ARTICLE

ACS chemical biology

Structural Characterization of *in Vitro* and *in Vivo* Intermediates on the Loading Module of Microcystin Synthetase

Leslie M. Hicks[†,¶], Michelle C. Moffitt[‡,¶], Laura L. Beer[‡], Bradley S. Moore[‡,§,*], and Neil L. Kelleher[†,*]

[†]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, [‡]Departments of Pharmacology & Toxicology, University of Arizona, Tucson, Arizona 85721, and [§]Scripps Institution of Oceanography and Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093

ABSTRACT The microcystin family of toxins is the most common cause of hepatotoxicity associated with water blooms of cyanobacterial genera. The biosynthetic assembly line producing the toxic cyclic peptide, microcystin, contains an adenylation–peptidyl carrier protein didomain (A–PCP) at the N-terminus of the initiator module McyG (295 kDa) that has been postulated to activate and load the starter unit phenylacetate for formation of the unusual aromatic β-amino acid residue, Adda, before subsequent extension. Characterization of the McyG A–PCP didomain (78 kDa) using ATP–PP_i exchange assays and mass spectrometry revealed that assorted phenylpropanoids are preferentially activated and loaded onto the PCP carrier domain rather than phenylacetate itself. For the first time, thioesters formed *in vivo* were detected directly using large molecule mass spectrometry. Additionally substrates were cleaved using a type II thioesterase for structural elucidation by small molecule mass spectrometry. Unprecedented features of the McyG A–PCP didomain include the *in vivo* acylation of the *holo* PCP with exogenous and endogenous substrates, along with the ability of the *apo* protein to retain the acyl-AMP intermediate during affinity purification. These results imply that phenylpropanoids are preferentially loaded onto the McyG PCP; however one carbon must be excised following extension of the starter unit with malonyl-CoA in order to generate the expected polyketide chain which leads us to ponder the novel biochemistry by which this occurs.

Toxic cyanobacterial waterblooms are of serious concern due to their production of potent hepatotoxins and neurotoxins in freshwater lakes and water reservoirs. Of the known toxins, the hepatotoxic microcystins are produced in various cyanobacterial genera, including *Microcystis*, *Anabaena*, *Nostoc*, *Oscillatoria*, and *Planktothrix* (*1–5*). Previous biochemical and genetic studies found that these cyclic heptapeptides are assembled on a large multienzyme complex *via* a mixed nonribosomal peptide synthetase/polyketide synthase (NRPS/PKS) pathway, the first such pathway identified in a cyanobacterium. NRPSs and PKSs are large, multifunctional protein complexes that catalyze the synthesis of nonribosomal peptides and polyketides, respectively, from amino acid and small chain carboxylic acid monomers *via* a thiotemplate mechanism (Figure 1) (*6, 7*). The microcystin synthetase (*mcy*) gene cluster was cloned from *Microcystis aeruginosa* (Figure 2) (*8–10*) and more recently from strains of the genera *Planktothrix* (*11*) and *Anabaena* (*12*). Through comparison of the sequence data from the *mcy* gene clusters with other NRPS and PKS systems, a biosynthetic scheme for microcystin was proposed (*8–10*). Microcystin synthetase contains an adenylation–peptidyl carrier protein (A–PCP) loading didomain at the N-terminus of the McyG PKS that has been postulated to activate and load the starter unit phenylacetate for subsequent polyketide extension in the formation of the novel aromatic β-amino acid (2S,3S,8S,9S)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-decadienoic acid (Adda) residue (Figure 2) (*9, 10*). While the structure of Adda suggests priming with phenylacetate, *in vivo* feeding experiments do not support the direct involvement of this starter unit (*13, 14*). Hence, on

*To whom correspondence should be addressed.
E-mail: kelleher@scs.uiuc.edu.
E-mail: bsmoore@ucsd.edu.

[¶]These authors contributed equally to this work.

Received for review November 18, 2005 and accepted March 2, 2006

Published online March 17, 2006

10.1021/cb500007v CCC: $33.50

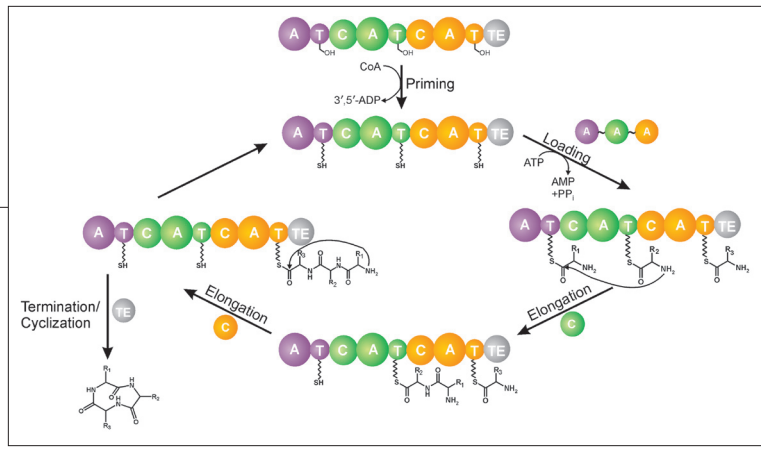

Figure 1. General mechanism of NRPS biosynthesis. Each module (denoted by color) contains the machinery to incorporate one monomer into the growing peptide chain. A minimal module contains an adenylation (A) domain, a condensation (C) domain, and a thiolation (T) domain (also known as peptidyl carrier protein or PCP) to which the intermediates are covalently tethered. The carrier protein domain contains an active site Ser to which a phosphopantetheine moiety is post-translationally loaded. This moiety then serves as the transporting group, moving the elongation intermediates between the catalytic centers. The A domains are responsible for recognizing and activating the requisite amino acid substrates and loading them onto the carrier protein domains. The C domains are responsible for catalyzing condensation, which forms the peptide bonds between the subunits. The final product composition depends on the number and arrangement of these modules, with the final product release step catalyzed by a thioesterase (TE) domain.

the basis of these incompatible *in vivo* data, we set out to investigate the substrate selectivity and specificity of the McyG didomain (McyG A-PCP).

Conventional methods to analyze the biosynthesis of NRPS/PKS complexes include quantification of individual intermediates present at stoichiometric levels by autoradiography and/or base hydrolysis of the thioester intermediates and analysis by radio-TLC (or high-performance liquid chromatography (HPLC)) (*15–18*). However, breakthroughs in electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) (*19, 20*) catalyzed a period of technology development that established a variety of instrumental modes for mass spectrometry (MS) to report on the molecular structure of small molecules, peptides, and intact proteins. Currently, the direct interrogation of enzyme intermediates by MS is illuminating a diverse array of chemistries used in nature (*21*). The development of hybrid quadrupole Fourier-transform mass spectrometry (Q-FTMS) instruments has enabled the detection of low-level components from complex mixtures and verification of their identity by high-resolution tandem MS (MS/MS) (*22*). The application of ESI coupled with Q-FTMS for the interrogation of thioester bound intermediates of the bacitracin, yersiniabactin, epothilone, and aminocoumarin pathways has further improved fundamental understanding of the mechanistic enzymology of these complex pathways (*23–27*). Probing of multiple modules in parallel has been achieved, along with substrate–activity relationships and quantification of active site occupancy. Additionally, proteolysis and liquid chromatography on line with ion-trap MS applied to natural and engineered proteins in the prototypical PKS system, 6-deoxyerythronolide B synthase (DEBS), has recently been reported (*28, 29*).

By use of diverse types of measurement strategies and instruments, a growing number of labs are utilizing the potential of MS to directly observe enzyme-bound intermediates. In this study, we report the *in vitro* and *in vivo* processing of the microcystin loading module (McyG A-PCP) with assorted primer units. Using conventional ATP-PP_i exchange assays and a combination of MS technologies, including MALDI time-of-flight (TOF) MS, ESI-Q-FTMS, and small molecule structural elucidation by gas chromatography/electron impact MS (GC/EI-MS), novel NRPS biochemistry is revealed.

RESULTS AND DISCUSSION

While sequence analysis of the *mcy* gene cluster has identified the mixed NRPS/PKS assembly complex proposed for Adda biosynthesis (*9, 10*), the mechanism of starter unit incorporation remains elusive. Phenylacetate was suggested to be the starter unit for Adda biosynthesis since it correlates with the structures of these related cyclic peptides and recent analysis of their respective biosynthetic gene clusters did not suggest otherwise. Previous feeding experiments, however, did not directly support the incorporation of phenylacetate and instead indicated that the starter unit was derived from L-phenylalanine (*13, 14*). Because of these intriguing observations, we set out to determine the priming mechanism of microcystin synthetase to deduce whether starter unit modifications take place before or after tethering to the assembly line.

The McyG A–PCP Loading Domain. Sequence analysis revealed that the *M. aeruginosa* PCC7820 *mcyG* gene was 99% identical to those from the other *M. aeruginosa* strains. The deduced protein sequence was identical in its arrangement of catalytic domains (Figure 2, panel c) with each containing the conserved active site residues required for activity. The amino acid sequence of the McyG A domain was aligned with the sequence of the GrsA A domain, for which the crystal structure has been solved; however, the presence of additional amino acids around the active site of the McyG A domain which were absent in the GrsA sequence made it difficult to postulate the substrate specificity based on sequence alignments alone

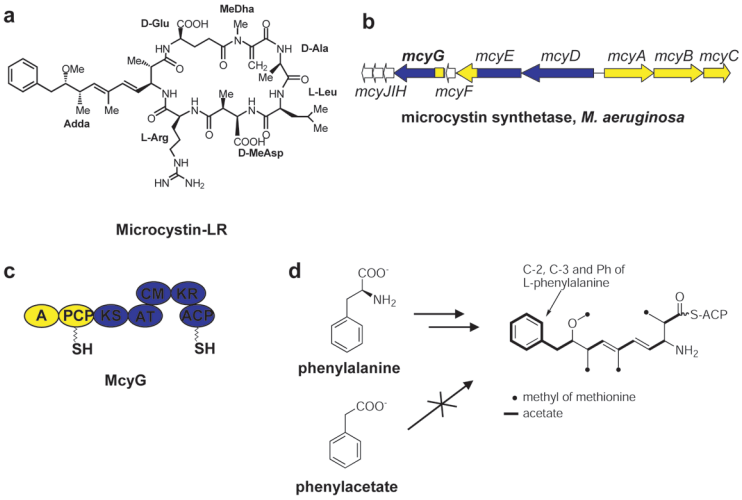

Figure 2. Overview of microcystin-LR structure, biosynthetic genes and proteins, and biosynthesis. a) Structure of microcystin-LR. Uncommon amino acid abbreviations: Adda (3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-decadienoic acid), MeDha (*N*-methyldehydroalanine), MeDhb (*N*-methyldehydrobutyrine), and D-MeAsp (D-*erythro*-β-methylaspartic acid). b) The microcystin (*mcy*) biosynthetic gene cluster from *M. aeruginosa*. The cluster contains 10 bidirectionally transcribed open reading frames comprised of six large multienzyme synthases/synthetases (McyA-E, G) that incorporate the amino acid and acyl-CoA precursors and four monofunctional proteins that are putatively involved in *O*-methylation (McyJ), epimerization (McyF), dehydration (McyI), and localization (McyH). c) The hybrid NRPS (yellow) and PKS proteins (blue) encoded by *mcyG*. The McyG A–PCP didomain (yellow) is the focus of this study. Abbreviations: A (adenylation), PCP (peptidyl carrier protein), KS (ketosynthase), AT (acyltransferase), CM (*C*-methyltransferase), KR (ketoreductase), and ACP (acyl carrier protein). d) Origin of the carbons in the Adda subunit of microcystin.

(*30, 31*). The *apo* McyG A–PCP didomain (78 kDa) was successfully expressed as described in the Methods section and is shown in Supplementary Figure 1.

Characterization of Adenylation Domain Specificity. To investigate the substrate specificity of the McyG didomain that is involved in the proposed loading of phenylacetate during the assembly of the Adda unit in the microcystin peptide, a range of substrates were tested against the *apo* McyG A–PCP didomain using the ATP-PP_i exchange assay (*32*). This enzyme accepted a wide range of phenylpropanoids (Figure 3), and the highest activity was clearly observed with *trans*-cinnamic acid. Other phenylpropanoids were utilized as possible substrates, including D-phenylalanine, *p*-coumarate (activities 30% and 17% relative to cinnamic acid), and to a lesser extent L-phenylalanine, phenylpyruvate, D-3-phenyllactate, and L-3-phenyllactate (less than 10% relative to cinnamic acid). Activity was less than 1% for the substrate phenylacetate along with benzoate, L-alanine, and water as negative controls.

***In Vivo* and *in Vitro* Formation of *Holo* McyG A–PCP.** In order to analyze substrate loading onto the McyG PCP, it was essential to generate the active *holo* form of the McyG PCP by transferring the phosphopantetheine moiety of CoA onto Ser 604. Since the *E. coli* strain used for heterologous expression of the McyG A–PCP protein does not contain a general phosphopantetheinyl transferase (PPTase) that is suitable for post-translational modification of foreign PCPs (*33*), a bifunctional $pHIS_8$-based expression plasmid was constructed to generate the *holo* PCP *in vivo* by coexpressing the octahistidyl-tagged McyG A–PCP didomain with the Svp PPTase protein from *Streptomyces verticillus* ATCC15003 (*34*). Absence of the poly-histidine tag attached to the Svp protein enabled direct purification of the tagged *holo* McyG A–PCP didomain (*holo* A–$PCP_{in\ vivo}$) by

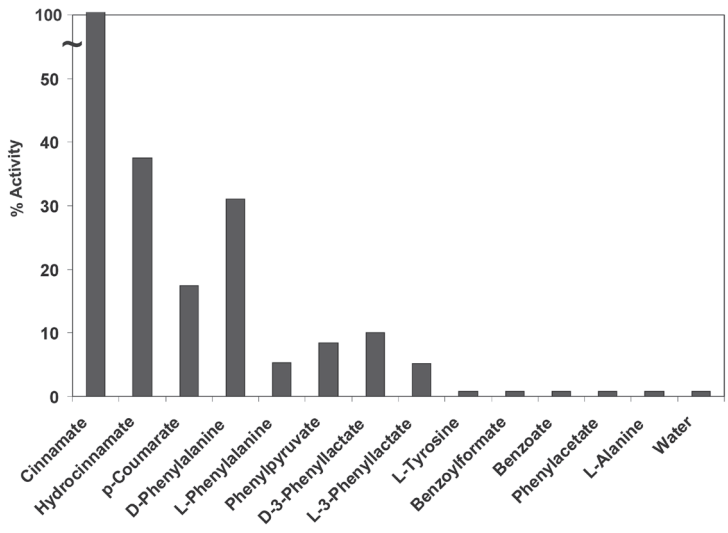

Figure 3. The McyG adenylation domain and its relative substrate-dependent ATP–PP_i exchange activities for a variety of substrates. Activity of the McyG didomain relative to cinnamic acid (100%).

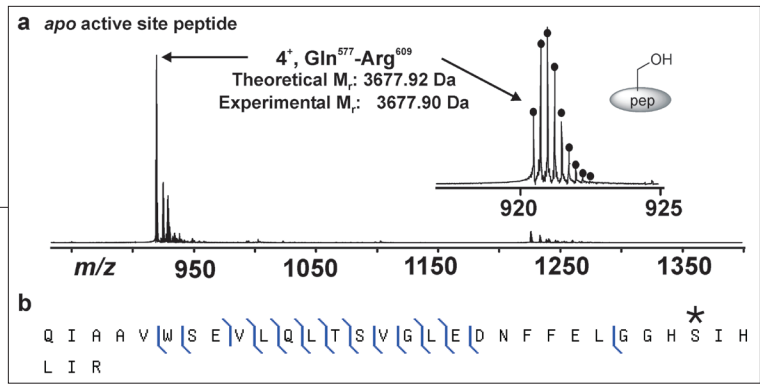

Figure 4. Identification of the tryptic peptide containing the McyG-PCP active site serine. a) Fourier-transform mass spectrum of the Q^{577}-R^{609} tryptic peptide containing the *apo* McyG-PCP active site serine (Ser^{604}). b) The fragment ion map correlates the observed MS/MS fragment ions (not shown) to the Q^{577}-R^{609} peptide; the active site serine at position 604 is marked with an asterisk.

Ni^{2+}-affinity chromatography. Alternatively, purified PPTases, Svp (*34*) or Sfp from *Bacillus subtilis* (*35*), were used to catalyze the post-translational modification of the *apo*-PCP *in vitro* to form *holo* A–$PCP_{in\ vitro}$.

Analysis of Intact McyG A-PCP by MALDI–TOF MS. Phosphopantetheinylation of the McyG PCP *in vivo* and *in vitro* was confirmed by MALDI–TOF MS analysis. The *apo* A–PCP protein had the expected mass; however, the *holo* A–$PCP_{in\ vivo}$ protein was found to be 511 Da greater than the mass of the *apo* protein, far greater than the anticipated 340 Da increase that would result from the attachment of the phosphopantetheine cofactor from coenzyme A. MALDI–TOF MS analysis of the *holo* A–$PCP_{in\ vitro}$ protein revealed the *holo* enzyme (78604 Da) along with this larger species (78751 Da) previously observed in the *in vivo* generated A–PCP sample (Supplementary Table 1). These data suggested that the overweight *holo* protein may reflect a common acylated form that was modified not only *in vivo* but also *in vitro* through the copurification of an acyl-AMP bound to the enzyme that when converted from *apo* to *holo* autoacylates. In order to characterize this enzyme-bound intermediate at isotopic resolution, proteolysis combined with ESI-Q-FTMS was utilized.

Q-FTMS Characterization of McyG A–PCP. HPLC separation of peptides generated by exhaustive trypsin digestion of the *apo* McyG A–PCP didomain followed by Q-FTMS analysis was performed in order to characterize the PCP active site (Figure 4). Initial experiments detected a peptide corresponding to Gln^{577}-Arg^{609} in the *apo* didomain (Figure 4, panel a). The identity of this PCP active site containing peptide was confirmed by tandem mass spectrometry (data not shown), the fragment ions from which are correlated to the peptide primary sequence as seen in the graphical fragment map (Figure 4, panel b). Analysis of all HPLC fractions indicated that the peptide was present only in its *apo*-form, revealing that the *E. coli* PPTases (EntD or ACPS) had no detectable activity during McyG A–PCP overexpression.

Acylation of *in Vitro*-Generated *Holo* McyG A–PCP Didomain. Purified *apo* A–PCP protein was incubated with Sfp and CoA in order to generate the *holo* A–$PCP_{in\ vitro}$ protein and subsequently used to analyze loading of substrates *in vitro* (Figure 5). In the absence of substrate, a small amount of *holo* peptide was identified; however, the majority was found to have a M_r value 132.0 Da greater than the *holo* form (Figure 5, panel a). This observation is consistent with the MALDI–TOF MS results, which identified that a mixture of *holo* and acylated protein was generated from *in vitro* phosphopantetheinylation. On the basis of the +132.0 Da mass shift (Δm) and the structural similarity to the active phenylpropanoid compounds in the ATP-PP_i exchange assay, the substrate loaded onto the PCP was hypothesized to be either hydrocinnamate or benzoylformate (both would add 132.0 Da).

To elucidate the structure of the loaded substrate, the acyl-*S*-PCP protein was incubated with the surfactin thioesterase II enzyme (*36*), resulting in the release of the hydrolyzed substrate and intact *holo* A–PCP protein. After precipitation of the protein, analysis of the supernatant by negative ion mode ESI-MS revealed that the hydrolyzed substrate had an *m/z* of 149, which was identical to the hydrocinnamate control (data not shown). The retention time and ion fragmentation pattern resulting from GC/EI-MS of the substrate were also identical to the hydrocinnamate standard (Supplementary Figure 2), confirming that hydrocinnamyl-*S*-PCP is generated in *E. coli*. Here, the utility of unambiguous structural assignment of unexpected intermediates by small molecule mass spectrometry was demonstrated. Revisiting the ATP-PP_i exchange assay with these two substrates further confirmed hydrocinnamate as the *in vivo* acylating molecule, as hydrocinnamate was active and benzoylformate was not (Figure 3). Lastly, when *holo* A–$PCP_{in\ vitro}$ protein was generated from the *apo* A–PCP protein in the absence of ATP, the hydrocinnamyl-*S*-PCP species is observed at >85% occupancy in repeated experiments (data not shown). We conclude that hydrocinnamate is adenylated *in vivo* to form hydrocinnamyl-AMP which is then retained within the adenylation active site during purification of the A–PCP didomain in a ~1:1 stoichiometry (*37*). Once the *holo* PCP is generated *in vitro*, the hydrocinnamyl substrate is then transferred from AMP to the thiol of the *holo* PCP, to form the hydrocinnamyl-*S*-PCP thioester. Only

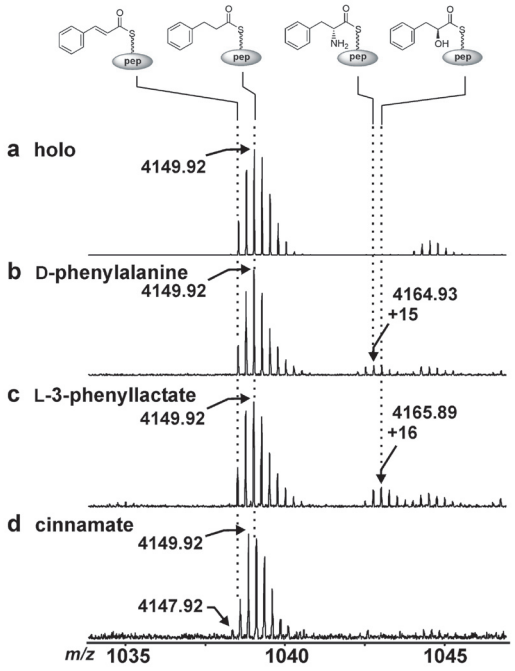

Figure 5. *In vitro* acyl-*S*-PCP peptide intermediates from McyG A–PCP generated with assorted phenylpropanoids: structures and mass shifts. Structures at the top of the figure from left to right show the cinnamyl- (+130 Da), hydrocinnamyl- (+132 Da), D-phenylalanyl- (+147 Da), and 3-phenyllactyl-*S*-PCP (+148 Da) derived Q^{577}-R^{609} peptides (pep). Mass shifts (*i.e.*, +XXX Da) are noted relative to the *holo* peptide. Shown are the *in vitro* loading on *holo* McyG-PCP (1035–1045 *m/z* region; 4+ ions) of a) hydrocinnamate derived from the growth medium, b) D-phenylalanine, c) L-3-phenyllactate, and d) cinnamate. In each case, the majority of the peptide has a MW of 4149.92, correlating to the acylation of the PCP with hydrocinnamyl-AMP following *in vitro* phosphopantetheinylation with Sfp.

a minor fraction (<5%) of the *in vitro* generated *holo* PCP contained the free thiol group, thereby limiting enzyme occupancy available for acylation.

Consistent with ATP-PP_i assays, when *holo* A–$PCP_{in\ vitro}$ was incubated with ATP and substrates D-phenylalanine and L-3-phenyllactate, the *holo* PCP was converted to the acylated form (Figure 5, panels b and c). However, in each case, the hydrocinnamyl-*S*-PCP was still present as the major species. *Holo* A–$PCP_{in\ vitro}$ was also incubated with ATP and cinnamate to generate cinnamyl-*S*-PCP (+130.0 Da); however identification was complicated by the presence of the overlapping isotopic distribution from hydrocinnamyl-*S*-PCP (+132.0 Da) (Figure 5, panel d).

For a more robust loading to be observed, the *holo* A–$PCP_{in\ vitro}$ reaction was incubated with surfactin thioesterase II enzyme to regenerate free, unacylated *holo* A–$PCP_{in\ vitro}$ enzyme. Subsequent loading experiments with various substrates were carried out, the results of which are summarized in Table 1. As expected, phenylpropanoids D-phenylalanine, L-[ring-$^2H_5$]-phenylalanine, L-3-phenyllactate, D-3-phenyllactate, phenylpyruvate, hydrocinnamate, and cinnamate were shown to load, with percent occupancies in a 20 min reaction varying from <5 to 60%, while phenylacetate did not load to any detectable occupancy in this assay. As a multichannel measurement approach with minimal bias, large molecule MS is proving to be a general approach for detection and discovery of noncovalently bound substrates and covalent intermediates on purified enzymes (*21, 37, 38*). For regeneration of *holo* carrier peptides, the type II TE domain allows gentle hydrolysis of the loaded species and re-acylation of the PCP domain by various substrates introduced *in vitro*.

Acylation of *in Vivo* Generated *Holo* McyG A–PCP Didomain. *Holo* A–$PCP_{in\ vivo}$ was similarly digested, fractionated, and analyzed by FTMS. Although the Gln^{577}-Arg^{609} peptide was not detected in the apo form, the *holo*-form was identified and the majority of the peptide was further shifted by +132.0 Da, indicating formation of an acyl-*S*-PCP *in vivo* (Figure 6, panel a). The +132.0 species was assigned as the hydrocinnamyl-*S*-PCP based on the *in vitro* results reported above. This represents, as far as the authors know, the first direct observation of a thioester

TABLE 1. *In vitro* reloading of McyG A–PCP after TE II hydrolysis

Substrate	% occupancy
Cinnamate	20
Hydrocinnamate	50
D-Phenylalanine	40
L-[ring-$^2H_5$]-Phenylalanine	5
Phenylpyruvate	<5
L-3-Phenyllactate	40
D-3-Phenyllactate	60
Phenylacetate	ND[a]

[a]ND, loaded species not detected.

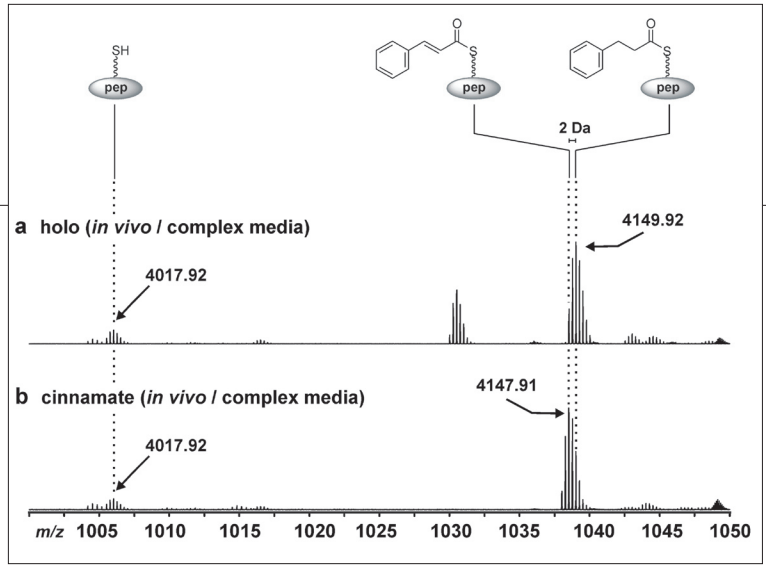

Figure 6. *In vivo* acyl-*S*-PCP peptides from *holo* McyG A–PCP (complex medium): structures and mass shifts. Structures at the top of the figure from left to right show the *holo*-, cinnamyl- (+130 Da), and hydrocinnamyl-*S*-PCP (+132 Da) derived Q^{577}-R^{609} peptides (pep) that correlate to the equivalent peaks in the mass spectra. Shown are the *in vivo* loading on *holo* McyG-PCP (1000–1050 *m/z* region; 4+ ions) of a) hydrocinnamate derived from the growth medium and b) supplemental cinnamate added during IPTG induction.

intermediate formed *in vivo* on an NRPS or PKS carrier domain, raising future possibilities for the detection of (rate limiting) intermediates covalently bound *in vivo* or stalled/misprimed synthetases from natural producing organisms for optimized combinatorial biogenesis of natural product analogues via substrate feeding during fermentation.

While hydrocinnamate is attached *in vivo* to the *holo* A–PCP when expressed in a TB growth medium, cinnamic acid had greater *in vitro* activity based on the $^{32}PP_i$ assay (Figure 3). Therefore, in order to determine if cinnamate can compete for PCP occupancy *in vivo*, it was added at 0.1 mM to a culture of *E. coli* upon induction of the *holo* enzyme. Here, substrate was available in the growth medium and could be actively transported into the cell (*39*). When the *E. coli* growth medium was supplemented with cinnamate, a −2 Da shift in the major species was easily detected by FTMS analysis correlating to formation of the cinnamyl-*S*-PCP *in vivo* (Figure 6, panel b). In a similar fashion, phenylacetate was added to the culture upon induction. However, formation of the phenylacetyl-*S*-PCP was not identified; rather, hydrocinnamyl-*S*-PCP was the major species detected. These results suggest that cinnamate is actively loaded onto the PCP *in vivo*, and despite the addition of the assumed starter unit phenylacetate, hydrocinnamate is preferentially loaded onto the PCP in direct correlation with the *in vitro* ATP-PP_i exchange assay results.

In order to further explore the *in vivo* loading properties of the McyG A–PCP didomain, the *holo* enzyme was expressed in the *E. coli* host grown in a defined, glucose-based medium. We suspected that hydrocinnamic acid, which is preferentially loaded *in vivo*, is derived from the Terrific Broth (TB) fermentation medium. By expressing the *holo* A–PCP protein in a defined glucose-based medium, it was then possible to create a growth environment devoid of these "unnatural" *E. coli* phenylpropanoids. Although expression levels of the didomain protein were lower in the defined medium relative to expression in TB, sufficient quantities were available for characterization. Analysis of the resultant Gln^{577}-Arg^{609} peptide did not yield the +132 Da shift correlating with hydrocinnamyl-*S*-PCP as previously observed in the TB-sample. Instead, a new +148 Da shift was detected (Figure 7, panel a). The mass of this acylated peptide correlates with the loading of 3-phenyllactate vs either phenylalanine or phenylpyruvate which are ~1 and 2 Da off, respectively. Upon addition of hydrocinnamic acid to the culture during induction, the PCP active site peptide was solely present as the hydrocinnamate loaded form (Figure 7, panel b). To investigate the origin of the 3-phenyllactate residue, L-[ring-$^2H_5$]phenylalanine was administered upon induction to the culture grown in the defined medium. Two phenylalanine residues are present in the Gln^{577}-Arg^{609} peptide, and analysis of the *holo* peptide showed the anticipated +10 Da isotope shift (Figure 7, panel c, left). On the other hand, the acylated peptide was shifted +15 Da (Figure 7, panel c, right), consistent with the incorporation of the two phenylalanine residues in the peptide chain and the phenylalanine-derived 3-phenyllactate unit. Using FTMS technology, we were able to identify some unusual characteristics of the McyG A–PCP didomain, which may not have been possible with traditional techniques. The resolution of the FTMS instrument was valuable for discrimination between various acylated species that were generated under specific growth conditions and differed in some cases by just 1 Da.

Utilization of MS Technologies To Elucidate Enzyme-Bound Intermediates. By employing a combination of MS technologies and *in vitro* biochemical assays, we establish through this study that the loading didomain of McyG has broad substrate selectivity for a variety of phenylpropanoids. Further, free

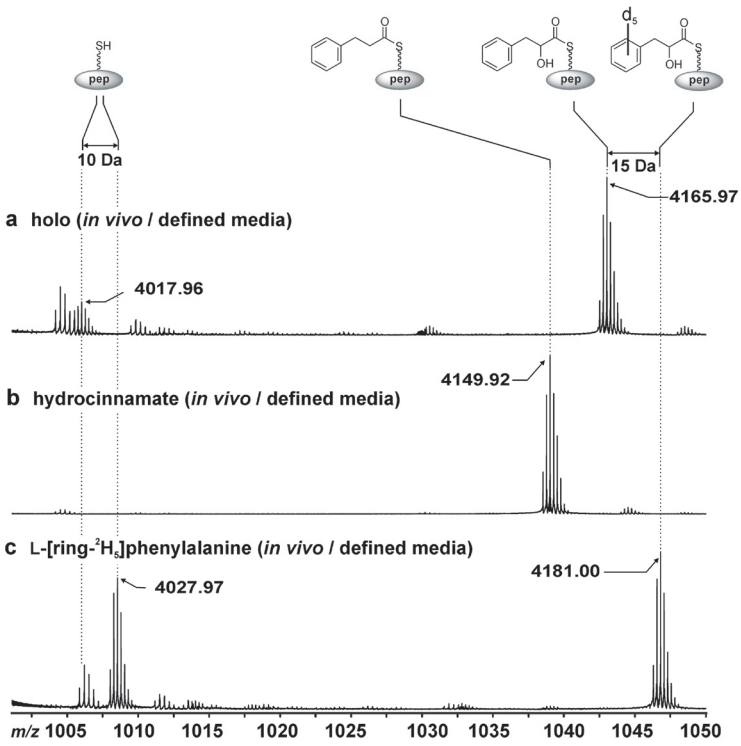

Figure 7. *In vivo* acyl-*S*-PCP peptides from *holo* McyG A–PCP (defined medium): structures and mass shifts. Structures at the top of the figure from left to right show the *holo*-, hydrocinnamyl- (+132 Da), and 3-phenyllactyl-*S*-PCP (+148 Da) derived Q^{577}-R^{609} peptides (pep). Shown are the *in vivo* loading on *holo* McyG-PCP (1000–1050 *m/z* region; 4+ ions) of a) biosynthetic 3-phenyllactate, b) supplemental dihydrocinnamate added during IPTG induction, and c) biosynthetic 3-[ring-$^2H_5$]phenyllactate derived *in vivo* from supplemental L-[ring-$^2H_5$]phenylalanine. The +10 Da shift of the *holo*-PCP derived peptide in panel c is consistent with two [ring-$^2H_5$]phenylalanine residues at positions 597 and 598, whereas the +15 Da shift of the 3-phenyllactyl-*S*-PCP species correlates to a third [ring-$^2H_5$]phenylalanine-derived residue in the form of loaded 3-[ring-$^2H_5$]phenyllactate.

phenylacetate is not a substrate and likely not the biosynthetic precursor of Adda. Results of ATP-PP_i exchange assays and MS demonstrated that substrate chain length was essential for adenylation activity of the McyG A domain, though substitutions at the α carbon were tolerated with little effect on activity. The results correlate with *Microcystis* feeding experiments (*11*) and suggest that the McyG A–PCP didomain has minimal editing function, consistent with previous analyses of other NRPS systems (*40*).

On the basis of these observations, a phenylpropanoid, and not phenylacetate, is the likely primer unit of the Adda polyketide. Preliminary data reveal that a phenylpropanoid primer unit is also incorporated into the Adda moiety of the structurally related cyanobacterial toxin, nodularin, by the NRPS module of NdaC, confirming this phenomenon is not restricted to McyG (*41*, M.C.M and B.S.M. unpublished data). A different phenylpropanoid, such as 3-phenyllactate which is derived from phenylalanine and was selectively activated *in vivo* from *E. coli*, may be the true starter unit. The processing of any phenylpropanoid once bound to the McyG synthetase, however, must undergo novel NRPS biochemistry involving the loss of the thioester carbonyl carbon. If the phenylpropanoid is chain extended with malonyl-CoA by the McyG PKS machinery, then the phenylalanine-derived carboxyl carbon would need to be excised, resulting in the cleavage of two C–C bonds and the formation of a new C–C bond. We are at present actively pursuing the product profile of the intact McyG protein by FTMS to determine the fate of various phenylpropanoid molecules upon PKS extension with malonyl-CoA.

Many structurally diverse bioactive natural products emanate from NRPS and PKS biosynthetic pathways. Thus, an astute understanding of how NRPSs and PKSs function at the molecular level is of high importance to further the ongoing effort to re-engineer these enzymes for efficient generation of novel therapeutics. Although substrate specificity can be predicted from the structure of the final product and the sequence of the biosynthetic gene cluster, this is based on reference to prototypical systems. Here, high-resolution MS allowed readout of *in vivo* and *in vitro* substrate preferences in a largely unbiased fashion. The ability to elucidate unexpected structures of enzyme-bound species in a discovery mode greatly complements more widespread biochemical techniques for determining substrate specificity and mechanism of thiotemplate biosynthesis, a process that increasingly defies highly stringent classification based on prototypical systems studied in detail to date.

METHODS

Bacterial Strains and Growth Conditions. *Escherichia coli* strain XL1-Blue was used for subcloning and grown in Luria broth liquid media or agar plates (Becton Dickinson). Unless otherwise stated, overexpression experiments were performed in Terrific Broth (TB) (*42*) using the expression strain *E. coli* BL21(DE3). *Microcystis aeruginosa* PCC7820 genomic DNA was provided by Professor R. E. Moore, University of Hawaii.

Design of McyG A–PCP Didomain Domain Expression Vectors. The 7857 bp *mcyG* gene fragment was amplified from *M. aeruginosa* PCC7820 using DNA primers based on the 5′- and 3′-ends of the *mcyG* gene from *M. aeruginosa* strains PCC7806 (*10*) and K-139 (*8, 9*). High-fidelity PCR reactions were performed using *Pfu* Turbo polymerase as described by the manufacturer (Stratagene). Specific forward and reverse oligonucleotide primers (synthesized by Qiagen) were added to a final concentration of 0.2 μM. Amplified DNA was cloned directly into the pGEM vector (Promega) and sequence confirmed by specific primer walking.

The *mcyG* gene was amplified using the primers McyG-BamHI (5′-ggatcccattccatcagttttggagg-3′) and GACP-SalIRev (5′-aagcttgtcgacttattcttgtaagattaaatc-3′). The 2014 bp ***mcyG*** gene fragment encoding the NRPS A–PCP didomain was amplified using the primers McyG-BamHI (5′-ggatcccattccatcagttttggagg-3′) and McyG-HindR2 (5′-ggaaaagcttcgtcattcaaataattgc-3′) and subcloned into pHIS$_8$ (*43*) using the restriction enzymes BamHI and HindIII. The resulting plasmid, pMAT701, was used for the expression and purification of *apo* McyG A–PCP didomain. The 2014 bp *mcyG* gene fragment encoding the NRPS A–PCP didomain was also directionally ligated into pBM58, the pHIS8-based vector which coexpresses Svp, the phosphopantetheinyl transferase protein from *Streptomyces verticillus* ATCC15003, from the T7 promoter (*44*). The resulting plasmid, pMAT692 (Svp), was used for the expression and purification of the *holo* McyG A–PCP$_{in\ vivo}$ didomain.

Overexpression and Purification of His-Tagged McyG A–PCP. For expression of McyG protein, 1.5 mL of an overnight culture of *E. coli* BL21(DE3), transformed with the appropriate vector, was inoculated into 75 mL of TB containing 50 μg mL^{-1} kanamycin. The culture was grown at 37 °C to an OD$_{600}$ of between 0.4 and 0.6. Expression of the protein was then induced *via* the addition of IPTG at a final concentration of 0.5 mM, and cell growth continued for 5 h at 30 °C. Where necessary, the culture media was supplemented with 1 mg of cinnamate, L-[ring-$^2H_5$]phenylalanine, or phenylacetate when protein expression was induced. Following expression, cells were harvested by centrifugation at 5000 rpm for 10 min and stored at −80 °C until use.

Alternatively, for the expression of *holo* McyG A–PCP$_{in\ vivo}$, 4.5 mL of an overnight culture of *E. coli* BL21(DE3), transformed with pMAT692 (Svp), cells was collected by centrifugation and the cell pellet washed twice with 4.5 mL of wash buffer (3 g of KH_2PO_4, 6 g of Na_2HPO_4, 0.5 g of NaCl per liter (pH 7.4)). The cells were then inoculated into 150 mL of modified M9 minimal media (pH 7.4) (*45*). The culture was grown at 37 °C to an OD$_{600}$ of 0.6. Expression of the protein was then induced *via* the addition of IPTG at a final concentration of 0.25 mM, and cell growth continued for 20 h at 30 °C. Where necessary, the culture medium was supplemented with 2 mg of hydrocinnamate when protein expression was induced. Alternatively, 2 mg of L-[ring-$^2H_5$]phenylalanine was added 3 times at 2 h intervals after the cells were inoculated into the media.

For purification of the N-terminal His$_8$ tagged protein, cells were thawed in 15 mL of purification buffer (50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol) containing 10 mM imidazole. Cells were lysed by sonication, and the insoluble fraction separated from the soluble proteins *via* centrifugation at 13000 rpm for 30 min. The total soluble fraction was applied to a column containing 0.5 mL of Ni–agarose (Qiagen) before washing three times with 5 mL of purification buffer containing 20 mM imidazole. The expressed protein containing the N-terminal His$_8$ tag was then eluted from the column with 5 × 250 μL of purification buffer containing 250 mM imidazole. Purified protein was then exchanged into Tris protein storage buffer (100 mM Tris (pH 7.2), 2 mM EDTA, 1 mM tris(2-carboxy-ethyl)phosphine hydrochloride (TCEP), 10% glycerol) using a PD-10 column (Pharmacia) and concentrated using a Microcon YM-50 centrifuge filter (Millipore). Protein was then used directly in assays or stored at −80 °C prior to use. All protein samples were analyzed by separation through a 7.5% or 10% SDS–polyacrylamide gel (BioRad) and visualized using Bio-Safe coomassie stain (BioRad). Amino acid numbering is based on the McyG sequence from *M. aeruginosa* PCC7806.

***In Vitro* ATP-PP$_i$ Exchange Assays.** Protein (0.75–10 μg) was combined with 1.5 mM substrate carboxylic acid and 2 mM sodium [^{32}P]pyrophosphate (PP$_i$) (activity of 500 nCi) in reaction buffer (75 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM ATP, 5 mM dithiothreitol (DTT)) at a final reaction volume of 100 μL and incubated for 45 min at room temperature. Reactions were terminated *via* the addition of 30 μL of charcoal solution (2.5% charcoal, 10% trichloroacetic acid, 20 mM Na_2HPO_4). The charcoal was pelleted by centrifugation at 13000 rpm for 5 min and washed twice with 200 μL of water. Charcoal was then resuspended in 150 μL of water, transferred to a scintillation vial with 2 mL of scintillant, and radioactivity counted.

***In Vitro* Phosphopantetheine Transfer and Substrate Loading.** Prior to substrate loading, the carrier domains were first phosphopantetheinylated (100 mM Tris (pH 7.2), 5–10 mM $MgCl_2$, 1–2 mM TCEP, 0.6 mg mL^{-1} Svp/Sfp, 160 μM coenzyme A, and 27 μM *apo* McyG A–PCP protein) in 100 μL reactions by incubating for 3–4 h at room temperature or 30 °C. For subsequent loading of substrates onto the *holo* McyG PCP domain, 5–10 mM ATP and 1.5 mM substrate (D-phenylalanine, L-[ring-$^2H_5$]phenylalanine, L-3-phenyllactate, D-3-phenyllactate, cinnamate, hydrocinnamate, or phenylpyruvate) were added to the reaction and incubation continued for 15 min to 1 h at room temperature or 30 °C. Samples were analyzed by MALDI-TOF MS (AZCC/SWEHSC proteomics core facility, University of Arizona). Alternatively, samples analyzed by FTMS were immediately digested using the protocol described below.

Enzymatic Hydrolysis of Substrates and Small Molecule Structural Elucidation. Covalent intermediates from the PCP domain of McyG A–PCP were hydrolyzed by the addition of a type II external thioesterase domain (SrfA-D) (*36, 46*) and incubation for 1–4 h at 30 °C. For small molecule characterization, proteins were precipitated by the addition of 50% TFA following incubation, and the supernatant was subsequently analyzed by negative ion mode ESI-MS (Micromass Quattro Tandem quadrupole/hexapole/quadrupole instrument) and/or GC/EI-MS (Micromass 70-VSE double focusing sector Instrument) for small molecule identification and structural validation. For McyG A–PCP reloading experiments, reactions were centrifuged in Microcon YM-50 centrifugal filter devices (Millipore) for 2 min at 9000 rpm and the ~30 μL of *holo* McyG A–PCP present in the supernatant was used for substrate loading experiments as described above.

Digestion of McyG A–PCP. Proteolysis was performed by the addition of TPCK-treated trypsin (Promega) to the target protein at protease to substrate ratios ranging from 1:5 to 1:10 w/w in 50 mM NH_4HCO_3, pH 7.8, and incubated at 30 °C for 5 min. Reactions were quenched by the addition of an equal volume of 10% formic acid (Acros) and applied to a wide-pore Jupiter C4 reversed-phase column (4.6 × 150 mm) (Phenomenex) with a linear gradient from 10% to 90% MeCN (0.1% TFA) over 60 min for fractionation/desalting prior to FTMS analysis. Samples were lyophilized before resuspension in 49% H_2O, 49% MeOH, and

2% formic acid for FTMS analysis. *Holo* and loaded active site containing HPLC fractions from 33 to 37 min were combined prior to FTMS analysis.

ESI-Q-FTMS and Reported Masses. ESI was used with a custom Q-FTMS instrument operating at 8.5 T (*22*). The ions were directed through a heated metal capillary, skimmer, quadrupole, and multiple ion guides into the ion cell (~10^{-9} Torr) of the FTMS. Scans were acquired every 1 s, and data were stored with a MIDAS datastation (*47*) as 512k data sets. Spectra were calibrated externally using bovine ubiquitin (MW = 8559.62 Da), and theoretical isotopic distributions were generated using Isopro v3.0 and fit to experimental data by least squares to assign the most abundant peak.

High-resolution mass spectrometry of large molecules results in isotopic distributions within the mass spectra, explanations of which have been described previously (*48, 49*). Briefly, while arrows in figures point to the most abundant isotopes, all molecular weight values (M_r) in this paper are reported as monoisotopic values, which refer to the molecular ion peak composed of the most abundant isotopes of the elements including the mass defect (*i.e.*, *C* = 12.000000, *N* = 14.00307, etc). Assignment of isotopic distributions to the corresponding enzyme intermediates involved correlating the experimental monoisotopic M_r values to the theoretical monoisotopic M_r values for the enzyme intermediates, with a maximum error of 20 ppm (Supplementary Table 2).

Accession Codes. The *mcyG* sequence, isolated from *Microcystis aeruginosa* PCC7820, reported in this paper has been deposited in GenBank under accession number AY910575.

Acknowledgment: We thank R. Moore for providing strains, C. Walsh for the SrfA-D plasmid, L. Miller for overexpression of the SrfA-D construct and providing Figure 1, R. Milberg for assistance with the GC/EI-MS analysis, the AZCC/SWEHSC proteomics core facility at the University of Arizona for MALDI TOF analyses (supported by NIEHS Grant ES06694 and NIH/NCI Grant CA023074-26), the DNA Sequencing Service at the University of Arizona for performing sequencing reactions, J. Becker for assistance, and J. Jez for critical reading of the manuscript. The Quattro and 70-VSE mass spectrometers were purchased in part with grants from the Division of Research Resources, National Institutes of Health (RR 07141 and RR 04648, respectively). Financial support is greatly appreciated from the Washington Sea Grant program (R/B-39) to B.S.M. and the University of Illinois and the National Institutes of Health (GM 067725) to N.L.K. L.M.H. is a recipient of an NSF Graduate Research fellowship. This article is dedicated in memory of Kenneth L. Rinehart at the University of Illinois.

Supporting Information Available: This material is available free of charge via the Internet.

REFERENCES

1. Rinehart, K. L., Harada, K., Namikoshi, M., Chen, C., Harvis, C. A., Munro, M. H. G., Blunt, J. W., Mulligan, P. E., Beasley, V. R., Dahlem, A. M., and Carmicheal, W. W. (1988) Nodularin, microcystin, and the configuration of Adda, *J. Am. Chem. Soc. 110*, 8557–8558.
2. Namikoshi, M., Yuan, M., Sivonen, K., Carmichael, W. W., Rinehart, K. L., Rouhiainen, L., Sun, F., Brittain, S., and Otsuki, A. (1998) Seven new microcystins possessing two L-glutamic acid units, isolated from *Anabaena* sp. strain 186, *Chem. Res. Toxicol. 11*, 143–149.
3. Beattie, K. A., Kaya, K., Sano, T., and Codd, G. A. (1998) Three dehydrobutyrine-containing microcystins from *Nostoc*, *Phytochemistry 47*, 1289–1292.
4. Sano, T., and Kaya, K. (1998) Two new (E)-2-amino-2-butenoic acid (Dhb)-containing microcystins isolated from *Oscillatoria agardhii*, *Tetrahedron 54*, 463–470.
5. Meriluoto, J. A. O., Sandstrom, A., Eriksson, J. E., Remaud, G., Grey, C. A., and Chattopadhyaya, J. (1989) Structure and toxicity of a peptide hepatotoxin from the cyanobacterium *Oscillatoria agardhii*, *Toxicon 27*, 1021–1034.
6. Schwarzer, D., Finking, R., and Marahiel, M. A. (2003) Nonribosomal peptides: from genes to products, *Nat. Prod. Rep. 20*, 275–287.
7. Walsh, C. T. (2004) Polyketide and nonribosomal peptide antibiotics: modularity and versatility, *Science 303*, 1805–1810.
8. Nishizawa, T., Asayama, M., Fujii, K., Harada, K., and Shirai, M. (1999) Genetic analysis of the peptide synthetase genes for a cyclic heptapeptide microcystin in *Microcystis* spp, *J. Biochem. (Tokyo) 126*, 520–529.
9. Nishizawa, T., Ueda, A., Asayama, M., Fujii, K., Harada, K., Ochi, K., and Shirai, M. (2000) Polyketide synthase gene coupled to the peptide synthetase module involved in the biosynthesis of the cyclic heptapeptide microcystin, *J. Biochem. (Tokyo) 127*, 779–789.
10. Tillett, D., Dittmann, E., Erhard, M., von Dohren, H., Borner, T., and Neilan, B.A. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system, (2000) *Chem. Biol. 7*, 753–764.
11. Christiansen, G., Fastner, J., Erhard, M., Borner, T., and Dittmann, E. Microcystin biosynthesis in *Planktothrix:* genes, evolution, and manipulation, (2003) *J. Bacteriol. 185*, 564–572.
12. Rouhiainen, L., Vakkilainen, T., Siemer, B. L., Buikema, W., Haselkorn, R., and Sivonen, K. (2004) Genes coding for hepatotoxic heptapeptides (microcystins) in the cyanobacterium *Anabaena* strain 90, *Appl. Environ. Microbiol. 70*, 686–692.
13. Moore, R. E., Chen, J. L., Moore, B. S., Patterson, G. M. L., and Carmichael, W. W. (1991) Biosynthesis of microcystin-LR. Origin of the carbons in the Adda and Masp units, *J. Am. Chem. Soc. 113*, 5083–5084.
14. Rinehart, K. L., Namikoshi, M., and Choi, B.W. (1994) Structure and biosynthesis of toxins from blue-green algae (cyanobacteria), *J. Appl. Phycol. 6*, 159–176.
15. Chen, H., O'Connor, S. E., Cane, D. E., and Walsh, C. T. (2001) Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. *Chem. Biol. 8*, 899–912.
16. Admiraal, S. J., Khosla, C., and Walsh, C. T. The loading and initial elongation modules of rifamycin synthetase collaborate to produce mixed aryl ketide products, *Biochemistry 41*, 5313–5324.
17. O'Connor, S. E., Chen, H., and Walsh, C. T. (2002) Enzymatic assembly of epothilones: the EpoC subunit and reconstitution of the EpoA-ACP/B/C polyketide and nonribosomal peptide interfaces, *Biochemistry 41*, 5685–5694.
18. Admiraal, S. J., Khosla, C., and Walsh, C. T. (2003) A switch for the transfer of substrate between nonribosomal peptide and polyketide modules of the rifamycin synthetase assembly line, *J Am Chem Soc 125*, 13664–13665.
19. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules, *Science 246*, 64–71.
20. Karas, M., and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons, *Anal. Chem. 60*, 2299–2301.
21. Kelleher, N. L., and Hicks, L. M. (2005) Contemporary mass spectrometry for the direct detection of enzyme intermediates, *Curr. Opin. Chem. Biol. 9*, 424–430.
22. Patrie, S. M., Charlebois, J. P., Whipple, D., Kelleher, N. L., Hendrickson, C. L., Quinn, J. P., Marshall, A. G., and Mukhopadhyay, 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.
23. Hicks, L., Weinreb, P., Konz, D., Marahiel, M. A., Walsh, C. T., and Kelleher, N. L. (2003) Fourier-transform mass spectromtery for detection of thioester-bound intermediates in unfractionated proteolytic mixtures of 80 and 191 kDa portions of Bacitracin A synthetase, *Anal. Chim. Acta 496*, 217–224.

24. 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*, 13393–13400.
25. 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*, 13265–13275.
26. 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*, 327–335.
27. 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*, 2770–2780.
28. 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*, 2373–2387.
29. Schnarr, N. A., Chen, A. Y., Cane, D. E., and Khosla, C. (2005) Analysis of Covalently Bound Polyketide Intermediates on the 6-Deoxyerythronolide B Synthase by Tandem Proteolysis - Mass Spectrometry, *Biochemistry 44*, 11836–11842.
30. 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.
31. 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.
32. Stachelhaus, T., and Marahiel, M. A. (1995) Modular Structure of Peptide Synthetases Revealed by Dissection of the Multifunctional Enzyme GrsA, *J. Biol. Chem. 270*, 6163–6169.
33. Gehring, A. M., Lambalot, R. H., Vogel, K. W., Drueckhammer, D. G., and Walsh, C. T. (1997) Ability of *Streptomyces* spp. acyl carrier proteins and coenzyme A analogs to serve as substrates in vitro for *E. coli* holo-ACP synthase, *Chem. Biol. 4*, 17–24.
34. Sanchez, C., Du, L., Edwards, D. J., Toney, M. D., and Shen, B. (2001) Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin. *Chem. Biol. 8*, 725–738.
35. Quadri, L. E. N., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P., and Walsh, C. T. (1998) Characterization of Sfp, a *Bacillus subtilis* Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases, *Biochemistry 37*, 1585–1595.
36. Schwarzer, D., Mootz, H. D., Linne, U., and Marahiel, M. A. (2002) Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases, *Proc. Natl. Acad. Sci. U.S.A. 99*, 14083–14088.
37. Miller, L. M., Mazur, M. T., McLoughlin, S. M., and Kelleher, N. L. (2005) Parallel Interrogation of Covalent Intermediates in the Biosynthesis of Gramicidin S Using High Resolution Mass Spectrometry, *Protein Sci. 14*, 2702–2712.
38. Burns, K. E., Xiang, Y., Kinsland, C. L., McLafferty, F. W., and Begley, T. P. (2005) Reconstitution and Biochemical Characterization of a New Pyridoxal-5′-Phosphate Biosynthetic Pathway, *J. Am. Chem. Soc. 127*, 3682–3683.
39. Diaz, E., Ferrandez, A., Prieto, M. A., and Garcia, J. L. (2001) Biodegradation of aromatic compounds by *Escherichia coli*, *Microbiol. Mol. Biol. Rev. 65*, 523–569.
40. Luo, L., Burkart, M. D., Stachelhaus, T., and Walsh, C. T. (2001) Substrate recognition and selection by the initiation module PheATE of gramicidin S synthetase, *J. Am. Chem. Soc. 123*, 11208–11218.
41. Moffitt, M. C., and Neilan, B. A. (2004) Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins, *Appl. Environ. Microbiol. 70*, 6353–6362.
42. Sambrook, J., and Russell, D. W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
43. Jez, J. M., Ferrer, J. L., Bowman, M. E., Dixon, R. A., and Noel, J. P. (2000) Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase, *Biochemistry 39*, 890–902.
44. Izumikawa, M., Cheng, Q., and Moore, B. S. (2006) Priming type II polyketide synthases via a type II nonribosomal peptide synthetase mechanism, *J. Am. Chem. Soc. 128*, 1428–1429.
45. Dorrestein, P. C., Zhai, H., McLafferty, F. W., and Begley, T. P. (2004) The biosynthesis of the thiazole phosphate moiety of thiamin: the sulfur transfer mediated by the sulfur carrier protein ThiS, *Chem. Biol. 11*, 1373–1381.
46. Yeh, E., Kohli, R. M., Bruner, S. D., and Walsh, C. T. (2004) Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated, *ChemBioChem 5*, 1290–1293.
47. 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.
48. Yergey, J., Heller, D., Hansen, G., Cotter, R. J., and Fenselau, C. (1983) Isotopic Distributions in Mass Spectra of Large Molecules, *Anal. Chem. 55*, 353–356.
49. McLafferty, F. W. (1994) High-Resolution Tandem FT Mass Spectrometry above 10 kDa, *Acc. Chem. Res. 27*, 379–386.