BIOCHEMISTRY

 Volume 45, Number 6 February 14, 2006

Accelerated Publications

Activity Screening of Carrier Domains within Nonribosomal Peptide Synthetases Using Complex Substrate Mixtures and Large Molecule Mass Spectrometry†

Pieter C. Dorrestein,‡ Jonathan Blackhall,‡ Paul D. Straight,§ Michael A. Fischbach,|| Sylvie Garneau-Tsodikova,|| Daniel J. Edwards,⊥ Shaun McLaughlin,‡ Myat Lin,‡ William H. Gerwick,@ Roberto Kolter,§ Christopher T. Walsh,|| and Neil L. Kelleher*,‡

Department of Chemistry, University of Illinois, Urbana, Illinois 61801, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093, and Department of Chemistry, California State University, Chico, California 95929

Received November 15, 2005; Revised Manuscript Received January 3, 2006

ABSTRACT: For screening a pool of potential substrates that load carrier domains found in nonribosomal peptide synthetases, large molecule mass spectrometry is shown to be a new, unbiased assay. Combining the high resolving power of Fourier transform mass spectrometry with the ability of adenylation domains to select their own substrates, the mass change that takes place upon formation of a covalent intermediate thus identifies the substrate. This assay has an advantage over traditional radiochemical assays in that many substrates, the substrate pool, can be screened simultaneously. Using proteins on the nikkomycin, clorobiocin, coumermycin A_1, yersiniabactin, pyochelin, and enterobactin biosynthetic pathways as proof of principle, preferred substrates are readily identified from substrate pools. Furthermore, this assay can be used to provide insight into the timing of tailoring events of biosynthetic pathways as demonstrated using the bromination reaction found on the jamaicamide biosynthetic pathway. Finally, this assay can provide insight into the role and function of orphan gene clusters for which the encoded natural product is unknown. This is demonstrated by identifying the substrates for two NRPS modules from the *pksN* and *pksJ* genes that are found on an orphan NRPS/PKS hybrid cluster from *Bacillus subtilis*. This new assay format is especially timely for activity screening in an era when new types of thiotemplate assembly lines that defy classification are being discovered at an accelerating rate.

Almost weekly, new nonribosomal peptide synthetase (NRPS)[1] gene clusters encoding proteins that biosynthesize bioactive compounds are discovered (*1*−*4*). Since many of the compounds produced by the nonribosomal peptide synthetase (NRPS) as well as polyketide synthase (PKS) paradigm have potent medicinal utility or are involved in

† This work was supported in part by NIH Grants GM 49338 (C.T.W.), 067725 (N.L.K.), GM 58213 (R.K.), and CA 83155 (W.H.G.), NIH Kirschstein NRSA Postdoctoral Fellowship F32-GM 073323-01 (P.C.D.), National Science Foundation Postdoctoral Fellowship in Microbial Biology DBI-0200307 (P.D.S.), and the Hertz Foundation Graduate Fellowship (M.A.F.).

* To whom correspondence should be addressed. E-mail: kelleher@scs.uiuc.edu. Fax: (217) 244-8068. Phone: (217) 244-3927.

‡ University of Illinois.

§ Department of Microbiology and Molecular Genetics, Harvard Medical School.

|| Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School.

⊥ California State University.

@ University of California at San Diego.

10.1021/bi052333k CCC: $33.50
Published on Web 01/24/2006

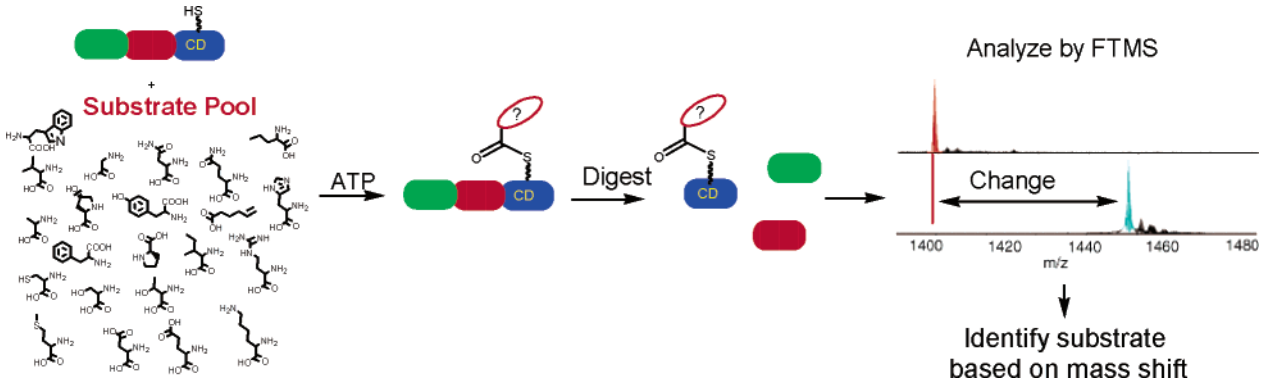

FIGURE 1: General approach of using FTMS to screen for acylation of a carrier domain using a substrate pool.

virulence and display unusual chemistry, they are of great academic and industrial interest. Therefore, we wanted to develop an alternative method for substrate screening to complement the more traditional radioactive assays (*5*, *6*). In NRPS and PKS systems, the substrates and intermediates are loaded onto and processed while attached to the pantetheinyl functionality on a carrier domain (*7*, *8*). Examples of NRPS or hybrid NRPS/PKS natural products for which substrates load onto carrier domains are the siderophore pyoverdine (*9*), a virulence factor excreted by pseudomonads, the antimicrobial agents penicillin (*10*), vancomycin (*11*), and gramicidin (*12*), and the antitumor agent calicheamycin (*13*). With ~300 genomes sequenced and available in the public domain, there are a large number of orphan PKS and NRPS gene clusters that have been identified (*14*, *15*). Even when the sequence of a gene cluster with a known natural product is available, progress in characterizing the proteins in vitro has been slow. Among the limiting factors in our understanding of how these NRPS systems biosynthesize their respective natural products is the inefficiency in obtaining in vitro activity to verify and identify the substrates for these enzymes. One of the many reasons for this inefficient activity screening includes the unavailability of the proper radiolabeled substrates or intermediates. While excellent advancements have been made using bioinformatics to predict the substrate for adenylating domains (*16*−*18*), caution should be taken when relying solely on the predictions to assign a specific substrate (*19*). Furthermore, new types of thiotemplate assembly lines that defy classification or thiotemplate assembly lines that do not follow the standard collinearity rules are being discovered at an accelerating rate (*1*−*4*, *20*). As a general strategy, all bioinformatic predictions should be confirmed in vitro. With the recent push to develop NRPS-derived bioactive compounds in vitro (*21*, *22*) and to manipulate these systems in vivo to generate new pharmaceutically useful bioactive compounds (*23*−*25*), the in vitro characterization must be performed more efficiently. One of the ways one can improve the ability to screen for substrates is by being able to screen many substrates simultaneously. This assay would ideally also avoid the use of radiolabeled substrates which are used to load carrier domains usually used in autoradiographic analysis. Using non-radiolabeled substrates significantly broadens the accessibility of substrates that are commercially available or synthetically feasible. Here we describe a mass spectrometry-based method of identifying substrates on the basis of mass changes that take place during acylation of a phosphopantetheinyl functionality on the carrier domain(s) of a NRPS protein from very complex substrate reaction mixtures. Observing such acylations by mass spectrometry on carrier domains is now becoming routine (*5*, *19*, *26*−*36*).

The general method of identifying covalently loaded substrates or intermediates is shown in Figure 1. Overproduced protein that contains a carrier domain is purified and incubated with Sfp, a promiscuous phosphopantetheinyl transferase from *Bacillus subtilis*, and CoA to generate the holo form of the carrier domain (*37*). A substrate pool and, when required, a separately purified activating domain are added to the holo carrier domain. Following an incubation period, the reaction mixture is digested, quenched using formic acid, and purified by HPLC prior to ESI-FTMS analysis.

RESULTS

NikP1, which is involved in the formation of the antibiotic nikkomycin (*6*), was incubated with Sfp and CoA to generate the holo form. Holo-NikP1 was incubated with ATP and all 19 proteinogenic L-amino acids, glycine, L-selenocysteine, L-cystine, and 4-*trans*-hydroxy-L-proline for 30 min before being quenched, digested using cyanogen bromide, and analyzed by FTMS. The FTMS data indicated a mass shift of 137.2 Da, in agreement with the loading of histidine onto the phosphopantetheinyl thiol (Figure 2C). When the same assay is repeated but L-histidine is omitted, NikP1 remained in its holo form, demonstrating that NikP1 is highly selective for the loading of histidine relative to all other substrates present in the assay mixture (Figure 2B).

To illustrate that this method was valid for not only NikP1, the substrate loadings for the following systems were examined: the CloN5/CloN4 carrier/adenylation domain pair involved in the formation of clorobiocin (Figure 3 and Table 1) (*5*), CouN5/CouN4 which is involved in the formation of the antibiotic coumermycin (*5*), EntB(ArCP)/EntE which is involved in the formation of the siderophore enterobactin (*24*), HMWP2 which is involved in the formation of the siderophore yersiniabactin (*27*), PchE/PchD which is involved in the biosynthesis of the siderophore pyochelin (*38*),

[1] Abbreviations: HPLC, high-pressure liquid chromatography; ESI, electrospray; FTMS, Fourier transform mass spectrometry; CoA, coenzyme A; NRPS, nonribosomal peptide synthetase; PKS, polyketide synthase; WT, wild-type; CNBr, cyanogen bromide; OCAD, octopole collisional activated dissociation; IRMPD, infrared multiphoton dissociation; SWIFT, stored waveform inverse Fourier transform; PP_i, pyrophosphate; TCEP, Tris(2-carboxyethyl)phosphine.

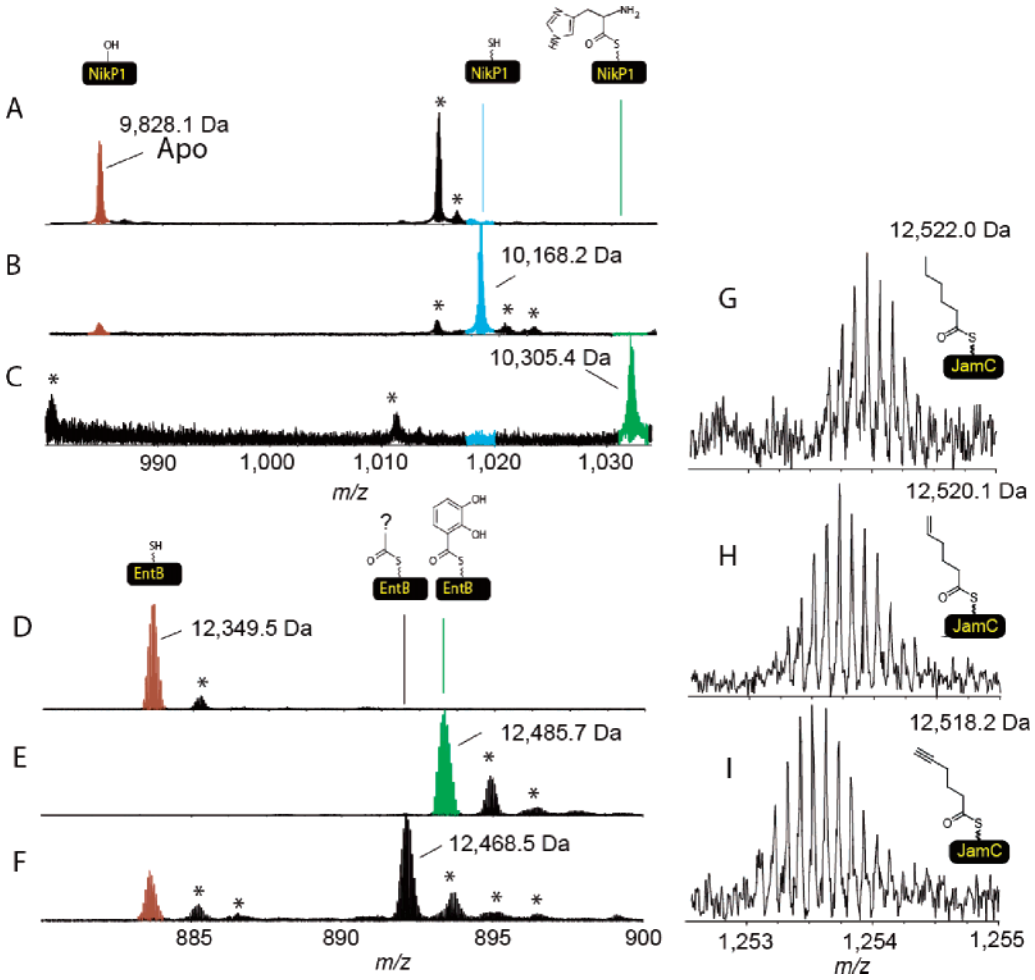

FIGURE 2: Identification of the preferred substrate by ESI-FTMS. (A) Apo form of the carrier site of NikP1. (B) Same as panel C but with L-histidine omitted. (C) Mass spectrum of the carrier site following incubation with CoA, Sfp, all 19 proteinogenic L-amino acids, glycine, L-selenocysteine, L-cystine, and 4-*trans*-hydroxy-L-proline. (D) Holo-EntB (ArCP). (E) Holo-EntB (ArCP) incubated with ATP and all 19 proteinogenic L-amino acids, glycine, L-selenocysteine, L-cystine, 4-*trans*-hydroxy-L-proline, and 2,3-dihydroxybenzoic acid. (F) Same as panel E but with 2,3-dihydroxybenzoic acid omitted. (G) Hexanoyl-S-JamC (only the ion with the N-terminal Met truncated is shown). (H) Hexenoyl-S-JamC. (I) Hexynoyl-S-JamC. An asterisk indicates salt adducts or non-active site peptides. The ions shown in panels A−C and G−I have a 10+ charge state. Those in panels D−F have a 14+ charge state.

JamC/JamA which is involved in the formation of the neurotoxin jamaicamide (*39*), and two NRPS modules from the orphan NRPS/PKS gene cluster on *B. subtilis* for which the product is unknown (*40*−*42*). The adenylation carrier didomains that are analyzed are from PksN (BG12652) and the second NRPS module from PksJ (BG10929, gene annotation from http://genolist.pasteur.fr/SubtiList/). All of the systems described above were incubated with substrate pools and analyzed by ESI-FTMS.

When PchE was presented with multiple substrates as well as the natural substrate salicylic acid and the activating enzyme PchD, the observed mass shift was 120 Da, in agreement with loading of salicylic acid (Table 1). When HMWP2 is incubated with six substrates known to independently load HMWP2 (*27*, *30*), the mass shift corresponding to the authentic substrate salicylic acid was the major species observed. When salicylic acid was omitted, mass shifts corresponding to methylsalicylic acid and benzoic acid are the main species observed. When CloN5/CloN4 or CouN5/CouN4 carrier protein/activating protein pairs are incubated with substrates not including the natural substrate L-proline, a mass shift of 113 Da, corresponding to the 4-*trans*-hydroxyproline, is observed. However, when proline is present, it is the only substrate loaded (+97 Da). In the case of EntB(ArCP)/EntE, a new species that is 119 Da larger than the holo form of EntB(ArCP) is formed when the authentic substrate is not present, but when the natural substrate is present, it is loaded exclusively (Table 1 and Figure 2E,F). This mass shift was confirmed to be 119 Da by tandem mass spectrometry (data not shown). Next, the early steps in the biosynthesis of jamaicamide were investigated. The apo-JamC protein showed two abundant forms of the protein, the WT and the post-translationally truncated form of the protein in which the N-terminal methionine has been removed (loss of 131 Da). Simultaneous incubation of holo-JamC, JamA, a substrate pool that included 6-bromo-5-hexynoic acid, 5-hexenoic acid, and ATP produced a mass shift that was 96 Da larger than that for holo-JamC, consistent with loading of hexenoic acid. Any attempts to load JamC with 6-bromo-5-hexynoic acid failed. 5-Hexanoic acid and 5-hexynoic acid are alternative substrates for this reaction (Figure 2G−I).

Even though as many as 24 substrates were screened simultaneously, the assay mixtures were still of a defined nature. To further probe the limits of the assay, CloN5/CloN4 and NikP1 were incubated with ATP and a commercially

Table 1: Identification of Substrate Loading from Defined Substrate Pools

system	digestion method	holo mass (calcd) (Da)	acylation substrate source	observed mass after acylation (major species only) (Da)	mass change (observed − $holo_{obs}$) (Da)	substrate name based on mass change (asterisk denotes a natural substrate)
nikkomycin	CNBr		*a*	10305.4	137.2	L-histidine*
NikP1		10168.2 (10168.3)	*b*	10168.2	0.0	N/A
clorobiocin	not digested		*a*	12616.9	96.8	L-proline*
CloN5/CloN4		12520.1 (12520.1)	*c*	12633.0	112.9	4-*trans*-hydroxyproline
coumermycin	not digested		*a*	12184.1	97.0	L-proline*
CouN5/CouN4		12087.1 (12087.1)	*c*	12200.2	113.1	4-*trans*-hydroxyproline
jamaicamide	not digested		*d*	12650.9	96.0	hexenoic acid*
JamC/JamA		12544.9 (12544.9)	*a*	12544.9	0.0	N/A
pyochelin	CNBr					
PchE/PchD		17653.3 (17653.8)	*e*	17773.7	120.4	salicylic acid*
yersiniabactin	trypsin		*f*	3531.73	120.08	salicylic acid*
HMWP2		3411.65 (3411.71)	*g*	3545.75	134.10	4-methylsalicylic acid
				3515.74	104.09	benzoic acid
enterobactin	not digested		*a*	12468.5	119.0	unknown
EntB(ArCP)/EntE		12349.5 (12349.3)	*h*	12485.7	136.2	2,3-dihydroxybenzoic acid*

[a] All 19 proteinogenic L-amino acids, glycine, L-selenocysteine, L-cystine, 4-*trans*-hydroxy-L-proline, and ATP. [b] Same as footnote a but with L-histidine omitted. [c] Same as footnote a but with L-proline omitted. [d] Same as footnote a but with 5-hexenoic acid and 6-bromo-5-hexynoic acid added. [e] Same as footnote a with salicylic acid added. [f] Salicylic acid, 4-methylsalicylic acid, benzoic acid, methylsalicylic acid, benzoic acid, 4-hydroxybenzoic acid, and *p*-toluic acid. [g] Same as footnote f but with salicylic acid omitted. [h] Same as footnote a but with 2,3-dihydroxybenzoate added.

Table 2: Identification of Substrate Loading from Undefined Substrate Pools

system	digestion method	acylation substrate source	new mass after acylation (Da)	mass change (observed − $holo_{calc}$) (Da)	substrate name based on mass change (asterisk denotes a natural substrate)
NikP1	CNBr	algal hydrolysate	10305.5	137.2	L-histidine*
		E. coli metabolome	10305.7	137.4	L-histidine*
CloN5/CloN4	not digested	algal hydrolysate	12617.1	97.0	L-proline*
		E. coli metabolome	12617.2	97.1	L-proline*
CouN5/CouN4	not digested	*E. coli* metabolome	12184.1	97.0	L-proline*
EntB(ArCP)/EntE	not digested	*E. coli* metabolome	NC[a]	NC[a]	not loaded
		E. coli metabolome[b]	NC[a]	NC[a]	not loaded

[a] No mass change observed. [b] Metabolome obtained from *E. coli* grown under iron-limiting conditions.

available algal amino acid hydrolysate mixture as a representative undefined substrate pool. In both cases, the correct mass shift corresponding to the natural substrate is observed (Table 2). This suggested that very complex mixtures and undefined assay conditions are possible. To further expand on this theme, a very complex substrate pool was tested, the *Escherichia coli* metabolome, on NikP1, CloN5/CloN4, CouN5/CouN4, and EntB (ArCP)/EntE. In three of the four cases, even though there is only partial acylation occupancy, the correct mass shift corresponding to the authentic substrate is observed. Loading onto holo-EntB (ArCP) was not observed, even when the metabolome is obtained from *E. coli* which was grown under iron-limited conditions, suggesting that there is not enough free substrate available for observation of the acylation reaction (Table 2).

To apply our method to orphan NRPS/PKS gene clusters, the adenylation and carrier didomains from PksN and PksJ were cloned and overproduced. Because the didomains are rather large (~90 kDa) to observe by FTMS with sub-dalton mass accuracy, the holo forms of the didomains were subjected to trypsin digestion and the active sites mapped by FTMS (Figure 4). The identities of the active sites were confirmed by tandem mass spectrometry as shown in Figure 4J for the PksJ didomain and in Figure 4K for the PksN didomain (*43*). Once mapped, the active sites were pantetheinylated using Sfp and CoA and compared to the apo form. In both cases, a mass shift of 340 Da is observed (Figure 4B,F). Subsequently, both of the proteins were incubated with ATP and an undefined substrate pool, the algal hydrolysate (Figure 4C,G). The PksJ domain increased in mass by 57 Da, while the PksN domain increased in mass by 71 Da. Incubation of PksJ with glycine and ATP resulted in the same mass shift (Figure 4D). Because the mass shift of PksN did not correspond to the predicted substrates cysteine and/or serine, alanine was incubated simultaneously with cysteine, serine, and threonine. Again the observed mass shift was 71 Da (Figure 4H). When the substrate pool consisted of only threonine and serine, a mass shift of 87 Da was observed (Figure 4I), while no acylation of the carrier domain was observed when the it was incubated with cysteine.

DISCUSSION

This paper has presented yet another tool in the arsenal for the in vitro characterization of the NRPS and PKS substrate specificity. ESI-FTMS has been used in conjunction with substrate pools to screen for substrates that can be loaded onto carrier domains. Although the assay is performed on an FTMS instrument, substrate screening can likely be done on other mass spectrometers as well, but one will have to keep in mind that the ease in mapping the active sites and the accuracy of the observed mass shifts increase with

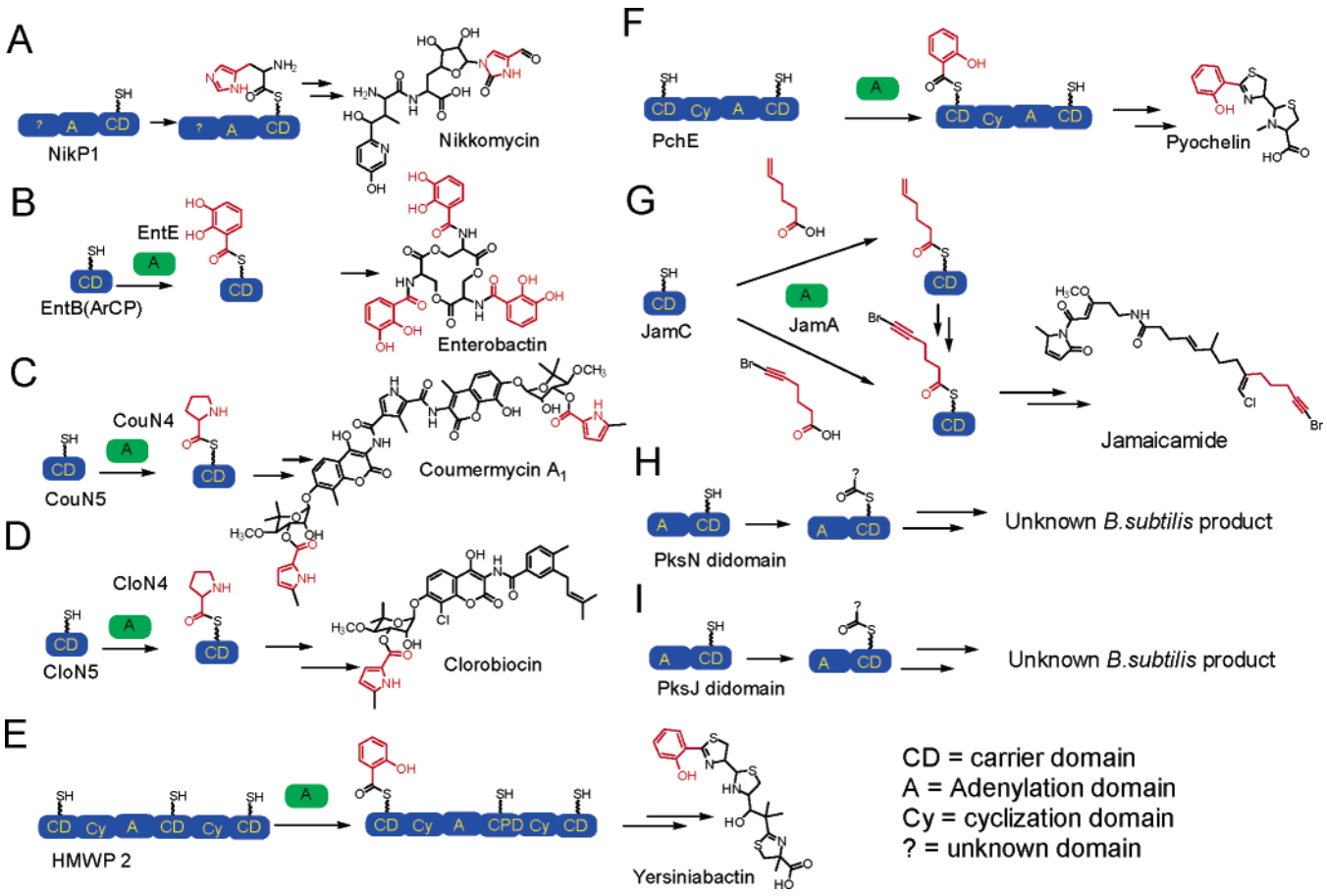

FIGURE 3: Systems interrogated in this study.

increasing resolution. Currently, to look at protein domains, FTMS, which requires some expertise, is the most accurate method for this. The custom-built instrument used in this study has a mass accuracy of 5−25 ppm when it is externally calibrated, while some commercial FTMS instruments now give mass accuracies to within 2 ppm. Analyzing substrates loaded onto carrier domains will be very difficult when the mass accuracy is lower than 100 ppm but has been done before (*34*−*36*). But as there are increasingly more FTMS instruments that are accessible to the scientific community, the implemention of our approach described here is likely to find wider application in the analysis of NRPS systems.

The generality of this method has many applications, and a few of these are demonstrated in this paper. The first application of the substrate pool is to identify the natural substrate. For the proof of concept experiments, NikP1, EntB, HMWP2, CouN5, CloN5, and PchE were loaded with their natural substrate. In each case that was investigated, as long as the natural substrate was present, substrate loading by noncognate substrates was negligible. In the absence of the natural substrate, there was one unexplained result where EntB (ArCP) displayed a 119 Da species and none of the 98−99% pure amino acids used as a substrate matched this mass upon acylation. Therefore, this is an unidentified noncognate substrate, which must come from the remaining 1−2% of the impurities found in commercially prepared amino acids. This is also in agreement with the observation that there was insufficient material for complete conversion of the holo form of the protein to the acylated species. When the authentic natural substrate, 2,3-dihydroxybenzoic acid, was present with this same substrate pool, it was the major peak observed in the spectrum (Figure 2E). The identification of substrates is also possible from very complex substrate pools of undefined nature such as the algal hydrolysate or the *E. coli* metabolome.

Sometimes loading with a noncognate substrate is desired, as would be the case in the development of new bioactive compounds. Replacing the pyrrole on clorobiocin or coumermycin may allow for the development of more soluble aminocoumarin antibiotics and therefore make them more applicable in a clinical setting (*23*). It was unknown if anything other than proline would load onto CloN5 or CouN5. When a substrate screen was done with 19 proteinogenic L-amino acids, glycine, L-selenocysteine, L-cystine, and 4-*trans*-hydroxy-L-proline but the native substrate, proline, was omitted, a mass shift of 113 Da was observed. This mass shift corresponds to the mass of 4-*trans*-hydroxy-L-proline being loaded. Subsequently, it was verified by the traditional radioactive pyrophosphate exchange assay that adenylating enzymes CloN4 and CouN4 could utilize 4-*trans*-hydroxy-L-proline as an alternative substrate (*5*). This makes 4-*trans*-hydroxy-L-proline a front-line candidate for generation of more soluble clorobiocin and coumermycin antibiotics and was identified from a single assay.

The third utility of this FTMS-based assay is to establish the timing of tailoring events which take place on NRPS and PKS modules. In the jamaicamide biosynthetic pathway, it was unknown if the bromination took place before or after the loading of hexenoic acid onto JamC. Simultaneous

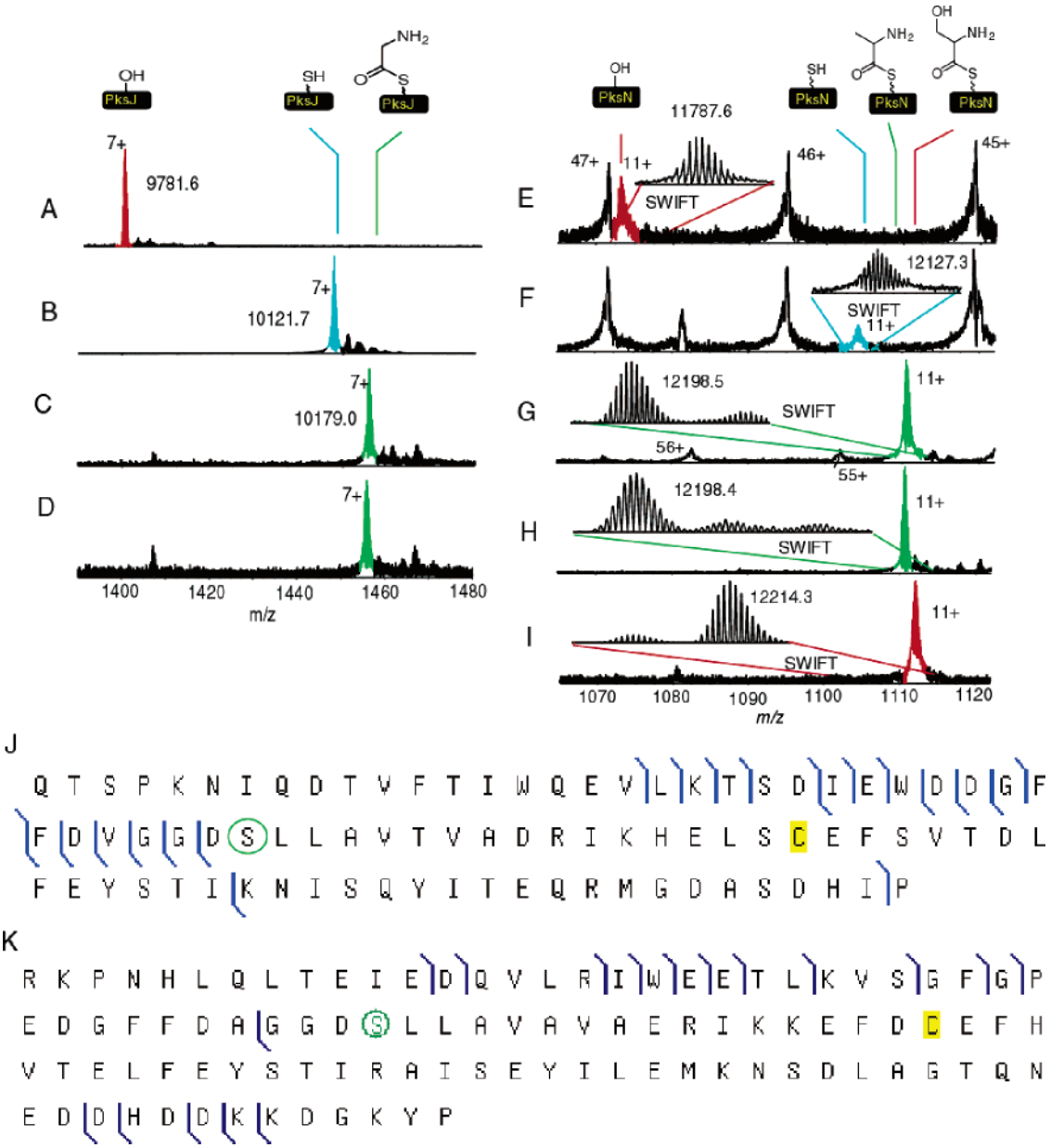

FIGURE 4: Active site mapping and substrate pool screening of the adenylation and carrier didomains from PksN and PksJ. (A) Apo-PksJ. (B) Holo-PksJ. (C) Algal hydrolysate substrate screen with holo-PksJ. (D) Glycyl-S-PksJ. (E) Apo-PksN. (F) Holo-PksN. (G) Algal hydrolysate substrate screen with holo-PksN. (H) Ala, Cyst, Ser, Thr screen with holo-PksN. (I) Ser, Thr screen with holo-PksN. (J) OCAD verification of the active site for holo-PksJ. (K) IRMPD verification of the active site for holo-PksN. The green circle indicates the location of the phosphopantetheinyl functionality.

incubation of holo-JamC, JamA, 6-bromo-5-hexynoic acid, 5-hexenoic acid, and ATP resulted in the loading of only hexenoic acid. Any attempt to load 6-bromo-5-hexynoic acid by itself failed as well. This means that the bromination reaction either takes place while the substrate is attached to JamC or otherwise on one of the other NRPS or PKS proteins found on the jamaicamide biosynthetic pathway or after the jamaicamide is fully assembled. 5-Hexanoic acid and 5-hexynoic acid were also alternative substrates, in agreement with the classical pyrophosphate exchange assay (*39*).

The fourth application is to use the substrate pool method as an activity screen. This is particulary relevant when the substrate cannot be readily predicted by bioinformatic means. One example is the amine donor source for the aminotransferase domain of MycA, a protein responsible for the generation of the β-amino acid found on the mycosubtilin biosynthetic pathway. Because the substrate amine donor was unknown, all possible substrate candidates were screened simultaneously to see if the protein was active. Armed with its activity, Gln was ultimately identified as the preferred amine donor in this reaction (*33*).

This method will be most useful in the characterization of gene clusters with unknown natural products (orphan gene clusters). As an example, *B. subtilis* has one orphan NRPS/PKS gene cluster (*40*, *41*). Even though this cluster has both NRPS and PKS modules, it has been suggested to be involved in the formation of difficidin (*20*, *41*). Difficidin, however, does not have an amino acid in its structure, and there is little experimental evidence to support the suggestion that difficidin is produced by this gene cluster (*41*). Therefore, the activity screen was used to establish if the NRPS domains of this cluster are functional. When activity was

observed, it was used to identify which substrates were loaded. Using bioinformatics, the substrate specificity for the second NRPS domain on PksJ was predicted to load glycine while PksN was predicted to load cysteine (*16*, *17*), and BLAST analysis indicated that the adenylation domain may also be similar to serine loading domains. Once the active sites had been mapped, they were pantetheinylated using Sfp and CoA. Subsequently, both of the proteins were incubated with ATP and an unbiased substrate source, the algal hydrolysate. The PksJ domain increased in mass by 57 Da, while the PksN domain increased in mass by 71 Da (Figure 4D,G). These mass shifts correspond to glycine and alanine, respectively. Incubation of PksJ with glycine and ATP resulted in the same mass shift. Because the mass shift of PksN did not correspond to the predicted substrate cysteine or serine, L-alanine was incubated simultaneously with L-cysteine, L-serine, and L-threonine (this substrate served as a negative control). Again, the only substrate that was loaded was alanine. When L-alanine was omitted, L-serine was a substrate for this protein. However, because alanine outcompetes both cysteine and serine, the predicted substrates, we favor alanine as the natural substrate. Having observed activity for the PksJ and PksN didomains provides some insight into the role of this gene cluster. The first important observation is that the NRPSs found on the orphan gene cluster of *B. subtilis* are active. Second, because we have observed activity, our data suggest that this gene cluster produces a product other than difficidin, since difficidin does not have a nitrogen in its structure that would have come from glycine or alanine. We cannot exclude the possibility that this gene cluster produces a modified form of difficidin or natural products analogous to difficidin such as hydroxymycotrienin A produced by *Bacillus* sp. BMJ958-62F4 (*44*), which has amino acid components in its structure. Last, this is the first demonstration of the utility of FTMS in providing insight into the functions of an orphan gene cluster and is going to be applicable to many other orphan NRPS gene clusters that have been sequenced. This method is applicable not only to NRPS systems but also to systems of polyketide origin, fatty acid biosynthesis, or any other system that can form covalent modifications to proteins.

MATERIALS AND METHODS

The sequencing grade trypsin was purchased from Promega, and 1 unit is defined as the amount of sequencing grade modified trypsin required to produce a ΔA_{253} of 0.001/min at 30 °C with the substrate α-benzoyl-L-arginine ethyl ester. Cyanogen bromide (CNBr), amino acids, HPLC solvents, CoA (trilithium salt), algal hydrolysate, and all other substrates were purchased from Sigma-Aldrich. 6-Bromo-5-alkynoic acid was synthesized as described previously (*45*). Superflow nickel affinity resin was from Qiagen. PD10 gel filtration columns were obtained from Amersham Biosciences. The HPLC column used for all desalting steps and separations a Jupiter 5 μm C4 300 Å column from Phenomenex was used. The freeware PAWS was obtained from Proteometrics.

Construction of JamC, PksJ, and PksN NRPS Constructs. The JamC gene was cloned from the pJam1 fosmid clone (*39*) using PfuI polymerase and the following primers: forward primer, 5′ CATGCCATGGAAAACTTAACCGTAG 3′; reverse primer, 5′ CCGCTCGAGTGCACCAAAGTGCTCTGC 3′. The resulting PCR product was cloned in frame with the carboxy six-His fusion at the NcoI and XhoI restriction sites of pET28a. The pksN and pksJ didomains were cloned in a similar fashion from *B. subtilis* strain NCIB 3610 (*49*). The pksJ-AT2 didomain was amplified using the following primers: forward, 5′ AGCTAGCTTTGAACTGTGGGAAACAGA 3′; and reverse, 5′ ACTCGAGTCATTTTTGCAATGTCCATAATCC 3′. The amplified fragment was digested with NheI and XhoI and cloned into pET28a to generate plasmid pPDS0372. The pksN-AT didomain was amplified using the following primers: forward, 5′ GAATTCACATATGGGCTTGCAAAAAGTGCTTG 3′; and reverse, 5′ ACTCGAGTCACGGGTATTTTCCATCTTTTTTG 3′. The amplified fragment was digested with NdeI and XhoI and cloned into pET28a to generate plasmid pPDS0374.

Protein Expression and Protein Purification. The proteins NikP1, HMWP2, PchE, PchD, JamA, CloN5, CloN4, CouN5, CouN4, EntB(ArCP), and EntE were purified as described previously (*5*, *26*, *27*, *30*, *39*). An expression strain [*E. coli* BL21(DE3) star transformed with a plasmid encoding JamC] was incubated at 37 °C until the OD_{600} reached 0.7. At this point, IPTG was added (50 mg/L) and the mixture allowed to incubate for 4−6 h at 28 °C. The cells were then harvested by centrifugation and lysed by the addition of lysozyme and sonication. The insoluble materials were pelleted by centrifugation, and the remaining supernatant was loaded onto a column containing NTA Superflow nickel affinity resin and the protein purified by following the instructions of the manufacturer (Qiagen). The overproduction and purification of *B. subtilis* PksJ and PksN didomains were carried out in a similar fashion. The purified proteins were buffer exchanged using PD-10 gel filtration columns that were equilibrated with 50 mM Tris buffer (pH 7.5) with 1 mM TCEP. Stock solutions containing 10% glycerol were prepared and stored at −80 °C.

Preparation of E. coli Metabolomes. E. coli was grown in 150 mL of Luria Broth (LB) at 37 °C to an OD_{600} of 0.6−0.8. The cells were harvested by centrifugation in a Sorvall RC-5C+ centrifuge (SLA-3000 rotor, 6000 rpm) at 4 °C for 6 min. The pellet was resuspended in 2.5 mL of 25 mM Tris (pH 7.6), and the cells were lysed by sonication in the presence of lysozyme. The lysate was clarified by centrifugation in a Sorvall RC-C5+ centrifuge (SS-34, 16 000 rpm, 4 °C for 25 min). Approximately 1 mL of lysed crude extract (CE) was gel filtered by using a PD10 column equilibrated with 25 mM Tris (pH 7.6). This was achieved by loading the column with 1.0 mL of CE, and it was allowed to flow through. Then 5.0 mL of 25 mM Tris-HCl (pH 7.6) was added and again allowed to flow through to remove all the proteins. After this volume, the small molecules of the lysed extract were collected in the following 3.0 mL. This 3.0 mL was collected as 1 mL fractions, frozen at −80 °C, and lyophilized. The dried sample was resuspended in 50−100 μL of 50 mM Tris (pH 7.6) so it could be used in the substrate identification studies.

Acylation of CloN5 Using Undefined Metabolite Pools. As a representative acylation, we describe the acylation of CloN5; all other assays were carried out in a similar fashion except that the substrate pool was varied. The acylation of CloN5 was carried out in two steps. First, the apo form of CloN5 at a concentration of 25 μM was reacted for 1 h at

Table 3: HPLC Gradient for Active Site Purification, where Solvent A Is Water (0.1% TFA) and Solvent B Is Acetonitrile (ACN with 0.1% TFA)[a]

	0.00 min	10.0 min	15.0 min	55.0 min	60.0 min	60.1 min	60.2 min	62.6 min	63.0 min	65.0 min	66.0 min
% A	90	90	70	30	10	10	95	95	5	5	90
% B	10	10	30	70	90	90	5	5	95	95	10

[a] The purpose of the unusual gradient above 60 min is to wash the column prior to the next injection to avoid contamination from a previous run.

room temperature in the presence of 3.6 μM Sfp, 8 mM $MgCl_2$, and 250 μM CoA in a total reaction volume of 100 μL, which resulted in the generation of the holo form of the enzyme. In the second step, 4 μL of 0.1 M ATP, 2 μL of 1.3 mg/mL CloN4, and 10−50 μL of an *E. coli* metabolome were added to the holoenzyme generated in the first step. This was incubated at room temperature for 30 min before the reaction was quenched in a 1:1 (v/v) manner with 10% formic acid. This was repeated for each (five total) *E. coli* metabolome. Other substrate pools included the defined mixtures of substrates which were present at 0.5−1 mM (often some precipitation would be observed) and the algal hydrolysate mixture present at 0.2−0.5 mg/mL.

Acylation and CNBr Digestion of NikP1. The acylation of NikP1 was also carried out in two steps. The generation of the holo form of the enzyme was performed as described for CloN5. In the second step, 4 μL of 0.1 M ATP and 5 μL of an *E. coli* metabolome were added to the solution. This was incubated at room temperature for 30 min before the reaction was stopped by quenching it in a 1:1 (v/v) manner with 10% formic acid. The acylated forms of NikP1 were purified by HPLC on an HP1090M HPLC system using a 30 min gradient of 90% water with 0.1% TFA and 10% ACN (0.1% TFA) to 5% water (0.1% TFA) and 95% ACN (0.1% TFA), collecting peaks between 21.5 and 23.0 min (this additional purification step of NikP1, before CNBr digestion, was done because the metabolome introduced contamination that overlapped with the active site after digestion; this step was not necessary when defined substrate sources were used in the substrate screen). The resulting samples were frozen at −80 °C and lyophilized overnight. The samples were redissolved in 100 mM NH_4Oac (pH 4), 6 M urea, 10% CH_3CN, and 10 mM TCEP and digested with 1.0 M CNBr in ACN (reaction volume of 400 μL) for 12−18 h in the dark. Each sample was then frozen at −80 °C, lyophilized overnight in the dark, and rechromatographed using the HPLC conditions described above, and the fractions eluting at 18−20 min were frozen, lyophilized, and analyzed by FTMS. As with CloN5, each *E. coli* metabolome was tested as well as a control containing 22 amino acids. PchE was digested using CNBr using an identical protocol as described for HMWP2 (*30*), and PksJ and PksN didomains were digested using trypsin in an identical fashion as described for MycA on the mycosubtilin biosynthetic pathway (*33*). All other proteins did not need digestion to be seen via ESI-FTMS.

Peptide Mapping of the PksJ and PksN Constructs. Each didomain of PksJ and PksN 1 mg/mL was incubated with 3.6 μM Sfp, 8 mM $MgCl_2$, and 320 μM coenzyme A for 1 h (reaction volume of 300 μL). The final product was digested with trypsin at pH 8.0 (85 units) for 10 min, and the reaction was quenced by acidifying the mixture with 50 μL of 10% formic acid. The reaction mixture was purified by HPLC on an HP1100 HPLC system using a 60 min gradient (Table 3) collecting fractions at 1 min intervals. A peak list of all the observed ions was generated using THRASH and/or manual deconvolution of the charge states. The generated peak list was imported into the freeware PAWS to identify the active sites. Once a match to the active site was obtained, the active sites were subjected to OCAD or IRMPD (*46*). The resulting fragment ions from OCAD and IRMPD were then analyzed using ProSight PTM to verify that they were indeed the active sites (*47*).

MS Analysis. HPLC fractions containing the active sites prepared as described above were redissolved in 100 μL of 78% ACN, 0.1% acetic acid or 49% methanol, and 1% formic acid and analyzed by ESI-FTMS. For mass spectrometric analysis, a custom 8.5 T ESI-FTMS mass spectrometer was used which was equipped with a front-end quadrupole (*46*). The samples were introduced into the FTMS instrument using a NanoMate 100 apparatus for automated nanospray (Advion Biosciences, Ithaca, NY). Typically, 500 ms ion accumulation per scan was used, and 50−200 scans were acquired per spectrum. The instrument was externally calibrated using ubiquitin, with monoisotopic M_r value of 8560.65 Da (Sigma). For the calculation of the masses of the proteins, the MIDAS analysis datastation was used (*48*). A mass peak list of all the observed ions was generated using THRASH (embedded in the MIDAS datastation) and/or manual deconvolution. All masses reported in this paper are reported as the neutral monoisotopic masses.

REFERENCES

1. Paulsen, I. T., Press, C. M., Ravel, J., Kobayashi, D. Y., Myers, G. S. A., Mavrodi, D. V., DeBoy, R. T., Seshadri, R., Ren, Q., Madupu, R., Dodson, R. J., Durkin, A. S., Brinkac, L. M., Daugherty, S. C., Sullivan, S. A., Rosovitz, M. J., Gwinn, M. L., Zhou, L., Schneider, D. J., Cartinhour, S. W., Nelson, W. C., Weidman, J., Watkins, K., Tran, K., Khouri, H., Pierson, E. A., Pierson, L. S., Thomashow, L. S., and Loper, J. E. (2005) Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5, *Nat. Biotechnol. 23* (7), 873−878.
2. Lautru, S., Deeth, R. J., Bailey, L. M., and Challis, G. L. (2005) Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining, *Nat. Chem. Biol. 1* (5), 265−269.
3. Liu, W., Nonaka, K., Nie, L., Zhang, J., Christenson, S. D., Bae, J., Van Lanen, S. G., Zazopoulos, E., Farnet, C. M., Yang, C. F., and Shen, B. (2005) The neocarzinostatin biosynthetic gene cluster from *Streptomyces carzinostaticus* ATCC 15944 involving two iterative type I polyketide synthases, *Chem. Biol. 12* (3), 293−302.
4. Rondon, M. R., Ballering, K. S., and Thomas, M. G. (2004) Identification and analysis of a siderophore biosynthetic gene cluster from *Agrobacterium tumefaciens* C58, *Microbiology 150* (11), 3857−3866.
5. Garneau, S., Dorrestein, P. C., Kelleher, N. L., and Walsh, C. T. (2005) Characterization of the formation of the pyrrole moiety during clorobiocin and coumermycin A1 biosynthesis, *Biochemistry 44* (8), 2770−2780.
6. Chen, H., Hubbard, B. K., O'Connor, S. E., and Walsh, C. T. (2002) Formation of β-hydroxy histidine in the biosynthesis of nikkomycin antibiotics, *Chem. Biol. 9* (1), 103−112.

7. Walsh, C. T. (2004) Polyketide and nonribosomal peptide antibiotics: Modularity and versatility, *Science 303* (5665), 1805−1810.
8. Finking, R., and Marahiel, M. A. (2004) Biosynthesis of nonribosomal peptides, *Annu. Rev. Microbiol. 58*, 453−488.
9. Smith, E. E., Sims, E. H., Spencer, D. H., Kaul, R., and Olson, M. V. (2005) Evidence for diversifying selection at the pyoverdine locus of *Pseudomonas aeruginosa*, *J. Bacteriol. 187* (6), 2138−2147.
10. Keszenman-Pereyra, D., Lawrence, S., Twfieg, M.-E., Price, J., and Turner, G. (2003) The npgA/cfwA gene encodes a putative 4′-phosphopantetheinyl transferase which is essential for penicillin biosynthesis in *Aspergillus nidulans*, *Curr. Genet. 43* (3), 186−190.
11. Zerbe, K., Woithe, K., Li, D. B., Vitali, F., Bigler, L., and Robinson, J. A. (2004) An oxidative phenol coupling reaction catalyzed by OxyB, a cytochrome P450 from the vancomycin-producing microorganism, *Angew. Chem., Int. Ed. 43* (48), 6709−6713.
12. Vater, J., and Stein, T. H. (1999) Structure, function, and biosynthesis of gramicidin S synthetase, *Compr. Nat. Prod. Chem. 4*, 319−352.
13. Ahlert, J., Shepard, E., Lomovskaya, N., Zazopoulos, E., Staffa, A., Bachmann, B. O., Huang, K., Fonstein, L., Czisny, A., Whitwam, R. E., Farnet, C. M., and Thorson, J. S. (2002) The calicheamicin gene cluster and its iterative type I enediyne PKS, *Science 297* (5584), 1173−1176.
14. Walsh, C. T. (2005) Natural insights for chemical biologists, *Nat. Chem. Biol. 1* (3), 122−124.
15. Fortman, J. L., and Sherman, D. H. (2005) Utilizing the power of microbial genetics to bridge the gap between the promise and the application of marine natural products, *ChemBioChem 6* (6), 960−978.
16. Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases, *Chem. Biol. 6* (8), 493−505.
17. 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* (3), 211−224.
18. Di Vincenzo, L., Grgurina, I., and Pascarella, S. (2005) In silico analysis of the adenylation domains of the freestanding enzymes belonging to the eucaryotic nonribosomal peptide synthetase-like family, *FEBS J. 272* (4), 929−941.
19. Van Lanen, S. G., Dorrestein, P. C., Christenson, S. D., Liu, W., Ju, J., Kelleher, N. L., and Shen, B. (2005) Biosynthesis of the β-amino acid moiety of the enediyne antitumor antibiotic C-1027 featuring β-amino acyl-S-carrier protein intermediates, *J. Am. Chem. Soc. 127* (33), 11594−11595.
20. Chang, Z., Sitachitta, N., Rossi, J. V., Roberts, M. A., Flatt, P. M., Jia, J., Sherman, D. H., and Gerwick, W. H. (2004) Biosynthetic pathway and gene cluster analysis of Curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*, *J. Nat. Prod. 67* (8), 1356−1367.
21. Kohli, R. M., Walsh, C. T., and Burkart, M. D. (2002) Biomimetic synthesis and optimization of cyclic peptide antibiotics, *Nature 418* (6898), 658−661.
22. Trauger, J. W., Kohli, R. M., Mootz, H. D., Marahiel, M. A., and Walsh, C. T. (2000) Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase, *Nature 407* (6801), 215−218.
23. Galm, U., Dessoy, M. A., Schmidt, J., Wessjohann, L. A., and Heide, L. (2004) In vitro and in vivo production of new aminocoumarins by a combined biochemical, genetic, and synthetic approach, *Chem. Biol. 11* (2), 173−183.
24. Eustaquio, A. S., Gust, B., Galm, U., Li, S.-M., Chater, K. F., and Heide, L. (2005) Heterologous expression of novobiocin and clorobiocin biosynthetic gene clusters, *Appl. Environ. Microbiol. 71* (5), 2452−2459.
25. Freitag, A., Galm, U., Li, S.-M., and Heide, L. (2004) New aminocoumarin antibiotics from a cloQ-defective mutant of the clorobiocin producer *Streptomyces roseochromogenes* DS12.976, *J. Antibiot. 57* (3), 205−209.
26. Shaw-Reid, C. A., Kelleher, N. L., Losey, H. C., Gehring, A. M., Berg, C., and Walsh, C. T. (1999) Assembly line enzymology by multimodular nonribosomal peptide synthetases: The thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization, *Chem. Biol. 6* (6), 385−400.
27. Mazur, M. T., Walsh, C. T., and Kelleher, N. L. (2003) Site-specific observation of acyl intermediate processing in thiotemplate biosynthesis by Fourier transform mass spectrometry: The polyketide module of yersiniabactin synthetase, *Biochemistry 42* (46), 13393−13400.
28. Hicks, L., Weinreb, P., Konz, D., Marahiel, M. A., Walsh, C. T., and Kelleher, N. L. (2003) Fourier-transform mass spectrometry for detection of thioester-bound intermediates in unfractionated proteolytic mixtures of 80 and 191 kDa portions of Bacitracin A synthetase, *Anal. Chim. Acta 496* (1−2), 217−224.
29. Hicks, L. M., O'Connor, S. E., Mazur, M. T., Walsh, C. T., and Kelleher, N. L. (2004) Mass spectrometric interrogation of thioester-bound intermediates in the initial stages of epothilone biosynthesis, *Chem. Biol. 11* (3), 327−335.
30. McLoughlin, S. M., and Kelleher, N. L. (2004) Kinetic and regiospecific interrogation of covalent intermediates in the nonribosomal peptide synthesis of yersiniabactin, *J. Am. Chem. Soc. 126* (41), 13265−13275.
31. Gatto, G. J., Jr., McLoughlin, S. M., Kelleher, N. L., and Walsh, C. T. (2005) Elucidating the substrate specificity and condensation domain activity of FkbP, the FK520 pipecolate-incorporating enzyme, *Biochemistry 44* (16), 5993−6002.
32. Dorrestein, P. C., Yeh, E., Garneau-Tsodikova, S., Kelleher, N. L., and Walsh, C. T. (2005) Dichlorination of a pyrrolyl-S-carrier protein by $FADH_2$-dependent halogenase PltA during pyoluteorin biosynthesis, *Proc. Natl. Acad. Sci. U.S.A. 102* (39), 13843−13848.
33. Aron, Z. D., Dorrestein, P. C., Blackhall, J. R., Kelleher, N. L., and Walsh, C. T. (2005) Characterization of a new tailoring domain in polyketide biogenesis: The amine transferase domain of MycA in the mycosubtilin gene cluster, *J. Am. Chem. Soc. 127* (43), 14986−14987.
34. Hong, H., Appleyard, A. N., Siskos, A. P., Garcia-Bernardo, J., Staunton, J., and Leadlay, P. F. (2005) Chain initiation on type I modular polyketide synthases revealed by limited proteolysis and ion-trap mass spectrometry, *FEBS J. 272* (10), 2373−2387.
35. Schnarr, N. A., Chen, A. Y., Cane, D. E., and Khosla, C. (2005) Analysis of covalently bound polyketide intermediates on 6-deoxyerythronolide B synthase by tandem proteolysis-mass spectrometry, *Biochemistry 44* (35), 11836−11842.
36. Stein, T., Vater, J., Kruft, V., Otto, A., Wittmann-Liebold, B., Franke, P., Panico, M., McDowell, R., and Morris, H. R. (1996) The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates, *J. Biol. Chem. 271* (26), 15428−15435.
37. Reuter, K., Mofid, M. R., Marahiel, M. A., and Ficner, R. (1999) Crystal structure of the surfactin synthetase-activating enzyme Sfp: A prototype of the 4′-phosphopantetheinyl transferase superfamily, *EMBO J. 18* (23), 6823−6831.
38. Quadri, L. E. N., Keating, T. A., Patel, H. M., and Walsh, C. T. (1999) Assembly of the *Pseudomonas aeruginosa* nonribosomal peptide siderophore pyochelin: In vitro reconstitution of aryl-4,2-bisthiazoline synthetase activity from PchD, PchE, and PchF, *Biochemistry 38* (45), 14941−14954.
39. Edwards, D. J., Marquez, B. L., Nogle, L. M., McPhail, K., Goeger, D. E., Roberts, M. A., and Gerwick, W. H. (2004) Structure and biosynthesis of the Jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*, *Chem. Biol. 11* (6), 817−833.
40. Kunst, F., Ogasawara, N., Moszer, I., et al. (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*, *Nature 390*, 249−256.
41. Hofemeister, J., Conrad, B., Adler, B., Hofemeister, B., Feesche, J., Kucheryava, N., Steinborn, G., Franke, P., Grammel, N., Zwintscher, A., Leenders, F., Hitzeroth, G., and Vater, J. (2004) Genetic analysis of the biosynthesis of non-ribosomal peptide- and polyketide-like antibiotics, iron uptake and biofilm formation by *Bacillus subtilis* A1/3, *Mol. Genet. Genomics 272* (4), 363−378.
42. Stein, T. (2005) *Bacillus subtilis* antibiotics: Structures, syntheses and specific functions, *Mol. Microbiol. 56* (4), 845−857.
43. Marshall, A. G., Hendrickson, C. L., and Jackson, G. S. (1998) Fourier transform ion cyclotron resonance mass spectrometry: A primer, *Mass Spectrom. Rev. 17* (1), 1−35.
44. Hosokowa, N., Naganawa, H., Hamada, M., Takeuchi, T., Ikeno, S., and Hori, M. (1996) Hydroxymycotrienins A and B, new ansamycin group antibiotics, *J. Antibiot. 49* (5), 425−431.
45. Gung, B. W., and Dickson, H. (2002) Total synthesis of (−)-Minquartynoic Acid: An anti-cancer, anti-HIV natural product, *Org. Lett. 4* (15), 2517−2519.
46. Patrie, S. M., Charlebois, J. P., Whipple, D., Kelleher, N. L., Hendrickson, C. L., Quinn, J. P., Marshall, A. G., and Mukho-

padhyay, B. (2004) Construction of a hybrid quadrupole/Fourier transform ion cyclotron resonance mass spectrometer for versatile MS/MS above 10 kDa, *J. Am. Soc. Mass Spectrom. 15*, 1099−1108.

47. LeDuc, R. D., Taylor, G. K., Kim, Y.-B., Januszyk, T. E., Bynum, L. H., Sola, J. V., Garavelli, J. S., and Kelleher, N. L. (2004) ProSight PTM: An integrated environment for protein identification and characterization by top-down mass spectrometry, *Nucleic Acids Res. 32*, W340−W345.
48. Senko, M. W., Canterbury, J. D., Guan, S., and Marshall, A. G. (1996) A high-performance modular data system for Fourier transform ion cyclotron resonance mass spectrometry, *Rapid Commun. Mass Spectrom. 10*, 1839−1844.
49. Branda, S. S., Gonzalez-Pastor, J. E., Ben-Yehuda, S., Losick, R., and Kolter, R. (2001) Fruiting body formation by *Bacillus subtilis*, *Proc. Natl. Acad. Sci. U.S.A. 98*, 11621−11626.

BI052333K