ACS chemical biology

Articles

pubs.acs.org/acschemicalbiology

Genetic Basis for the Biosynthesis of the Pharmaceutically Important Class of Epoxyketone Proteasome Inhibitors

Michelle Schorn,[†] Judith Zettler,[∥,⊥] Joseph P. Noel,[‡] Pieter C. Dorrestein,[§] Bradley S. Moore,*[,†,§] and Leonard Kaysser*[,†,∥,⊥]

[†]Scripps Institution of Oceanography, University of California, San Diego, California 92093, United States of America

[‡]Jack H. Skirball Center for Chemical Biology and Proteomics, Salk Institute for Biological Studies, La Jolla, California 92037, United States of America

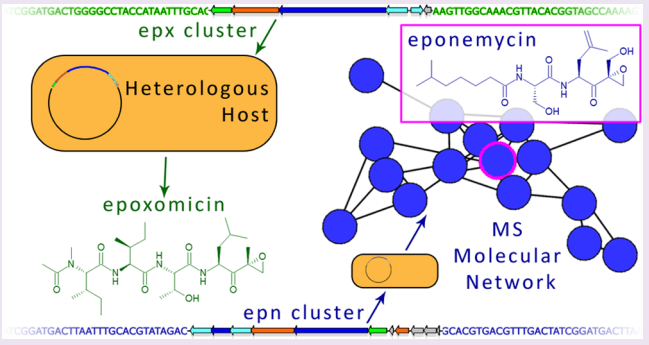
[§]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, California 92093, United States of America

[∥]Pharmaceutical Biology, Eberhard Karls University Tübingen, 72076 Tübingen, Germany

[⊥]German Center for Infection Research (DZIF), partner site Tübingen, 72076 Tübingen, Germany

Supporting Information

ABSTRACT: The epoxyketone proteasome inhibitors are an established class of therapeutic agents for the treatment of cancer. Their unique α',β'-epoxyketone pharmacophore allows binding to the catalytic β-subunits of the proteasome with extraordinary specificity. Here, we report the characterization of the first gene clusters for the biosynthesis of natural peptidyl-epoxyketones. The clusters for epoxomicin, the lead compound for the anticancer drug Kyprolis, and for eponemycin were identified in the actinobacterial producer strains ATCC 53904 and *Streptomyces hygroscopicus* ATCC 53709, respectively, using a modified protocol for Ion Torrent PGM genome sequencing. Both gene clusters code for a hybrid nonribosomal peptide synthetase/polyketide synthase multifunctional enzyme complex and homologous redox enzymes. Epoxomicin and eponemycin were heterologously produced in *Streptomyces albus* J1046 via whole pathway expression. Moreover, we employed mass spectral molecular networking for a new comparative metabolomics approach in a heterologous system and discovered a number of putative epoxyketone derivatives. With this study, we have definitively linked epoxyketone proteasome inhibitors and their biosynthesis genes for the first time in any organism, which will now allow for their detailed biochemical investigation.



The 26S proteasome is the essential enzymatic complex for nonlysosomal proteolytic degradation in eukaryotes.[1] It mediates levels of key factors in a variety of essential cellular processes that are deregulated in cancer cells and pivotal elements in carcinogenesis and tumorigenesis. Thus, the inhibition of the proteasome specifically targets heavily proliferating cells over quiescent cells.[2,3] The first proteasome inhibitor bortezomib (Figure 1, **1**) (marketed as Velcade by Millenium Pharmaceuticals) was approved by the U.S. Food and Drug Administration (FDA) in 2003. It is currently applied as a first-line treatment for multiple myeloma and mantle cell lymphoma. However, intravenous administration of the drug is associated with significant side effects. The development of proteasome inhibitors with improved properties is therefore an ongoing effort.

A number of potent proteasome inhibitors have been isolated from nature, predominantly from microorganisms.[4] The most prominent class are the peptide epoxyketones, which comprise epoxomicin (**2**),[5] eponemycin (**3**),[6] and several related compounds (**4**–**8**)[7−9] (Figure 1). All these molecules consist of a short peptidic core structure with a terminal C_3-extended leucine derivative. Compound **2** is particularly potent with IC_{50} values against the proteasome as low as 2.5 nM.[7] The compound has been used as a lead for the development of carfilzomib (**9**, Kyprolis, Onyx Pharmaceuticals), which was granted accelerated approval by the FDA in July 2012 for the treatment of refractory and relapsed multiple myeloma.[10] The drug appears to be better tolerated by patients than **1** and can therefore be applied in higher and more effective doses.[11,12] Beside their usage as anticancer drugs, epoxyketones have shown excellent activity against parasites.[13] In particular, *Plasmodium falciparum*, the deadly malaria pathogen, is vitally affected by this class of proteasome inhibitors at different stages of its life cycle.[14] Co-crystallization experiments with **2** and the

Received: September 11, 2013
Accepted: October 29, 2013
Published: October 29, 2013

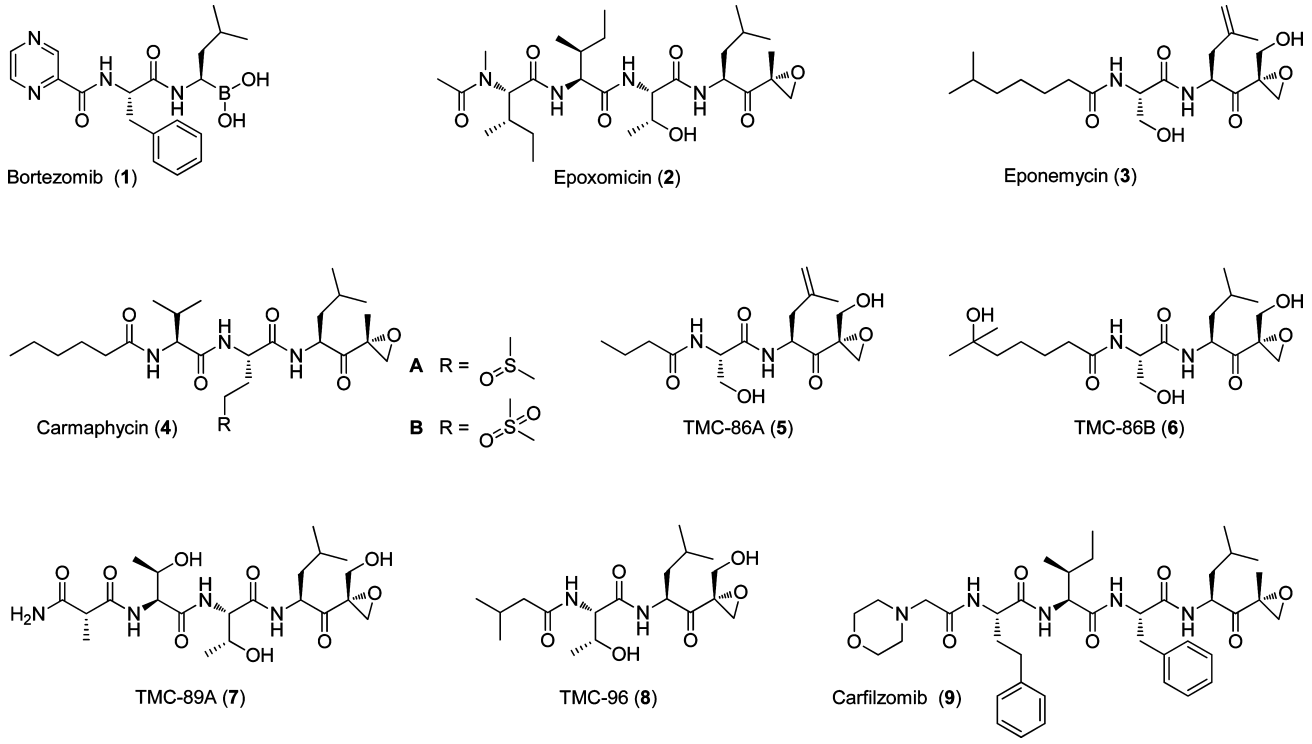


Figure 1. Chemical structures of proteasome inhibitors.

yeast proteasome revealed an irreversible two-step binding mechanism and the essential role of the α',β'-epoxyketone warhead.[15] The unprecedented terminal α',β'-position of the C3-carbonyl and the ring-strained epoxide constitutes two strongly electrophilic groups in the immediate proximity to each other, which are very accessible for nucleophilic attack. We were thus highly interested in the biosynthesis of these compounds and their unique pharmacophore. The biotransformation of the carboxy-terminus of a polyketide intermediate to an epoxide is biochemically not trivial and may involve either new polyketide biochemistry and/or unique enzymatic redox reactions.

In this study, we present the epoxomicin and the eponemycin gene cluster from an unspecified actinomycete strain ATCC 53904 and *Streptomyces hygroscopicus* strain ATCC 53709, respectively. Both compounds were produced by heterologous pathway expression in *S. albus* J1074 to definitively link epoxyketone proteasome inhibitors and their biosynthesis genes for the first time in any organism. This genetic linkage allowed us to locate homologous orphan gene clusters in various bacteria that promise the discovery of new bioactive derivative molecules.

RESULTS AND DISCUSSION

Identification of the Epoxomicin and Eponemycin Gene Clusters by Ion PGM Genome Sequencing. To investigate the biosynthetic pathways of epoxyketone proteasome inhibitors, we attempted to isolate the genes responsible for the formation of the prototypes **2** and **3**. To this end, we subjected genomic DNA of the producer strains ATCC 53904,[5] an unspecified actinomycete, and *S. hygroscopicus* ATCC 53709[6] to semiconductor sequencing from Ion Torrent.[16] Recently, we employed Ion Torrent technology in the *de novo* sequencing of the draft genome of *Thalassospira* sp. CNJ-328, which has a GC-content at around 50%.[17] However, we were unsuccessful in the sequencing of DNA with high GC-content such as from actinobacteria using standard protocols provided by the manufacturer. To address the sequencing problems, we slightly modified the procedure for manual template preparation as described in the Ion PGM 200 Xpress Template Kit. Betaine has been shown previously to substantially improve the amplification of difficult GC-rich DNA sequences.[18] As the Ion PGM template preparation is PCR based, we thus added betaine to a final concentration of 1 M to the amplification mix. After template preparation, enrichment, and sequencing with the Ion PGM system, the assembly of the obtained sequence data resulted in the generation of two draft genomes. The assembled sequence of the epoxomicin producer ATCC 53904 genome contains 8.9 Mb with a GC-content of 71.8% and was presented on 426 contigs with a 66-fold coverage. Similarly, the 9.8 Mb assembled genome sequence of the eponemycin producer ATCC 53709 has a GC-content of 71.1% and was presented on 490 contigs with a 79-fold coverage. Our modified protocol proved efficient and might thus facilitate the future application of semiconductor technology for the genome sequencing of other high-GC bacteria.

To assess the secondary metabolomic potential of both strains, we submitted the draft genome sequences to *in silico* analysis via the software antiSMASH.[19] Among 52 and 70 preliminary putative biosynthetic gene clusters, we identified potential clusters for the formation of **2** and **3** at 27.9 kb and 23.8 kb, respectively (Figure 2A). The homologous clusters encode hybrid nonribosomal peptide synthetase (NRPS)/polyketide synthase (PKS) multifunctional enzymes consistent with the putative formation of the core structure of the compounds. In addition, we identified analogous genes for a putative P450 monooxygenase and a conserved acyl-CoA dehydrogenase. A detailed summary of the proposed function of the genes can be found in the Supporting Information, Table S1.

dx.doi.org/10.1021/cb400699p | *ACS Chem. Biol.* 2014, 9, 301−309

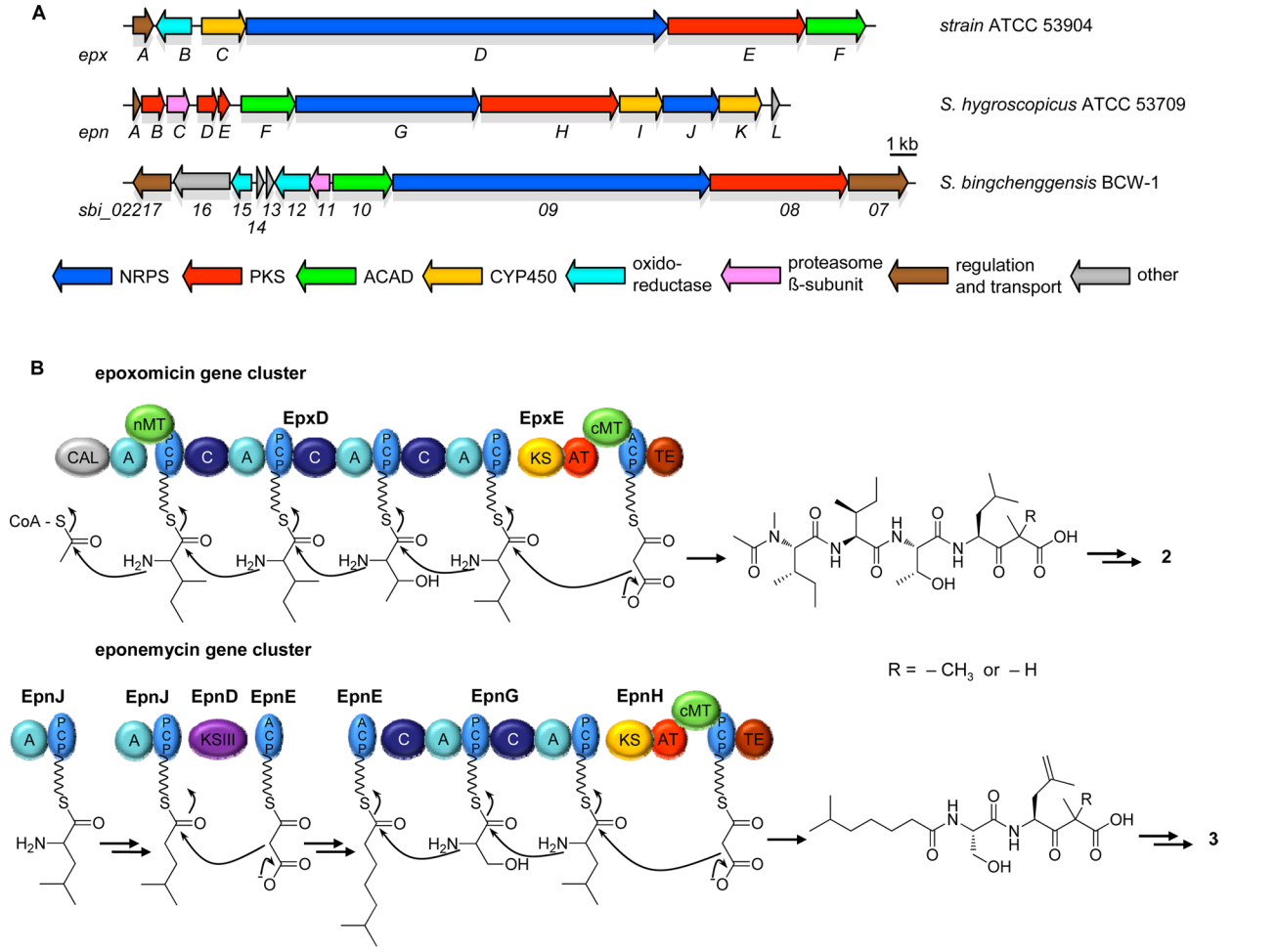


Figure 2. Biosynthesis of natural epoxyketones. (A) Relative organization of the epoxomicin (*epx*), eponemycin (*epn*) and orphan *S. bingchenggensis* gene clusters. (B) Predicted NRPS/PKS (nonribosomal peptide synthetase/polyketide synthase) assembly line synthesis of epoxomicin (**2**) and eponemycin (**3**). Abbreviations: ACAD, acyl-CoA dehydrogenase; CYP450, cytochrome P450; A, adenylation domain; ACP, acyl carrier protein; PCP, peptidyl carrier protein; C, condensation domain; MT, methyltransferase; TE, thioesterase; KS ketosynthase; AT acyltransferase domain.

These observations and the presence of a gene putatively encoding a resistant β-proteasome subunit homologous to the salinosporamide resistance enzyme[20] further suggested that the identified gene clusters code for epoxomicin (*epx*) and eponemycin (*epn*) production.

Heterologous Production of Epoxomicin and Eponemycin in *S. albus*. To confirm the suspected functions of the *epx* and *epn* gene clusters, we designed experiments to produce **2** and **3** in a surrogate host organism. To this end, we generated two fosmid libraries from the genomic DNA of both producer strains, ATCC 53904 and ATCC 53709, to isolate the gene clusters. The genomic libraries comprised ~1800 individual clones each and were screened by PCR. Both clusters were found intact on single fosmids, the epoxomicin gene cluster on fosmid 15C3 and the eponemycin gene cluster on fosmid 2H4. A heterologous expression approach was used to confirm the identity of the clusters.

For this purpose, we replaced the chloramphenicol resistance gene in the fosmid backbone by λ-Red-mediated recombination with an integration cassette we generated previously.[21] The cassette int_neo contains the attP attachment site and the integrase gene (*int*) of phage ΦC31, a kanamycin resistance gene (*neo*) and an origin of transfer (*oriT*), and allows site-specific integration in most *Streptomyces* chromosomes.[22] The resulting fosmids were named epxMS01 and epnLK01. In order to express the introduced pathway, the transcriptional and translational machinery of the host strain must recognize the promoter, ribosomal binding sites (RBS) and the regulatory system of the gene cluster. Consequently, the transfer of a biosynthetic gene cluster into a phylogenetically related strain is preferable for heterologous expression. While the eponemycin producer ATCC 53709 belongs to the genus *Streptomyces*, the taxonomic specification of the epoxomicin producer ATCC 53904 was not defined. We thus analyzed two phylogenetic markers, the 16S rRNA and the *rpoB* gene, for their relatedness to genes from other bacteria. Both markers classify the strain as a member of the Actinobacteria and the taxonomic family Pseudonocardiaceae. The *rpoB* gene shows highest homology (89% identity) to *Actinosynnema mirum* DSM 43827. The 16S rRNA confirms the close relatedness to *A. mirum* (96% identity) but indicates an even greater similarity to *Goodfellowia coeruleoviolacea* NRRL-B 24058 at 99% sequence identity.[23] Hence, we strongly consider ATCC 53904 to be a *Goodfellowia* species.

Both the epoxomicin and the eponemycin gene cluster were transferred into *S. albus* J1046 by intergeneric conjugation,[24] as this strain is one of the few in which non-*Streptomyces*-derived gene clusters have before been successfully expressed.[25] Three

dx.doi.org/10.1021/cb400699p | *ACS Chem. Biol.* 2014, 9, 301–309

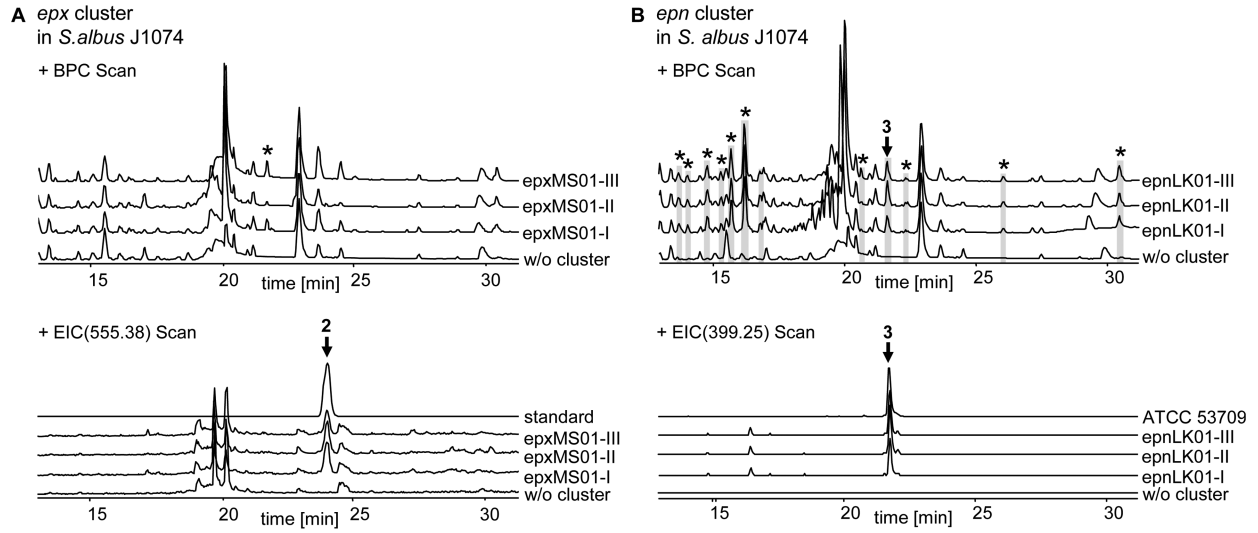


Figure 3. HPLC-MS analysis of (A) *S. albus* epxMS01-(1−3) expressing the epoxomicin gene cluster and (B) *S. albus* epnLK01-(1−3) expressing the eponemycin gene cluster. LC-MS base peak chromatograms (BPC, top) and extracted ion chromatograms (EIC, bottom) are depicted. Mass peaks of epoxomicin (**2**) and eponemycin (**3**) are indicated as well as unique peaks (*) in the *S. albus* heterologous mutants correlating to derivatives of **2** and **3** (see Supporting Information (SI) Table S2).

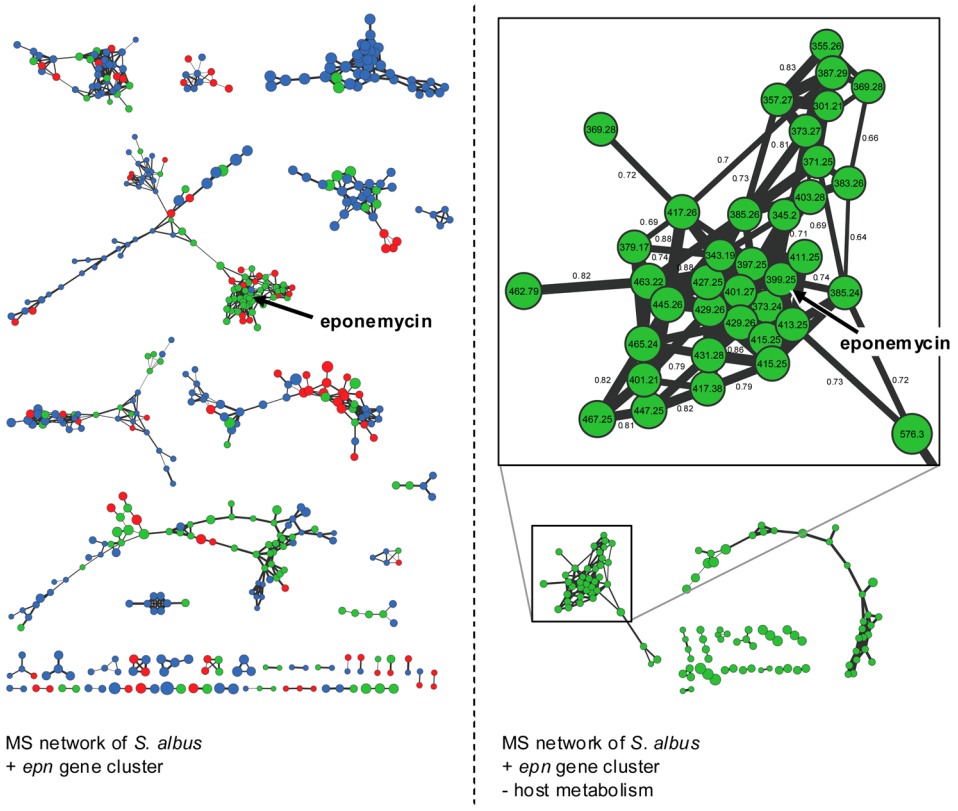


Figure 4. Molecular networks of mass spectra from *S. albus* containing the eponemycin gene cluster. Node (circle) colors indicate the source of the ions: blue (found in *S. albus* J1046), red (found only in *S. albus* epnLK01−1, −2, or −3), and green (found only in all three *S. albus* epnLK01-1, -2, and -3 strains). Node size indicates mass range of parent ions (m/z 200.17 − 1329.51). Edge line width indicates relatedness of MS/MS spectra represented by two connected nodes (cos 0.59−0.99). Selected masses and cosine values are noted in the eponemycin clade.

individual kanamycin resistant clones were selected and named *S. albus* epxMS01-(1−3) and *S. albus* epnLK01-(1−3), respectively. Ethyl acetate culture extracts of the mutant strains were analyzed by HPLC-MS and compared to the metabolic profiles of native *S. albus* J1064 and the producer strains ATCC 53904 and ATCC 53709 (Figure 3).

In the case of *S. albus* epnLK01, we discovered a new chromatographic peak not observed in *S. albus* J1064 (**3**, Figure 3B). Based on its chromatographic properties, its high-resolution mass, and its mass spectral fragmentation fingerprint (SI Figure S1), the product of the *S. albus*-expressed *epn* gene cluster is eponemycin (m/z calcd. 399.2495 $[M+H]^+$; obsvd. 399.2500). Moreover, we clearly identified at least 11 additional

dx.doi.org/10.1021/cb400699p | *ACS Chem. Biol.* 2014, 9, 301−309

epn-based products suggestive of the production of a series of eponemycin derivatives (Figure 3B and SI Table S2).

Analysis of the *S. albus* epxMS01 extracts (Figure 3A) confirmed the accumulation of **2** (m/z calcd. 555.3758 $[M+H]^+$; obsvd. 555.3758). However, heterologous production of **2** was low and detected only by HPLC retention time and MS/MS comparisons with authentic material. A possible explanation for its low production is the phylogenetic distance of *S. albus* and the epoxomicin producer ATCC 53904 and the resulting incompatibility of their expression system and regulatory networks. However, a distinct mass peak present in the ion chromatograms of the mutant strains clearly indicates the accumulation of an epoxomicin congener (Figure 3B and SI Table S2). Our genetic experiments successfully linked the biosynthesis of **2** and **3** to the *epx* and *epn* gene sets, respectively. This experimental outcome represents the first genes-to-molecules confirmation of the pharmaceutically important family of proteasome epoxyketone inhibitors.

MS Molecular Networking Reveals New Eponemycin Derivatives. While examining the base peak chromatograms of *S. albus* containing the epoxyketone gene clusters, we observed a number of additional unique mass peaks (Figure 3B), particularly in the *epn* mutant. Motivated by the prospect of new eponemycin derivatives, we subjected *S. albus* epnLK01 to MS molecular networking for comparative metabolic profiling. The spectral networks paradigm was originally developed for application in proteomics[26] but has recently been adapted as a general MS/MS-data analysis tool.[27,28] A molecular network is created based on the relationships of MS/MS spectra for any molecule that is captured by mass spectrometry, even across multiple experiments. The network we generated from the combined data of *S. albus* J1046 with and without the *epn* gene cluster displayed ten major MS/MS clusters comprising six or more nodes (circles) of distinct mass fragmentation (Figure 4). These clusters likely represent structurally related molecules and are also referred to as molecular families (MFs).[29] Upon subtracting mass ions consistent with the *S. albus* J1046 wildtype or with only a subset of the three heterologous mutants, eight of the ten major MFs were eliminated. One of the two remaining clusters forms a tight network centered on a node with m/z 399.246 representing **3** (Figure 4). Notably, at least twelve individual mass ions are directly related to this node.

The analysis of the MS/MS spectra incorporated in the network revealed a number of structural derivatives of **3** that are produced by the heterologous mutants (SI Figure S2). This is especially interesting as **3** has been reported as a single compound from the wildtype producer ATCC 53709. Based on the fragmentation patterns, we postulate that most of the variation applies to the length and oxidation status of the fatty acid side chain. The heterologously produced eponemycin analogues contain shorter (C_4) or longer (C_9) acyl moieties with double bonds, hydroxyl and/or keto groups consistent with distinct HPLC retention times (SI Table S2 and Figure S2). This observation suggests that the enzyme responsible for the attachment of the fatty acid group to the peptide is promiscuous and therefore may facilitate future bioengineering efforts. In addition, some of the mass spectra suggest the production of congeners of **3** with an altered epoxyketone pharmacophore. Notably, di- and tetrahydro derivatives are so far only known as synthetic compounds[6] but may be prominent metabolites in the *S. albus* epnLK01 extracts (SI Table S2). We plan to report the structures and biological properties of new eponemycin analogues separately.

Biosynthetic Pathways for Epoxomicin and Eponemycin. The peptidic backbone of the two epoxyketone compounds is assembled by hybrid nonribosomal peptide synthetase/polyketide synthase multifunctional enzymes (Figure 2B). EpxD consists of a tetra-modular NRPS and analysis of the adenylation (A) domain substrate specificities strongly correlated A_1 and A_3 to the activation of isoleucine and threonine, respectively.[30,31] The specificities for the A_2 and A_4 domains were less evident (SI Table S3). The megasynthetase begins with a putative primer fatty acyl condensation (C) domain[32] known to transfer fatty acids onto the initial amino acid residue.[33] We thus postulate that this domain is responsible for the installation of the acetyl moiety in **2**. The methyltransferase (MT) domain in the first module of EpxD probably further modifies the Ile1 residue by introducing the *N*-methyl group.

Interestingly, the corresponding eponemycin NRPS, EpnG, does not contain a specific primer C-domain. This observation rather suggests that the eponemycin fatty acyl moiety is not constructed in the same way as in **2**. Branched odd-chain fatty acids in the C_6–C_{10} range, such as 6-methyl heptanoic acid (6-MHA), can be found in natural products[34,35] but are not common in bacterial primary metabolism. Hence, a dedicated pathway for the generation and incorporation of 6-MHA might be encoded in the *epn* cluster. Normally, the biosynthesis of branched chain fatty acids involves FabH (KAS III), which accepts small CoA-activated acyl groups derived from leucine, valine, or isoleucine.[36] FabH catalyzes the Claisen condensation of these substrates with one unit of malonyl-ACP to initiate type II fatty acid synthesis (FAS). For the formation of 6-MHA, one would typically expect that valine-derived isobutyrate undergoes two rounds of malonate extension, one round with the help of FabH and the other carried out by common type II FAS enzymes. Notably, EpnD is a FabH-homologue similar to other enzymes that are occasionally encoded in secondary metabolite gene clusters and participate in the priming of type II PKS systems with unusual starter units[37] or the generation of acyl side chains in lipopeptide biosynthesis.[35] However, the presence of the discrete NRPS tridomain EpnJ, consisting of an mbtH-like protein, a leu-specific A-domain, and a peptidyl carrier protein (PCP), might indicate a more unusual mechanism for the biosynthesis of 6-MHA (Figure 2B). Here, EpnJ-bound leucine may be subjected to deamination and reduction prior to C_2-extension by the FabH homologue EpnD using malonyl-EpnE. Further reduction to afford 6-MHA would likely be performed by primary fatty acid synthase enzymes.

The terminal PKS-modules of the epoxomicin and the eponemycin assembly lines, EpxE and EpnH, respectively, are identically organized. Both comprise a malonyl-CoA-specific acyltransferase (AT)-domain and a putative C-methyltransferase (cMT) domain, which suggests that the substituted epoxy moiety does not derive from methylmalonyl-CoA but rather from malonyl-CoA and S-adenosylmethionine (Figure 2B). A C-terminal thioesterase (TE) domain in EpxE and EpnH intimates that the peptide-polyketide hybrid product is released from the enzyme as the carboxylic acid. In this case, the construction of the rare α',β'-epoxyketone unit would have to be mediated by auxiliary enzymes likely acting subsequent or *in trans* to the assembly of the core structure. The transformation from the free acid to the epoxide may include two reductions, a

dx.doi.org/10.1021/cb400699p | *ACS Chem. Biol.* 2014, 9, 301–309

dehydration and an epoxidation. Consequently, we analyzed the conserved biosynthesis genes in the epoxomicin and eponemycin loci, and gratifyingly, we identified two homologous genes common to both clusters—the putative acyl-CoA dehydrogenases (ACADs) *epxF*/*epnF* and the cytochrome P450 (CYP) monooxygenases *epxC*/*epnI*.

Cytochrome P450 enzymes are known to catalyze epoxidation reactions.[38] EpxC and EpnI are therefore plausible candidates to introduce an epoxy group into an unsaturated precursor molecule. Consequently, the reduction of the acid to the alcohol or the olefin may be performed by the ACADs EpxF/EpnF. Bacterial 4e$^-$ reductases that catalyze such reactions have been studied in the reductive off-loading of NRPSs[39] and in wax formation.[40] However, these enzymes rather belong to the short-chain dehydrogenase/reductase (SDR) superfamily, which is clearly distinct from the flavin adenine dinucleotide (FAD)-dependent ACADs. An alternative biosynthetic route to the epoxyketone moiety may involve the EpxE/EpnH cMT-domain if it introduces two methyl groups at C2. This scenario has been proposed in various polyketides with gem-dimethyl groups.[41] Subsequently, decarboxylation of the terminal carboxylic acid group to the dimethylketone may initiate a concerted reaction involving EpxF/EpnF and EpxC/EpnI to install the epoxide. On the basis of the biosynthetic features of the *epx* and *epn* gene clusters, the construction of the α',β'-epoxyketone pharmacophore is anticipated to involve new biochemical reactions. Detailed investigations of the pathway are therefore now underway. The distinct C8-OH group in eponemycin is likely implemented by the second CYP EpnK, which is unique to the *epn* cluster.

Genome Mining Identifies Homologous Gene Clusters in Various Bacteria. We next explored other microorganisms with the prospect to identify analogous pathways to new proteasome inhibitors. EpxF, the ACAD we predict to be essential for epoxyketone formation, was employed as a probe for a BLAST sequence homology search in the National Center for Biotechnology Information (NCBI) database. Strikingly, the genes with highest similarity to *epxF* (Expect (*E*) value < 1 e^{-70}) are all colocated with other genes from secondary metabolism. Most of these orphan biosynthetic pathways involve small hybrid NRPS/PKS systems and are present in a variety of different bacterial families (SI Figure S3). Notably, such homologous gene clusters can be found in a rhizosphere-associated *S. canus* 299MFChir4.1 strain as well as the human pathogen *Nocardia cyriacigeorgica* GUH-2. It might therefore be interesting to investigate if the compounds produced by these clusters have similar functions in symbiosis, virulence, and predation as the proteasome inhibitor family of syrbactins.[42,43] The cluster from *S. bingchenggensis* BCW-1 is particularly intriguing because it is highly analogous to the epoxomicin gene cluster (Figure 2A). We previously identified this set of genes and suggested it encodes a proteasome inhibitor, as it contains a secondary proteasome ß-subunit.[20] However, the *S. bingchenggensis* gene cluster lacks CYP homologues such as EpxC/EpnC that we propose are involved in the epoxidation reaction. We thus predict that this cluster may alternatively direct the formation of an acylated tripeptide in which the terminal amino acid moiety is modified not to an epoxyketone but rather as a vinylketone (enone) functional group. We recently showed in a separate study that synthetic carmaphycin enone derivatives are potent and irreversible proteasome inhibitors (B.S. Moore, unpublished data). Thus, we postulate that the orphan pathway in *S. bingchenggensis* BCW-1 encodes an unprecedented peptidic proteasome inhibitor.

In conclusion, we identified the epoxomicin and the eponemycin gene clusters, the first clusters for the biosynthesis of natural peptidyl epoxyketones. The genetic information suggests that their powerful pharmacophore is generated by a series of unprecedented biotransformations. With this study, we set the genetic basis to study the formation of the natural epoxyketone proteasome inhibitors in detail. Unexpectedly, we found a set of eponemycin congeners in extracts of a surrogate host organism with the *epn* cluster by molecular networking, demonstrating the benefits of this new technique for comparative metabolomics. Moreover, the presence of homologous gene clusters in other bacteria may facilitate the discovery of more new bioactive derivatives in the near future.

METHODS

Bacterial Strains and General Methods. Chemicals, microbiological, and molecular biological agents were purchased from standard commercial sources. Actinomycete strain ATCC 53904, *Streptomyces hygroscopicus* ATCC 53709, *Streptomyces albus* J1046, and their respective derivatives were maintained and grown on either MS agar (2% (w/v) soy flour, 2% (w/v) mannitol, 2% (w/v) agar; components purchased from Becton Dickinson) or TSB medium (Becton Dickinson). *Escherichia coli* strains were cultivated in LB medium (components purchased from Becton Dickinson) supplemented with appropriate antibiotics. DNA isolation and manipulations were carried out according to standard methods for *E. coli*[44] and *Streptomyces*.[45]

MS data were collected with an Agilent 6530 Accurate-Mass Quadrupole Time-of-flight (QTOF) LC-MS instrument (Agilent Technologies), and the analytes were separated with a reversed-phase C_{18} column (Phenomenex Luna 5 μ C18(2), 4.6 mm ×150 mm) on a 1260 Infinity LC-System (Agilent Technologies) using a flow rate of 0.1 mL/min.

Production of Epoxyketone Proteasome Inhibitors. TSB broth (10 mL) was inoculated with a spore suspension of *Streptomyces albus* J1046/epxMS01 or *Streptomyces albus* J1046/epnLK01 and incubated for 2 days at 30 °C and 200 rpm. Then 1% of the culture was transferred to 50 mL of R5 medium (103 g L^{-1} sucrose, 0.25 g L^{-1} K_2SO_4, 10.12 g L^{-1} $MgCl_2{\cdot}6H_2O$, 10 g L^{-1} glucose, 0.1 g L^{-1} casaminoacids, 5 g L^{-1} yeast extract, 5.73 g L^{-1} TES (*N*-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid), 80 μg L^{-1} $ZnCl_2$, 400 μg L^{-1} $FeCl_3{\cdot}6H_2O$, 20 μg L^{-1} $CuCl_2{\cdot}2H_2O$, 20 μg L^{-1} $MnCl_2{\cdot}4H_2O$, 20 μg L^{-1} $Na_2B_4O_7{\cdot}10H_2O$, 20 μg L^{-1} $(NH_4)_6Mo_7O_{24}{\cdot}4H_2O$, 50 mg L^{-1} KH_2PO_4, 3 g L^{-1} L-proline, 2.94 g L^{-1} $CaCl_2$, and 280 μg L^{-1} NaOH). After 6 days of incubation at 28 °C, 50 mL of EtOAc was added and incubated for 1 h at 200 rpm, and the EtOAc layer was recovered. The solvent was evaporated under reduced pressure. The residue was dissolved in 1 mL of MeCN, and the solution was filtered through a C_{18} sorbent (Spice C18 Sample Preparation Cartridges, Analtech). The filtrate was evaporated under reduced pressure in a 14-mL scintillation vial, and the residue was stored at −20 °C until LC/MS analysis.

Analysis of Culture Extracts and MS Molecular Networking. The residue was dissolved in 1000 μL of MeCN, and 5 μL of the dissolved residue was injected onto a reversed-phase HPLC column coupled to an mass spectrometer with an electrospray ionization interface (ESI) interface (heated capillary temperature 320 °C; sheath and collision gas nitrogen). The following solvent composition was used to separate the analytes: 10% (v/v) MeCN in H_2O for 4 min, 10−100% (v/v) MeCN in H_2O for 36 min, 100% (v/v) MeCN in H_2O for 3 min, 100−10% (v/v) MeCN in H_2O for 2 min. HR-MS data were acquired in positive mode ((+)-ESI). MS and MS/MS spectra were recorded with a scan rate of one and four spectra/second, respectively. Collision energy was 10 eV. Molecular formulas were calculated from monoisotopic masses using ChemCalc.[46]

To construct molecular networks MS/MS spectra recorded from extracts of *S. albus* J1046/epnLK01-(1−3) and *S. albus* J1046 wildtype were clustered using MS-Cluster.[47] Cluster-consensus spectra were further processed as described by Watrous et al.[28] Each spectrum comprised the 10 highest-cosine alignments in both directions. To define the MS/MS network pair wise alignments were considered with cosine ≥0.55 and ≥6 matched peaks. Custom scripts and the attributes to the molecular network were added as described.[28] The MS networks were visualized with Cytoscape (2.8.3.).[48]

DNA Sequencing and Bioinformatic Analysis. Ion Torrent libraries were prepared from genomic DNA (gDNA) of strains ATCC 53904 (epoxomicin producer) and ATCC 53709 (eponemycin producer), respectively, using the Ion Plus Fragment Library kit (Life Technologies) with a gDNA input of 1 μg. An S220 Focused-ultrasonicator (Covaris Inc.) was used to shear gDNA to obtain fragments of 100−250 bp size. Separation and extraction of DNA fragments was performed on the electrophoresis platform Pippin Prep (Sage Science Inc.). The Ion Library Quantitation Kit (Life Technologies) was used to quantify the libraries by quantitative real-time PCR on a LightCycler 480 (Roche Applied Science). No additional library amplification was performed. Samples were prepared manually using the Ion PGM 200 Xpress Template Kit (Life Technologies). To facilitate the amplification of the high GC content DNA, betaine was added to the amplification mix to a final concentration of 1M. The thermo profile was modified (95 °C for 10 min; 15 cycles, 95 °C for 30 s and 68 °C for 4 min; 30 cycles, 95 °C for 30 s and 68 °C for 6 min). Enrichment was performed on the Ion OneTouch ES (Life Technologies). Two sequencing runs per library were conducted on an Ion Personal Genome Machine (PGM) System using Ion 316 chips and the Ion PGM Sequencing 200 Kit v2 (Life Technologies). Sequencing data was assembled with the CLC Genomics Workbench software version 5.01 (CLC Bio). The draft genomes were subjected to the online tool antibiotics and Secondary Metabolite Analysis Shell (antiSMASH).[19] The *epx* and *epn* gene clusters were both found split on two contigs. The sequence gaps were closed by Sanger sequencing using primer walking and the shotgun method (GenoTech, Baejeon, Korea). Manual in silico sequence analysis was performed using GC frame-plot[49] and BLAST.[50] The Geneious software package (Biomatters Ltd.) and Artemis (Wellcome Trust Genome Campus) were used for sequence analysis and annotation.

Generation and Screening of Fosmid Libraries. The genomic fosmid libraries were constructed for strains ATCC 53904 and *Streptomyces hygroscopicus* ATCC 53709. High-molecular weight chromosomal DNA was randomly sheared to obtain fragments of ~40 kb size and cloned into pCC1FOS (Epicentre Biotechnologies). Fosmid libraries with ~1800 clones each were generated in *E. coli* EPI300 according to the manufacturers' instructions. For identification of the biosynthetic gene clusters the fosmid libraries were screened by PCR with primer pairs epxA_f GAATCTCAAGCGCGAGGGG/epxA_r GGTGTCGCGGAAGTAGTCC and epxF_f GCGCAC-CATGTCGCTGTTG/epxF_r GTAGTCGGGTGTCTCCTCC (library ATCC 53904) as well as epnA_f GTGTGGCCGTGAGCG-GATTC/epnA_r GCGGCCACGTTCCGATCTTG and epnK_f CAGCATGCTGCTGCAAGCCC/epnK_r CCCGGATGAAGTTC-GACCGC (library ATCC 53709). The primers were designed to amplify small specific fragments (0.3−0.5 kb) from the borders of the respective gene clusters.

Heterologous Expression of the *epx* and the *epn* Gene Clusters. An *Xba*I restriction fragment from merLK01 was generated representing an integration cassette (int_neo) for stable chromosomal integration. int_neo was used to replace *cat* in fosmids 15C3 (*epx* cluster) and 2H4 (*epn* cluster), as described previously.[21] The resulting fosmids epxMS01 and epnLK01 were verified by restriction analysis. The fosmids were transferred into *E. coli* ET12567 and introduced into *S. albus* J1046 by triparental intergeneric conjugation with the help of *E. coli* ET12567/pUB307. Kanamycin resistance mutants were selected and designated as *S. albus* J1046/epxMS01-(1−3) and *S. albus* J1046/epnLK01-(1−3), respectively.

■ ASSOCIATED CONTENT

Supporting Information

Supporting tables and graphics. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

The complete sequences of the epoxomicin (27908 bp) and the eponemycin (23794 bp) biosynthetic gene clusters were deposited at NCBI under the accession numbers KF647219 and KF647220, respectively.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: leonard.kaysser@pharm.uni-tuebingen.de.
*E-mail: bsmoore@ucsd.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. R. Kersten for assistance with mass spectrometry analysis. We are grateful to Dr. A. Lechner and H. Sun for help with molecular networking and semiconductor sequencing, respectively. This work was supported by the National Institutes of Health (NIH) (CA127622 to B.S.M. and GM097509 to P.C.D. and B.S.M.), a Feodor Lynen postdoctoral fellowship from the Alexander von Humboldt Foundation to L.K., and an NIH instrument grant (S10-OD010640).

■ REFERENCES

(1) Ciechanover, A. (1994) The ubiquitin-proteasome proteolytic pathway. *Cell 79*, 13−21.

(2) Lopes, U. G., Erhardt, P., Yao, R., and Cooper, G. M. (1997) p53-dependent induction of apoptosis by proteasome inhibitors. *J. Biol. Chem. 272*, 12893−12896.

(3) Drexler, H. C. (1997) Activation of the cell death program by inhibition of proteasome function. *Proc. Natl. Acad. Sci. U.S.A. 94*, 855−860.

(4) Grawert, M. A., and Groll, M. (2012) Exploiting nature's rich source of proteasome inhibitors as starting points in drug development. *Chem. Commun. (Camb.) 48*, 1364−1378.

(5) Hanada, M., Sugawara, K., Kaneta, K., Toda, S., Nishiyama, Y., Tomita, K., Yamamoto, H., Konishi, M., and Oki, T. (1992) Epoxomicin, a new antitumor agent of microbial origin. *J. Antibiot. (Tokyo) 45*, 1746−1752.

(6) Sugawara, K., Hatori, M., Nishiyama, Y., Tomita, K., Kamei, H., Konishi, M., and Oki, T. (1990) Eponemycin, a new antibiotic active against B16 melanoma. I. Production, isolation, structure, and biological activity. *J. Antibiot. (Tokyo) 43*, 8−18.

(7) Pereira, A. R., Kale, A. J., Fenley, A. T., Byrum, T., Debonsi, H. M., Gilson, M. K., Valeriote, F. A., Moore, B. S., and Gerwick, W. H. (2012) The carmaphycins: New proteasome inhibitors exhibiting an *α,β*-epoxyketone warhead from a marine cyanobacterium. *Chembiochem 13*, 810−817.

(8) Koguchi, Y., Kohno, J., Suzuki, S., Nishio, M., Takahashi, K., Ohnuki, T., and Komatsubara, S. (2000) TMC-86A, B and TMC-96, new proteasome inhibitors from *Streptomyces sp.* TC 1084 and *Saccharothrix sp.* TC 1094. II. Physico-chemical properties and structure determination. *J. Antibiot. (Tokyo) 53*, 63−65.

(9) Koguchi, Y., Nishio, M., Suzuki, S., Takahashi, K., Ohnuki, T., and Komatsubara, S. (2000) TMC-89A and B, new proteasome inhibitors from *Streptomyces sp.* TC 1087. *J. Antibiot. (Tokyo) 53*, 967−972.

(10) Kuhn, D. J., Chen, Q., Voorhees, P. M., Strader, J. S., Shenk, K. D., Sun, C. M., Demo, S. D., Bennett, M. K., van Leeuwen, F. W., Chanan-Khan, A. A., and Orlowski, R. Z. (2007) Potent activity of carfilzomib, a novel, irreversible inhibitor of the ubiquitin-proteasome

pathway, against preclinical models of multiple myeloma. *Blood 110*, 3281−3290.
(11) McCormack, P. L. (2012) Carfilzomib: In relapsed, or relapsed and refractory, multiple myeloma. *Drugs 72*, 2023−2032.
(12) Curran, M. P., and McKeage, K. (2009) Bortezomib: A review of its use in patients with multiple myeloma. *Drugs 69*, 859−888.
(13) Glenn, R. J., Pemberton, A. J., Royle, H. J., Spackman, R. W., Smith, E., Jennifer Rivett, A., and Steverding, D. (2004) Trypanocidal effect of α',β'-epoxyketones indicates that trypanosomes are particularly sensitive to inhibitors of proteasome trypsin-like activity. *Int. J. Antimicrob. Agents 24*, 286−289.
(14) Czesny, B., Goshu, S., Cook, J. L., and Williamson, K. C. (2009) The proteasome inhibitor epoxomicin has potent *Plasmodium falciparum* gametocytocidal activity. *Antimicrob. Agents Chemother. 53*, 4080−4085.
(15) Groll, M., Kim, K. B., Kairies, N., Huber, R., and Crews, C. M. (2000) Crystal structure of epoxomicin: 20S proteasome reveals a molecular basis for selectivity of α',β'-epoxyketone proteasome inhibitors. *J. Am. Chem. Soc. 122*, 1237−1238.
(16) Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Clark, T. A., Huber, M., Branciforte, J. T., Stoner, I. B., Cawley, S. E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J. A., Namsaraev, E., McKernan, K. J., Williams, A., Roth, G. T., and Bustillo, J. (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature 475*, 348−352.
(17) Ross, A. C., Xu, Y., Lu, L., Kersten, R. D., Shao, Z., Al-Suwailem, A. M., Dorrestein, P. C., Qian, P. Y., and Moore, B. S. (2013) Biosynthetic multitasking facilitates thalassospiramide structural diversity in marine bacteria. *J. Am. Chem. Soc. 135*, 1155−1162.
(18) Henke, W., Herdel, K., Jung, K., Schnorr, D., and Loening, S. A. (1997) Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res. 25*, 3957−3958.
(19) Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., and Breitling, R. (2011) antiSMASH: Rapid identification, annotation, and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res. 39*, 339−346.
(20) Kale, A. J., McGlinchey, R. P., Lechner, A., and Moore, B. S. (2011) Bacterial self-resistance to the natural proteasome inhibitor salinosporamide A. *ACS Chem. Biol. 6*, 1257−1264.
(21) Kaysser, L., Bernhardt, P., Nam, S. J., Loesgen, S., Ruby, J. G., Skewes-Cox, P., Jensen, P. R., Fenical, W., and Moore, B. S. (2012) Merochlorins A−D, cyclic meroterpenoid antibiotics biosynthesized in divergent pathways with vanadium-dependent chloroperoxidases. *J. Am. Chem. Soc. 134*, 11988−11991.
(22) Bierman, M., Logan, R., O'Brien, K., Seno, E. T., Rao, R. N., and Schoner, B. E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces spp*. *Gene 116*, 43−49.
(23) Labeda, D. P., and Kroppenstedt, R. M. (2006) *Goodfellowia gen. nov.*, a new genus of the Pseudonocardineae related to Actinoalloteichus, containing *Goodfellowia coeruleoviolacea gen. nov.*, comb. nov. *Int. J. Syst. Evol. Microbiol. 56*, 1203−1207.
(24) Flett, F., Mersinias, V., and Smith, C. P. (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett. 155*, 223−229.
(25) Lombo, F., Velasco, A., Castro, A., de la Calle, F., Brana, A. F., Sanchez-Puelles, J. M., Mendez, C., and Salas, J. A. (2006) Deciphering the biosynthesis pathway of the antitumor thiocoraline from a marine actinomycete and its expression in two streptomyces species. *ChemBioChem 7*, 366−376.
(26) Bandeira, N., Tsur, D., Frank, A., and Pevzner, P. A. (2007) Protein identification by spectral networks analysis. *Proc. Natl. Acad. Sci. U.S.A. 104*, 6140−6145.
(27) Guthals, A., Watrous, J. D., Dorrestein, P. C., and Bandeira, N. (2012) The spectral networks paradigm in high throughput mass spectrometry. *Mol. Biosyst. 8*, 2535−2544.
(28) Watrous, J., Roach, P., Alexandrov, T., Heath, B. S., Yang, J. Y., Kersten, R. D., van der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J. M., Moore, B. S., Laskin, J., Bandeira, N., and Dorrestein, P. C. (2012) Mass spectral molecular networking of living microbial colonies. *Proc. Natl. Acad. Sci. U.S.A. 109*, 1743−1752.
(29) Nguyen, D. D., Wu, C. H., Moree, W. J., Lamsa, A., Medema, M. H., Zhao, X., Gavilan, R. G., Aparicio, M., Atencio, L., Jackson, C., Ballesteros, J., Sanchez, J., Watrous, J. D., Phelan, V. V., van de Wiel, C., Kersten, R. D., Mehnaz, S., De Mot, R., Shank, E. A., Charusanti, P., Nagarajan, H., Duggan, B. M., Moore, B. S., Bandeira, N., Palsson, B. O., Pogliano, K., Gutierrez, M., and Dorrestein, P. C. (2013) MS/MS networking guided analysis of molecule and gene cluster families. *Proc. Natl. Acad. Sci. U.S.A. 110*, 2611−2620.
(30) Challis, G. L., Ravel, J., and Townsend, C. A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol. 7*, 211−224.
(31) Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol. 6*, 493−505.
(32) Rausch, C., Hoof, I., Weber, T., Wohlleben, W., and Huson, D. H. (2007) Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol. 7*, 78.
(33) Kraas, F. I., Giessen, T. W., and Marahiel, M. A. (2012) Exploring the mechanism of lipid transfer during biosynthesis of the acidic lipopeptide antibiotic CDA. *FEBS Lett. 586*, 283−288.
(34) Wilkinson, S., and Lowe, L. A. (1964) Structure of polymyxin B1. *Nature 202*, 1211.
(35) Powell, A., Borg, M., Amir-Heidari, B., Neary, J. M., Thirlway, J., Wilkinson, B., Smith, C. P., and Micklefield, J. (2007) Engineered biosynthesis of nonribosomal lipopeptides with modified fatty acid side chains. *J. Am. Chem. Soc. 129*, 15182−15191.
(36) Revill, W. P., Bibb, M. J., Scheu, A. K., Kieser, H. J., and Hopwood, D. A. (2001) β-ketoacyl acyl carrier protein synthase III (FabH) is essential for fatty acid biosynthesis in *Streptomyces coelicolor* A3(2). *J. Bacteriol. 183*, 3526−3530.
(37) Moore, B. S., and Hertweck, C. (2002) Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat. Prod. Rep. 19*, 70−99.
(38) Thibodeaux, C. J., Chang, W. C., and Liu, H. W. (2012) Enzymatic chemistry of cyclopropane, epoxide, and aziridine biosynthesis. *Chem. Rev. 112*, 1681−1709.
(39) Read, J. A., and Walsh, C. T. (2007) The lyngbyatoxin biosynthetic assembly line: Chain release by four-electron reduction of a dipeptidyl thioester to the corresponding alcohol. *J. Am. Chem. Soc. 129*, 15762−15763.
(40) Hofvander, P., Doan, T. T., and Hamberg, M. (2011) A prokaryotic acyl-CoA reductase performing reduction of fatty acyl-CoA to fatty alcohol. *FEBS Lett. 585*, 3538−3543.
(41) Gulder, T. A., Freeman, M. F., and Piel, J. (2011) The catalytic diversity of multimodular polyketide synthases: Natural product biosynthesis beyond textbook assembly rules. *Top. Curr. Chem.*, DOI: 10.1007/128_2010_113.
(42) Stein, M. L., Beck, P., Kaiser, M., Dudler, R., Becker, C. F., and Groll, M. (2012) One-shot NMR analysis of microbial secretions identifies highly potent proteasome inhibitor. *Proc. Natl. Acad. Sci. U.S.A. 109*, 18367−18371.
(43) Krahn, D., Ottmann, C., and Kaiser, M. (2011) The chemistry and biology of syringolins, glidobactins, and cepafungins (syrbactins). *Nat. Prod. Rep. 28*, 1854−1867.
(44) Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York.
(45) Kieser, T., Bibb, M., Buttner, M., Chater, K., and Hopwood, D. (2000) *Practical Streptomyces Genetics*, The John Innes Foundation, Norwich, U.K.

(46) Patiny, L., and Borel, A. (2013) ChemCalc: A building block for tomorrow's chemical infrastructure. *J. Chem. Inf. Model 53*, 1223−1228.
(47) Frank, A. M., Bandeira, N., Shen, Z., Tanner, S., Briggs, S. P., Smith, R. D., and Pevzner, P. A. (2008) Clustering millions of tandem mass spectra. *J. Proteome Res. 7*, 113−122.
(48) Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., and Ideker, T. (2011) Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics 27*, 431−432.
(49) Bibb, M. J., Findlay, P. R., and Johnson, M. W. (1984) The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. *Gene 30*, 157−166.
(50) Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res. 25*, 3389−3402.