazeutische Biotechnologie, Universität des Saarlandes
Campus, Gebäude C2.3, 66123 Saarbrücken (Deutschland)
E-mail: rom@helmholtz-hzi.de
Homepage: http://www.helmholtz-hzi.de/hips

[+] These authors contributed equally to this work.

[**] We thank Gregory Velicer of Indiana University for providing strain *Myxococcus xanthus* A2 and Charles Moore for critically reading this manuscript. We thank Thomas Hoffmann and Eva Luxenburger for help with the HPLC-MS analyses, Wolfgang Kessler and Stephan Hüttel for the large-scale fermentation, and Gabriela Zurek and Aiko Barsch (Bruker Daltonik) for their advice. N.S.C. thanks the Deutscher Akademischer Austauschdienst for a fellowship. Research in the R.M. group was funded by the Bundesministerium für Bildung und Forschung and the Deutsche Forschungsgemeinschaft.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201106305.

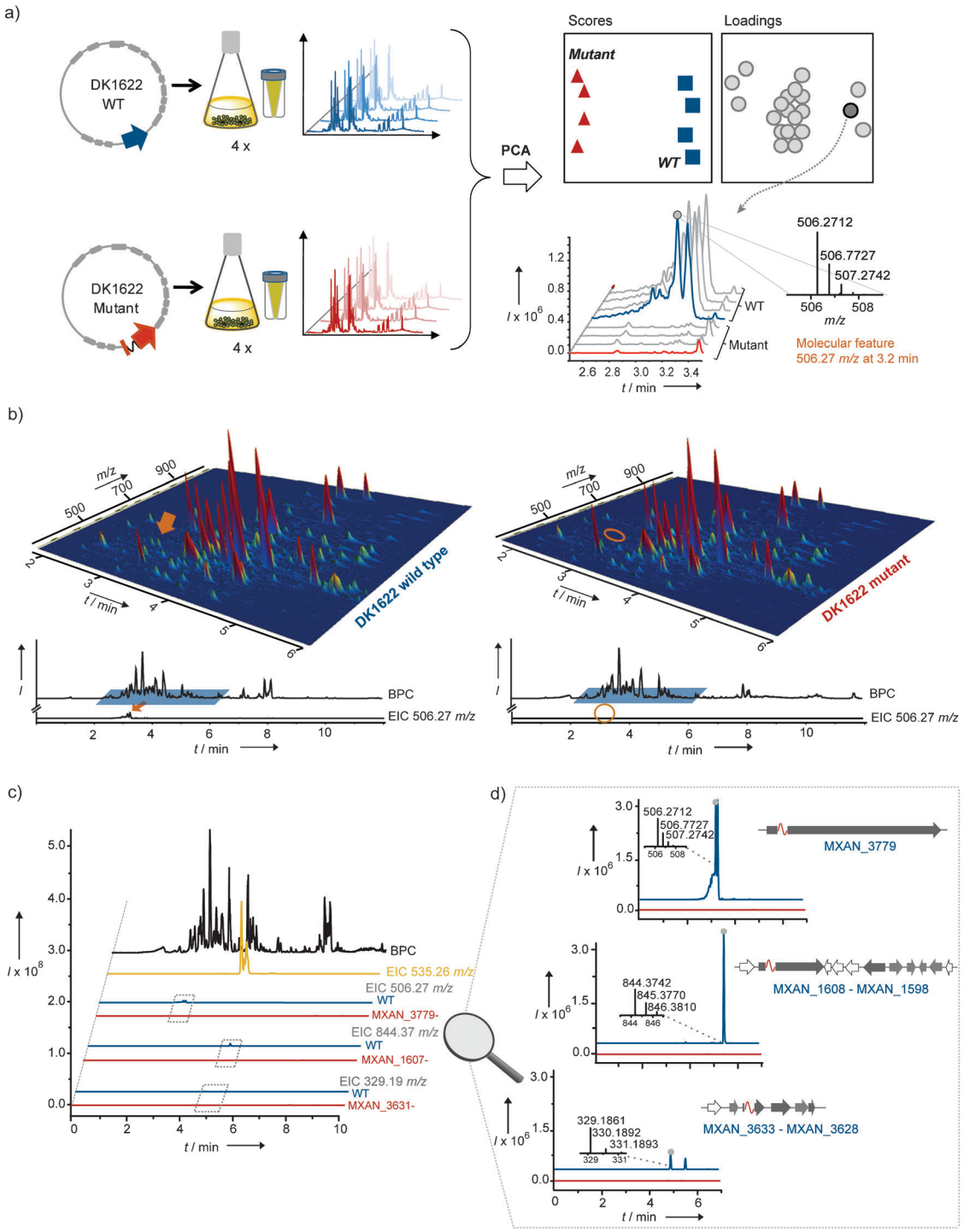


Figure 1. The gene-to-compound approach by comprehensive secondary-metabolome mining. a) Target gene inactivation, replicate cultivation of WT and mutant strains, LC-HRMS analysis and statistical evaluation by PCA, resulting in a PCA scores and loadings plot. Each dot on the loadings plot corresponds to a molecular feature represented by an m/z–t_r pair. As an example, the molecular feature m/z 506.27 eluting at $t_r = 3.2$ min is highlighted. b) Comparison of DK1622 WT and mutant samples (chromatogram and partial 3D representation), showing the position of molecular feature m/z 506.27 (orange arrow, orange circle) within the original LC-MS data. c) Assignment of three molecular features to gene clusters as revealed by comparative statistical analysis. EICs of c506 ($[M+2H]^{2+}$ m/z 506.27, **1**), c844 ($[M+H]^+$ m/z 844.37), and c329 ($[M+H]^+$ m/z 329.19) from the WT (blue) and mutant (red), the BPC of the crude extract, and the EIC of DKxanthene-534 ($[M+H]^+$ m/z 535.26) demonstrate the disparity of intensities between the more abundant known *M. xanthus* compounds and the newly assigned compounds. d) Magnified EICs of c506, c844, and c329. Chromatographic profiles are displayed alongside mass spectra and inactivated biosynthetic gene clusters (red squiggly line: location of the targeted gene inactivation). BPC: base peak chromatogram, EIC: extracted ion chromatogram, PCA: principal component analysis, WT: wild-type, t_r = retention time.

(Figure 1). PCA reduces the dimensionality of a multivariate dataset, while retaining the relevant information in terms of variance. Thereby PCA can act as a statistical filter, highlighting groups of observations, trends, and outliers within a complex sample set.[20] Statistical analysis was performed pairwise using the wild-type sample set and the individual mutant sample sets, and the molecular features determined to be stringently absent from the sample replicates of all mutant strains were included in a shortlist of candidate compounds.

Using this approach, three new secondary metabolites—c506, c844, and c329—were revealed and assigned to their biosynthetic gene clusters in *Myxococcus xanthus* DK1622 (Figure 1c). To obtain initial insights regarding the presumed structures of the novel compounds, the DNA sequences of the corresponding pathways were analyzed. Compound c506 is a product of the hybrid NRPS/PKS enzyme encoded by the gene MXAN_3779 (42825 base pairs) which encodes the largest single and continuous hybrid enzyme (predicted molecular weight: 1.55 MDa) in DK1622 that is presumed to synthesize and release an entire structural backbone. Annotation revealed the presence of eleven NRPS modules, two PKS modules, and a terminal thioesterase (TE) domain. The detailed bioinformatic analysis of the biosynthetic gene clusters underlying the formation of compounds c506, c844, and c329 can be found in the Supporting Information (Tables 3–8 and Figure 5).

After the assigment of these three novel classes of compounds to biosynthetic gene clusters we chose the candidate compound c506, which was named myxoprincomide (**1**: HR-ESIMS: *m/z* 506.2712, $[M+2H]^{2+}$; calcd for $C_{45}H_{76}N_{10}O_{16}$: *m/z* 506.2721, $[M+2H]^{2+}$), as a target for isolation and structure elucidation. Difficulties to predict its scaffold based on genetic information suggested that this novel secondary metabolite would exhibit uncommon structural features. LC-MS analysis indicated that myxoprincomide was present in rather low amounts in crude methanol extracts of DK1622/XAD cultivations in comparison to the levels of secondary metabolites previously reported from this strain (Figure 1c); thus we were presented with the challenge to obtain sufficient amounts of purified compound for structural elucidation. To address this problem, we investigated our collection of 98 wild-type *M. xanthus* isolates used in our pilot study on the intraspecies diversity of *M. xanthus* secondary metabolites, and by using semiquantitative analysis of metabolomic data we identified *M. xanthus* A2 as a better producer of **1** (Figure 2a).[21] Subsequently, strain A2 was genetically engineered to artificially overexpress its MXAN_3799 homologue by insertion of the T7A1 promoter upstream of the gene, generating the optimized strain A2.c506 that produced **1** at levels 30 times higher than DK1622 (Figure 2b, Figure 6 in the Supporting Information). A MeOH extract from a 100 L fermentation of strain A2.c506 was fractionated by sephadex and HPLC, yielding 1.2 mg of myxoprincomide (**1**). Its structure was determined using extensive 2D NMR specrotroscopy and ESI-MS/MS techniques (see the Supporting Information). Myxoprincomide (**1**) is an unusual linear peptide containing nine amino acid residues, including the atypical 2-oxo-β-leucine, γ-OH-valine, and β-lysine residues.

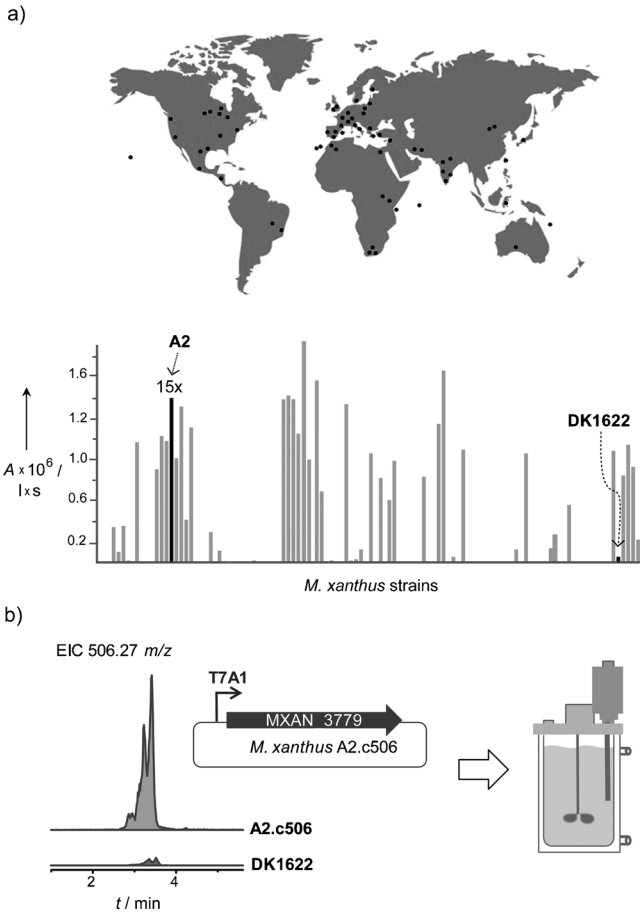


Figure 2. a) Selection of a better producer of **1** from a worldwide collection of 98 *M. xanthus* strains based on LC-MS quantitation (*A*, peak area). b) Optimization of production by promotor insertion in *M. xanthus* A2 prior to upscaling for compound isolation.

Based on the structure and the results of feeding experiments, we are able to propose a biosynthetic model for the formation of myxoprincomide by the giant NRPS/PKS hybrid assembly line Mxp, encoded by MXAN_3779 (Figure 3a, Figure 9 in the Supporting Information). Because the condensation (C) domain of module 1 apparently lacks the conserved motifs required for the activity of C domains, biosynthesis is proposed to begin with the incorporation and subsequent N-methylation of L-serine. Chain extension continues with the incorporation of L-leucine, γ-OH-valine, L-serine, and L-valine by Mxp modules 2–5. PKS module 6 presumably facilitates the elongation by extension with malonyl-CoA. Next, an unusual oxidation at the position α to the thioester function might take place, most likely catalyzed by the oxidation domain in Mxp module 7 (Figure 17 in the Supporting Information). Tautomerization could result in the formation of an α-keto functional group which is nucleophilically attacked by the amino group of the downstream L-serine precursor. Feeding with ^{13}C-labeled acetate provides evidence that only the carbon at position 2 is maintained in myxoprincomide (Table 10 in the Supporting Information). Thus, we hypothesize that the elimination of a moiety containing only C1 occurs by a yet to be understood mechanism during myxoprincomide assembly. Modules 9–11

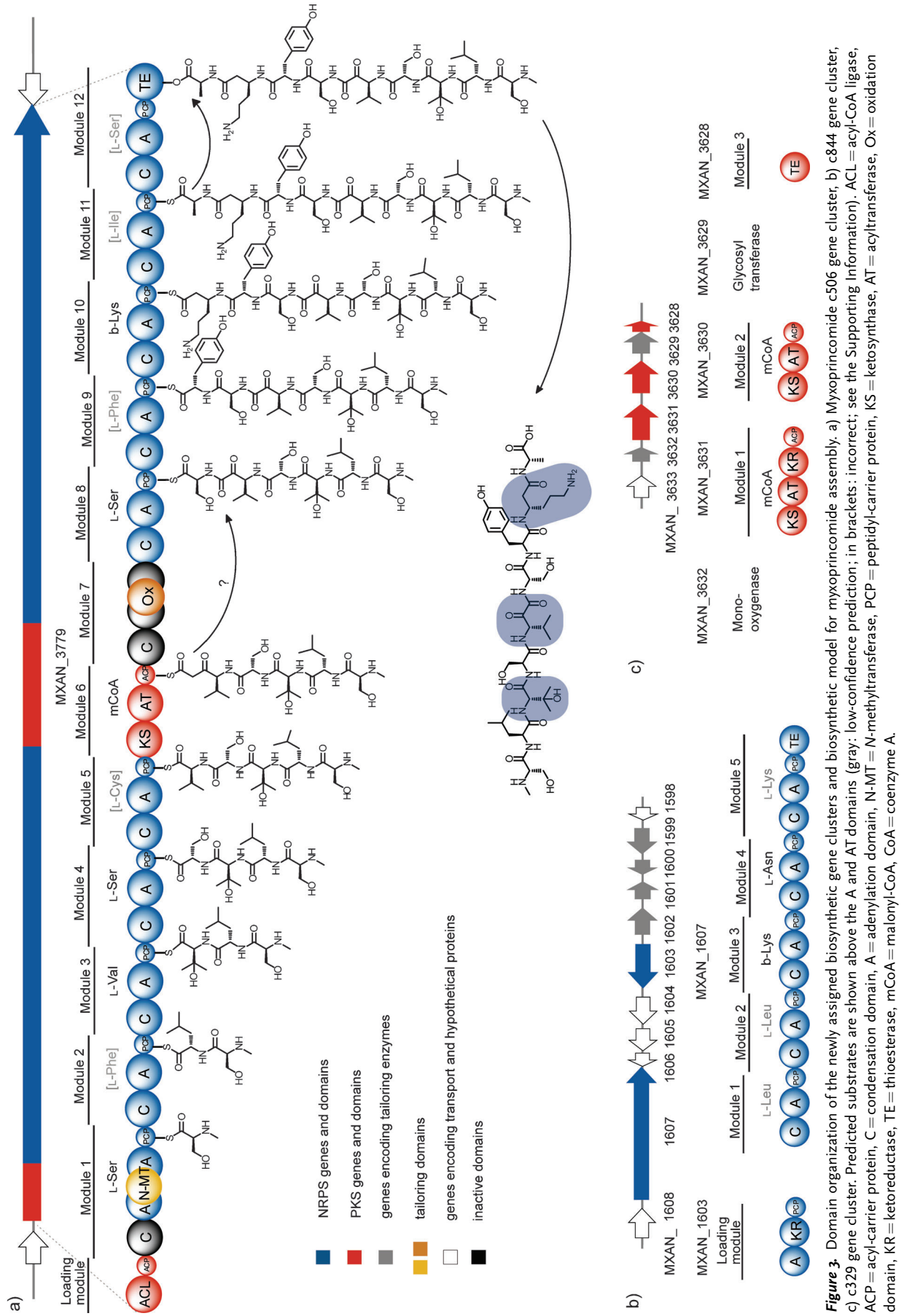


Figure 3. Domain organization of the newly assigned biosynthetic gene clusters and biosynthetic model for myxoprincomide assembly. a) Myxoprincomide c506 gene cluster, b) c844 gene cluster, c) c329 gene cluster. Predicted substrates are shown above the A and AT domains (gray: low-confidence prediction; in brackets: incorrect; see the Supporting Information). ACL = acyl-CoA ligase, ACP = acyl-carrier protein, C = condensation domain, A = adenylation domain, N-MT = *N*-methyltransferase, PCP = peptidyl-carrier protein, KS = ketosynthase, AT = acyltransferase, Ox = oxidation domain, KR = ketoreductase, TE = thioesterase, mCoA = malonyl-CoA, CoA = coenzyme A.

catalyze the chain extension with the incorporation of L-tyrosine, β-lysine, and L-alanine. The peptide chain is then passed on to the TE domain, apparently skipping the final NRPS module, and is released as a linear chain. The discovery and characterization of myxoprincomide and its correlation to the Mxp pathway now sets the stage for the investigation of the biochemistry underlying its unusual stuctural features.

Since the start of the postgenomic era, the targeted inactivation of secondary-metabolite biosynthetic gene clusters in combination with comparative analysis has frequently proven successful to establish the link between a specific biosynthetic pathway and its corresponding small-molecule product.[22,23] However, the number of discovered compounds often falls short of a strain's genetic capability for the production of secondary metabolites—a discrepancy which may be at least in part attributed to the fact that the "knockout and differential profiling" approach commonly relies on readily accessible analytical readouts such as appreciable UV/Vis absorption or distinct MS signal intensity in order to highlight the compound of interest. In this regard, most compound classes reported from DK1622 to date did not present a challenge, owing to their amenable analytical properties and because the respective biosynthetic pathways have been (partially or in full) previously characterized from other myxobacteria, thus greatly facilitating their identification in DK1622. In contrast, detection of the three new metabolites in this study and their correlation to biosynthetic pathways by our gene knockout/secondary-metabolome-mining strategy was more challenging because of the lack of prominent analytical signals (i.e. UV/Vis absorption and MS intensity), coinciding with their low abundance, and therefore required a more sophisticated approach.

The particular strength of our approach is that there are minimal requirements regarding the compounds to be discovered, that is, their compatibility with the chosen analytical setup. By keeping the first-pass analysis as unbiased as possible, it is the role of comprehensive statistical treatment to deliver an initial readout, thereby creating an entry point into the compound identification workflow.[24] As soon as candidate compounds are defined, follow-up techniques such as feeding studies and analysis of tandem-MS data can be readily applied (see the Supporting Information).[25] It might appear as an appealing alternative to mine the secondary metabolomes of promoter insertion vs. knockout pairs instead of wild-type vs. knockout pairs by comparative LC-MS analysis, but it should be noted that the intricate operon organizations of biosynthetic gene clusters as frequently found in bacterial genomes would make this undertaking generally difficult.[26] In fact, we attribute our inability to apply certain published genome-mining techniques to the unassigned DK1622 biosynthetic gene clusters at least in part to the observation that especially myxobacterial biosynthetic machineries most frequently do not follow textbook models and thus do not allow prediction of the compound's structure from analysis of the DNA sequence alone.[6]

It is worthwhile to note that the combined metabolome mining/genome mining strategy used here has the capability to reveal eight NRPS/PKS-derived compound classes from DK1622 (the three novel products—myxoprincomide, c844, and c329—plus the five known compound families (data not shown)), thus covering approximately half of the strains' presumed NRPS/PKS-related secondary metabolome by the use of only one single analytical setup. However, the expression of additional biosynthetic pathways under standard laboratory conditions is underpinned by transcriptomics and proteomics studies.[14,15] This apparent discrepancy may be in part due to the fact that the products of these remaining pathways are produced at early stages but degraded later during the life cycle of the bacteria, thus requiring analysis of samples from multiple timepoints during cultivation.[27] Furthermore, we anticipate that statistical evaluation of LC-MS data acquired using alternative cultivation parameters, workup procedures, and analytical setups (LC separation conditions, MS ionization modes) could yield additional candidates for novel compounds.

In conclusion, we have shown that the combination of gene-inactivation studies with a downstream holistic comparison of secondary metabolomes lead to the complete characterization of myxoprincomide and the discovery of two additional new classes of secondary metabolites from DK1622. Moreover, we have demonstrated how the combined use of whole-genome information, genetic manipulation, and species-wide metabolomic data helped to establish an optimized producer strain for myxoprincomide, thus enabling us to uncover this previously "hidden" natural product from the secondary metabolome of *M. xanthus*. We believe that the comprehensive metabolome-mining strategy presented here has a high potential to facilitate the discovery of additional secondary metabolites from myxobacteria and other microbial sources of natural products, in particular in cases of low-abundance compounds.

Experimental Section

The cultivation and genetic manipulation of DK1622 have been reported previously. Full experimental details, including the construction and verification of DK1622 knockout and overproducer mutants, bioinformatic analysis, LC-HRMS analysis conditions, and statistical treatment of data, as well as myxoprincomide production optimization, isolation, and structural elucidation can be found in the Supporting Information.

Received: September 6, 2011
Revised: October 17, 2011
Published online: December 7, 2011

Keywords: genome mining · mass spectrometry · metabolomics · myxobacteria · natural products

[1] C. T. Walsh, M. A. Fischbach, *J. Am. Chem. Soc.* **2010**, *132*, 2469 – 2493.
[2] B. S. Deshpande, S. S. Ambedkar, J. G. Shewale, *Enzyme Microb. Technol.* **1988**, *10*, 455 – 473.
[3] H. Budzikiewicz, *FEMS Microbiol. Rev.* **1993**, *10*, 209 – 228.
[4] E. Katz, A. L. Demain, *Bacteriol. Rev.* **1977**, *41*, 449 – 474.
[5] K. J. Weissman, R. Müller, *Nat. Prod. Rep.* **2010**, *27*, 1276 – 1295.
[6] S. C. Wenzel, R. Müller, *Nat. Prod. Rep.* **2009**, *26*, 1385 – 1407.
[7] *Myxobacteria: Multicellularity and differentiation* (Ed.: D. Whitworth), ASM Press, Chicago, **2007**.

[8] H. B. Bode, R. Müller, *Angew. Chem.* **2005**, *117*, 6988 – 7007; *Angew. Chem. Int. Ed.* **2005**, *44*, 6828 – 6846.
[9] H. B. Bode, P. Meiser, T. Klefisch, N. S. Cortina, D. Krug, A. Göhring, G. Schwär, T. Mahmud, Y. A. Elnakady, R. Müller, *ChemBioChem* **2007**, *8*, 2139 – 2144.
[10] C. M. Ronning, W. C. Nierman in *Myxobacteria: Multicellularity and differentiation* (Ed.: D. Whitworth), ASM Press, Chicago, **2007**, pp. 285 – 298.
[11] S. C. Wenzel, P. Meiser, T. Binz, T. Mahmud, R. Müller, *Angew. Chem.* **2006**, *118*, 2354 – 2360; *Angew. Chem. Int. Ed.* **2006**, *45*, 2296 – 2301.
[12] V. Simunovic, J. Zapp, S. Rachid, D. Krug, P. Meiser, R. Müller, *ChemBioChem* **2006**, *7*, 1206 – 1220.
[13] P. Meiser, K. J. Weissman, H. B. Bode, D. Krug, J. S. Dickschat, A. Sandmann, R. Müller, *Chem. Biol.* **2008**, *15*, 771 – 781.
[14] H. B. Bode, M. W. Ring, G. Schwär, M. O. Altmeyer, C. Kegler, I. R. Jose, M. Singer, R. Müller, *ChemBioChem* **2009**, *10*, 128 – 140.
[15] C. Schley, M. O. Altmeyer, R. Swart, R. Müller, C. G. Huber, *J. Proteome Res.* **2006**, *5*, 2760 – 2768.
[16] B. S. Goldman, W. C. Nierman, D. Kaiser, S. C. Slater, A. S. Durkin, J. Eisen, C. M. Ronning, W. B. Barbazuk, M. Blanchard, C. Field, C. Halling, G. Hinkle, O. Iartchuk, H. S. Kim, C. Mackenzie, R. Madupu, N. Miller, A. Shvartsbeyn, S. A. Sullivan, M. Vaudin, R. Wiegand, H. B. Kaplan, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15200 – 15205.
[17] D. Krug, G. Zurek, B. Schneider, R. Garcia, R. Müller, *Anal. Chim. Acta* **2008**, *624*, 97 – 106.
[18] C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, *Anal. Chem.* **2006**, *78*, 779 – 787.
[19] K. M. Aberg, R. J. O. Torgrip, J. Kolmert, I. Schuppe-Koistinen, J. Lindberg, *J. Chromatogr. A* **2008**, *1192*, 139 – 146.
[20] B. K. Lavine in *Encyclopedia of Analytical Chemistry: Applications, Theory, and Instrumentation* (Ed.: R. A. Meyers), Wiley, New York **2000**, pp. 1 – 20.
[21] D. Krug, G. Zurek, O. Revermann, M. Vos, G. J. Velicer, R. Müller, *Appl. Environ. Microbiol.* **2008**, *74*, 3058 – 3068.
[22] M. Zerikly, G. L. Challis, *ChemBioChem* **2009**, *10*, 625 – 632.
[23] H. Gross, *Appl. Microbiol. Biotechnol.* **2007**, *75*, 267 – 277.
[24] N. Vinayavekhin, A. Saghatelian, *ACS Chem. Biol.* **2009**, *4*, 617 – 623.
[25] H. Gross, V. O. Stockwell, M. D. Henkels, B. Nowak-Thompson, J. E. Loper, W. H. Gerwick, *Chem. Biol.* **2007**, *14*, 53 – 63.
[26] E. K. Schmitt, C. M. Moore, P. Krastel, F. Petersen, *Curr. Opin. Chem. Biol.* **2011**, *15*, 497 – 504.
[27] R. Müller, K. Gerth, *J. Biotechnol.* **2006**, *121*, 192 – 200.

