JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

COMMUNICATION

pubs.acs.org/JACS

Proteomics-Based Discovery of Koranimine, a Cyclic Imine Natural Product

Bradley S. Evans,[†] Ioanna Ntai,[‡] Yunqiu Chen,[‡] Sarah J. Robinson,[‡] and Neil L. Kelleher*[,‡]

[†]Institute of Genomic Biology, University of Illinois Urbana−Champaign, 1206 West Gregory Drive, Urbana, Illinois 61801, United States

[‡]Department of Chemistry and Chemistry of Life Processes Institute, Northwestern University, 2145 North Sheridan Road, Evanston, Illinois 60208, United States

Supporting Information

ABSTRACT: Nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) are large enzymes responsible for the biosynthesis of medically and ecologically important secondary metabolites. In a previous report, we described a proteomics approach to screen for expressed NRPSs or PKSs from bacteria with or without sequenced genomes. Here we used this proteome mining approach to discover a novel natural product arising from rare adenylation (A) and reductase (Red) domains in its biosynthetic machinery. We also cloned the entire gene cluster and elucidated the biosynthesis of the new compound, which is produced by an unsequenced *Bacillus* sp. isolated from soil collected in Koran, Louisiana.

The potent bioactivities of peptides produced by nonribosomal biosynthesis range from antibiotics such as penicillins and vancomycin to anticancer drugs such as bleomycin.[1] Members of this chemical family include immunosuppressants such as cyclosporin,[2] pathogenicity determinants such as yersiniabactin[3] and aureusimine,[4] and powerful mycotoxins such as HC-toxin.[5] With the deluge of microbial genome sequences becoming available, it is apparent that secondary metabolite gene clusters are widely dispersed and largely uncharacterized and that they often encode NRPSs or PKSs.[6] Despite some successes[7,8] in harnessing the genetic potential seen in microbial genomes, forced expression of the 30−80 kilobases of an entire gene cluster has not evolved into a generally robust method for exploiting their rich biosynthetic capacity. The proteomic approach described herein instead seeks to screen native hosts under a wide range of culture conditions. Recently, proteomic approaches for studying secondary metabolism have proven useful for the detection and characterization of thiotemplate biosynthetic pathways.[9,10] In the present study, 23 unsequenced environmental isolates were cultured under nine different sets of conditions and then sampled at three time points before screening by simple SDS-PAGE, as described previously.[10] Of these 23 strains, 20 showed some evidence of >150 kDa protein bands; these tend to be large NRPS or PKS enzymes, which often are >1300 amino acids in length.

A mass spectrometric analysis of four high-molecular-weight protein bands from one environmental *Bacillus* isolate is shown in Figure S1 in the Supporting Information. Direct peptide sequencing by tandem mass spectrometry (i.e., the spectrum in Figure S1b) gave the amino acid subsequences that guided the design of PCR primers for amplification of DNA corresponding to the region between the identified peptides and the conserved core regions of NRPS adenylation domains (Figure S1c). A single amplicon was our hook into the unknown gene cluster. On the basis of the low sequence identity of this amplicon's translated amino acid sequence (42%) to the closest sequence in GenBank, it was possible that we had uncovered a yet-unsequenced biosynthetic pathway. Eventual DNA sequencing of this >30 kilobase gene cluster and prediction of its functional elements led us quickly to predict and then detect a new peptide natural product named koranimine (**1**) (*m*/*z* 804.502; Figure 1).

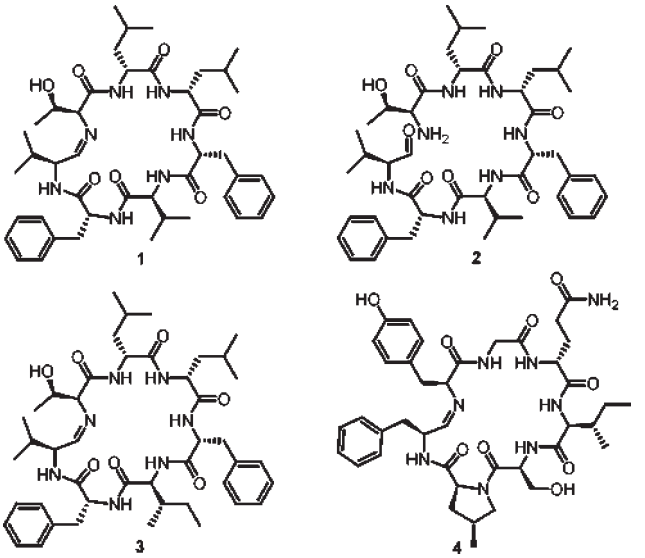
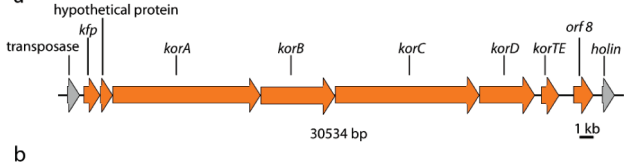


Figure 1. Koranimine (**1**) is the major product. Compound **2** is the direct result of reductive release from the assembly line that forms the imine structure in **1**. Compound **3** is a minor product that has isoleucine substituted for valine at the fifth amino acid in the peptide. Compound **4** is nostocyclopeptide A1.

When the entire 30 kb cluster (Figure 2) was cloned from our environmental strain (NK2003) and heterologously expressed in

Received: February 19, 2011
Published: April 26, 2011

a



b

Genes of the koranimine biosynthetic pathway

orf name	bp	amino acids	function	Homologue (% AA identity)
kfp	714	237	Phosphopantetheinyl transferase	BAH44970.1 (33%)
korA	7815	2605	NRPS	BAH43927.1 (32%)
korB	4710	1545	NRPS	ACC81529.1 (35%)
korC	7707	2568	NRPS	EDZ50535.1 (37%)
korD	2889	962	NRPS	ACK73815.1 (35%)
korTE	732	243	Type II thioesterase	EEM05992.1 (35%)

Figure 2. (a) Open reading frame (ORF) map for the koranimine gene cluster. (b) Annotation of the koranimine genes.

Bacillus subtilis 168, a time-dependent increase in the amount of **1** was observed at 25 min in the total ion chromatogram of a liquid chromatography/Fourier transform mass spectrometry (LC/FT-MS) run (Figure S2; m/z 804.502). Additionally, the species at m/z 804.502 was not present in the culture supernatant of the control strain (*B. subtilis* 168; Figure S2). These data proved the association between the koranimine biosynthetic gene cluster (30 kb) and its small-molecule product. The gene cluster of **1** is not highly related to any other NRPS clusters present in GenBank at this time, a noteworthy result after more than a decade of shotgun sequencing of bacterial genomes.

Initial tandem mass spectrometric analysis of **1** identified the six amino acid sequence tag Thr-Leu/Ile-Leu/Ile-Phe-Val-Phe (Figure S3). Stable isotope incorporation studies revealed the presence of one threonine residue, two leucine residues, two phenylalanine residues, and two valine residues in the koranimine molecule (Figure S4). Only minor levels of misincorporation of Ile instead of Leu were observed (Figure S4c). However, the feeding results using $^{13}C_6$-labeled isoleucine verified the production of a related compound incorporating Ile instead of Val at the fifth amino acid position (**3** in Figure 1). Analysis of D_{10}-leucine- and D_8-$^{15}N_1$-phenylalanine-labeled koranimine revealed complete abstraction of the α-deuterium in both amino acids, indicating that the domains predicted to have epimerization function were in fact converting all instances of L-Leu and L-Phe to their D-Leu and D-Phe stereoisomers, respectively (Figure S5).

Isolation of 1.1 mg of **1** from 20 L of culture followed by structural interpretation using 2D NMR techniques confirmed the presence of Val, Phe, and Leu and some minor incorporation of Ile vs Val (Figure S6), all consistent with the mass spectrometric data in Figures S3, S4, and S5. Because of the limited amount of material, indirect-detection gHMBC experiments were used to detect signatures for Phe, Leu, and Val (Figure S6). A main structural aspect of koranimine exhibiting apparent chemical variation in both MS and NMR analyses is the terminal imine functionality, which arises from decidedly nonstandard NRPS termination chemistry (see below). By MS, signals consistent with both the cyclic imine form (**1**) and the open aldehyde form (**2**; +18.01 Da) (see Figure 1) were observed. The NMR data did not show the expected imine ^1H at ~6.8 ppm or the imine carbon at ~167 ppm (Figure S6). Subjecting the purified product to inductively coupled plasma MS (ICP-MS) showed no trace of any bound mono-, di-, or trivalent metal. However, reduction by $NaCNBH_3$ showed a clear increase of 2.012 Da (Figure S7), leading us to assign with confidence the imine structure for compound **1**.

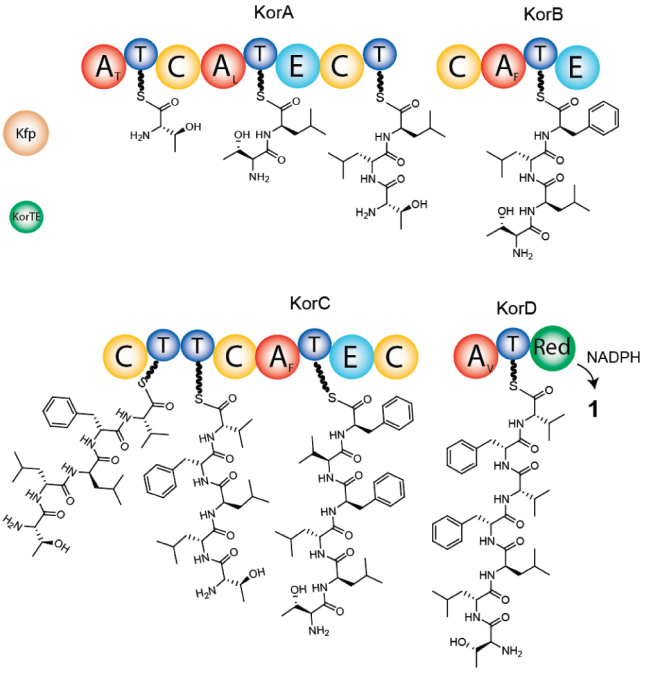


Figure 3. Domain structures of the synthetases and the biosynthetic scheme for production of koranimine. Symbols: A, adenylation; T, thiolation; C, condensation; E, epimerization; Red, reductase.

The biosynthetic mechanism for koranimine production is shown in Figure 3, with several unusual aspects of the domains and their overall organization. Only after the structure of **1** was determined were we able to make a proposal for the entire biosynthetic pathway with confidence. The first peculiarity is that the natural product was determined to be a heptapeptide, but the gene cluster contains only five adenylation domains. Prediction of substrates for adenylation domains is routine in the NRPS field.[11] This exercise accurately predicted only the first, second, and seventh amino acids to be threonine, leucine, and valine, which are installed by the adenylation domains KorA-A_1 and KorA-A_2 and the adenylation domain of KorD, respectively. The adenylation domains of KorB and KorC are both only 42% identical at the amino acid level to their nearest homologues (gb: ACC81021.1 and gb: AAO23333.1, respectively), and standard algorithms produced no substrate predictions, indicating that these enzymes contain highly unusual adenylation domain signatures. The adenylation domains of KorB and KorC were ultimately implicated in the incorporation of Phe by metabolic labeling (Figure S4d).

Another peculiarity in the koranimine system is KorA, where a condensation−thiolation (C−T) didomain lacks a directly linked adenylation (A) or epimerization (E) domain. Deductions from the koranimine structure indicate that KorA incorporates and epimerizes the second and third amino acids (both leucine). Therefore, a single adenylation domain likely services two carrier sites. While an adenylation domain acting in this way is unconventional, it is not without precedent.[12] This particular example opens up two mechanistic possibilities, both unprecedented. In path a of Figure S8, the adenylation domain services both carrier sites (either in cis or in trans), forcing the single epimerization domain to act at both sites. Alternatively, in path b of Figure S8, the D-Leu−acyl-S−KorA-T_2−enzyme intermediate would form, after which an unprecedented shuttling of this

dx.doi.org/10.1021/ja2015795 |*J. Am. Chem. Soc.* 2011, 133, 7316–7319

intermediate by the condensation domain would move the D-Leu intermediate to KorA-T_3, opening up KorA-T_2 for a second round of adenylation and epimerization. Next, KorA-C_2 would switch roles to catalyze the standard type of condensation and form the amide bond between two D-Leu residues.

In KorC, there is yet another unusual domain structure containing a T−T−C tridomain, also with no directly linked adenylation domain. Biochemical experiments showed that the tandem T domains of KorC are both charged by the adenylation domain of KorD (Figure S9). This long-range transactivation, similar to the long-range transactivation of HMWP1 by HMWP2 in the yersiniabactin pathway,[12] may be a rate-limiting step and could explain the presence of the tandem T domains in the fourth module. Tandem T domains have previously been implicated in increasing the flux through thiotemplate systems.[13,14]

We propose that KorTE has an editing function like that of other type-II thioesterases, namely, removing mis-acylated carrier sites from the pathway and increasing the efficiency of the pathway.[15,16] KorTE did decrease the occupancy of acyl-S−KorC (Figure S9g). We envisioned that the long-range transaminoacylation of KorC by KorD might be facilitated by the action of a shuttling function of the orf 8 enzyme, predicted to be an α/β hydrolase family enzyme (in a fashion similar to SyrC of the syringomycin pathway[17]). In vitro experiments showed that the orf 8 enzyme does have hydrolytic activity toward acyl-S−enzyme intermediates (Figure S10f). However, it did not stimulate substrate transfer from KorD to KorC in vitro (Figure S9f).

The sixth and final NRPS module, in addition to charging the fourth module and incorporating the seventh residue, valine, performs a two-electron, NADPH-dependent reduction of the C-terminal acyl-S-thioester acid to the aldehyde. This terminal aldehyde functionality has previously been shown to be involved in nonenzymatic macrocyclization[18] yielding the final imine product, **1**. The only precedent for this is the nostocyclopeptide system[19] (nostocyclopeptide A1, **4**), where the first reductive chain release was shown. However, the nostocyclopeptides have completely different amino acid sequences (Figure 1). The imine functionality for nostocyclopeptide has been shown to be reversible in an NMR tube,[18] and evidence points to a spontaneous formation upon reductive release of an aldehyde from the NRPS assembly line. The functional impact of having a sequence dissimilar to nostocyclopeptide[18,20] is unknown at this time, though a functional study in 2010 demonstrated that a nostocyclopeptide antagonizes the toxic effect of microcystin on hepatocytes[21] and blocks hepatocyte drug trransporters.[22] Koranimine was not found to inhibit the growth of *Escherichia coli*, *Staphylococcus aureus*, *Enterococcus faecalis*, *Klebsiella pneumonia*, or *Saccharomyces cerevisiae*. Further bioassessment of **1** to probe the limited structural discrepancies between **1** and **4** is underway.

In summary, the structure of a new natural product was determined using multistage tandem MS (MS^n), feeding studies with stable isotope tracers, NMR spectroscopy, and in vitro enzyme reconstitution. The original discovery of the compound began with a new "proteome-first" strategy that scanned microbial proteomes for *expressed* gene clusters, even those not yet uncovered by genome sequencing efforts such as koranimine. Furthermore, the approach can be used to screen conditions where typically cryptic gene clusters can be teased into expression. Future efforts for expression screening at the protein level will become more efficient and sensitive and serve as a high-value sieve for large numbers of environmental isolates in the discovery of novel natural products and their biosynthetic pathways. This discovery platform is structure-based, does not rely on bioassays, and can even be applied to exploration of cross- and mixed systems of microbial ecology (i.e., "metaproteomics").

■ ASSOCIATED CONTENT

Supporting Information. Supplementary figures; methods for molecular biological techniques, compound isolation, and mass spectrometry; and supplemental references. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
n-kelleher@northwestern.edu

■ ACKNOWLEDGMENT

We thank William Metcalf, Huimin Zhao, Wilfred van der Donk, Satish Nair, Doug Mitchell, and Karl Scheidt for comments. We also thank the Roy J. Carver Charitable Trust, the University of Illinois, and the National Institutes of Health (GM R01 067725 and GM P01 077596), which have supported the development of PrISM.

■ REFERENCES

(1) Du, L. C.; Sanchez, C.; Chen, M.; Edwards, D. J.; Shen, B. *Chem. Biol.* **2000**, *7*, 623–642.
(2) Weber, G.; Schorgendorfer, K.; Schneiderscherzer, E.; Leitner, E. *Curr. Genet.* **1994**, *26*, 120–125.
(3) Pelludat, C.; Rakin, A.; Jacobi, C. A.; Schubert, S.; Heesemann, J. *J. Bacteriol.* **1998**, *180*, 538–546.
(4) Wyatt, M. A.; Wang, W.; Roux, C. M.; Beasley, F. C.; Heinrichs, D. E.; Dunman, P. M.; Magarvey, N. A. *Science* **2010**, *329*, 294–296.
(5) Walton, J. D. *Phytochemistry* **2006**, *67*, 1406–1413.
(6) Donadio, S.; Monciardini, P.; Sosio, M. *Nat. Prod. Rep.* **2007**, *24*, 1073–1109.
(7) Knappe, T. A.; Knappe, T. A.; Linne, U.; Zirah, S.; Rebuffat, S.; Xie, X.; Marahiel, M. A. *J. Am. Chem. Soc.* **2008**, *130*, 11446–11454.
(8) Lautru, S.; Deeth, R. J.; Bailey, L. M.; Challis, G. L. *Nat. Chem. Biol.* **2005**, *1*, 265–269.
(9) Meier, J. L.; Niessen, S.; Hoover, H. S.; Foley, T. L.; Cravatt, B. F.; Burkart, M. D. *ACS Chem. Biol.* **2009**, *11*, 948–957.
(10) Bumpus, S. B.; Evans, B. S.; Thomas, P. M.; Ntai, I.; Kelleher, N. L. *Nat. Biotechnol.* **2009**, *27*, 951–956.
(11) Rausch, C.; Weber, T.; Kohlbacher, O.; Wohlleben, W.; Huson, D. H. *Nucleic Acids Res.* **2005**, *33*, 5799–5808.
(12) Suo, Z.; Tseng, C. C.; Walsh, C. T. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 99–104.
(13) Jiang, H.; Zirkle, R.; Metz, J. G.; Braun, L.; Richter, L.; Van Lanen, S. G.; Shen, B. *J. Am. Chem. Soc.* **2008**, *130*, 6336–6337.
(14) Rahman, A. S.; Hothersall, J.; Crosby, J.; Simpson, T. J.; Thomas, C. M. *J. Biol. Chem.* **2005**, *280*, 6399–6408.
(15) Kim, B. S.; Cropp, T. A.; Beck, B. J.; Sherman, D. H.; Reynolds, K. A. *J. Biol. Chem.* **2002**, *277*, 48028–48034.
(16) Schwarzer, D.; Mootz, H. D.; Linne, U.; Marahiel, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14083–14088.
(17) Singh, G. M.; Vaillancourt, F. H.; Yin, J.; Walsh, C. T. *Chem. Biol.* **2007**, *14*, 31–40.
(18) Enck, S.; Kopp, F.; Marahiel, M. A.; Geyer, A. *ChemBioChem* **2008**, *9*, 2597–2601.
(19) Becker, J. E.; Moore, R. E.; Moore, B. S. *Gene* **2004**, *325*, 35–42.
(20) Kopp, F.; Mahlert, C.; Grunewald, J.; Marahiel, M. A. *J. Am. Chem. Soc.* **2006**, *128*, 16478–16479.

(21) Jokela, J.; Herfindal, L.; Wahlsten, M.; Permi, P.; Selheim, F.; Vasconçelos, V.; Døskeland, S. O.; Sivonen, K. *ChemBioChem* **2010**, *11*, 1594–1599.
(22) Herfindal, L.; Myhren, L.; Kleppe, R.; Krakstad, C.; Selheim, F.; Jokela, J.; Sivonen, K.; Døskeland, S. O. *Mol. Pharmaceutics* **2011**, *8*, 360–367.