

Mass Spectrometric Interrogation of Thioester-Bound Intermediates in the Initial Stages of Epothilone Biosynthesis

Leslie M. Hicks,¹ Sarah E. O'Connor,² Matthew T. Mazur,¹ Christopher T. Walsh,² and Neil L. Kelleher^{1,*}

¹Department of Chemistry
600 South Mathews Avenue
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

²Department of Biological Chemistry and Molecular Pharmacology
200 Longwood Avenue
Harvard Medical School
Boston, Massachusetts 02115

Summary

Direct detection of thioester intermediate mixtures bound to EpoC, a 195 kDa polyketide synthase, has been achieved using limited proteolysis and Fourier-transform mass spectrometry (FTMS). Incubation with various *N*-acetylcysteamine thioester (S-NAC) substrate mimics produced mass shifts on the EpoC ACP domain consistent with their condensation with an enzyme-bound carbanion produced by the decarboxylation of methylmalonyl-S-EpoC. Reconstitution of EpoA-ACP, EpoB, and EpoC gave a +165.0 Da mass shift consistent with the formation of the methylthiazolyl-methacrylyl product by incorporation of acetyl-CoA, cysteine, and methylmalonyl-CoA. Thioester-templated reaction intermediates and products are typically characterized by quantifying radioactive substrates, either enzyme bound or chemically hydrolyzed. In contrast, the MS-based methodology described here provides semiquantifiable ratios of free enzyme, intermediate, and product occupancy and reveals that certain substrates result in a >50% formation of nonproductive intermediates.

Introduction

Many therapeutic compounds emanate from multifunctional, multidomain enzymes known as nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS). These remarkable proteins synthesize immunosuppressants, such as cyclosporin and FK506, as well as precursors to critical antibiotics, such as erythromycin and vancomycin. The basic motif of thiotemplate biosynthesis [1, 2] involves three general steps: monomer loading/initiation, condensation/translocation, and termination/cyclization. After ATP-driven acylation of precursors onto free thiols of cofactors covalently bound to peptidyl carrier protein (PCP) domains in NRPS [3] or acyl-S-transfer from CoA derivatives onto acyl carrier protein (ACP) domains in PKS [4], the elongation process proceeds by N- to C-terminal movement of biosynthetic intermediates as they grow in complexity. Further-

more, a host of tailoring domains are incorporated in the enzyme primary structure to accessorize thioester-bound intermediates via reduction, dehydration, amino acid cyclization, epimerization, and oxidation reactions [5]. With much of the biosynthetic logic of NRPS and PKS organized at the level of primary sequence, such enzymes offer a rational strategy for harnessing their biosynthetic potential [6].

NRPS and PKS systems contain multiple domains organized into (~100–200 kDa) modules for activation and condensation of particular substrates. With most biosynthetic intermediates covalently bound to their exceptionally large primary structures, the carrier domains (ACP or PCP) can exist in many states that all differ in their molecular weight (M_r) values: no cofactor loaded (apo), phosphopantetheinyl cofactor loaded (holo), extender unit tethered as a thioester (e.g., methylmalonyl-S-ACP), and complicated mixtures of more advanced biosynthetic intermediates. Typically, monitoring of individual intermediates present at stoichiometric levels has previously been accomplished by autoradiography and/or base hydrolysis of the thioester-linked compounds and analysis by radio-TLC (or HPLC) [7, 8]. However, unconventional approaches utilizing mass spectrometry (MS) to analyze intermediates have been used to characterize acid-hydrolyzed intermediates in the penicillin biosynthesis pathway via HPLC and MALDI-TOF MS [9], while, prior to MS, a lengthy HPLC procedure was used to elucidate intermediates in the gramicidin S system [2]. Various combinations of MS have enabled the detection of the 4' phosphopantetheine cofactor at the thioester binding site for L-valine of gramicidin S synthetase 2 [1], the confirmation of loading of various coenzyme A derivatives onto apo-ACPs via holo-ACP synthase [10], and the identification of active-site cysteine residues in trichodiene synthase [11]. More contemporary incarnations of MS afford the potential for more direct analysis of complex proteolytic mixtures of 5–50 kDa peptides without loss of mass accuracy [12]. Specifically, Fourier-transform mass spectrometry (FTMS) has been forwarded as an informative method to discern the molecular heterogeneity covalently tethered to these large synthetases [7].

Electrospray ionization (ESI) coupled to FTMS offers 10⁵ resolving power, especially valuable for measuring complex mixtures directly [13]. For example, the ESI/FTMS combination has enabled the semiquantification of heterogeneous species present in recombinant thiaminase I from *E. coli* [14] and the direct analysis of oxoester bound intermediates kinetically trapped on EntF from the enterobactin pathway in *E. coli* [13]. Interrogation of enzyme-bound intermediates using “large molecule” MS offers direct and semiquantifiable readout of relative ratios (% occupancies) of intermediate mixtures without radiolabels or synthesis of standard compounds, provided that key carrier peptides can be identified from complex proteolytic mixtures. The development of Quadrupole-FTMS hybrid instruments now enables extended ability to both detect low-level com-

*Correspondence: kelleher@scs.uiuc.edu

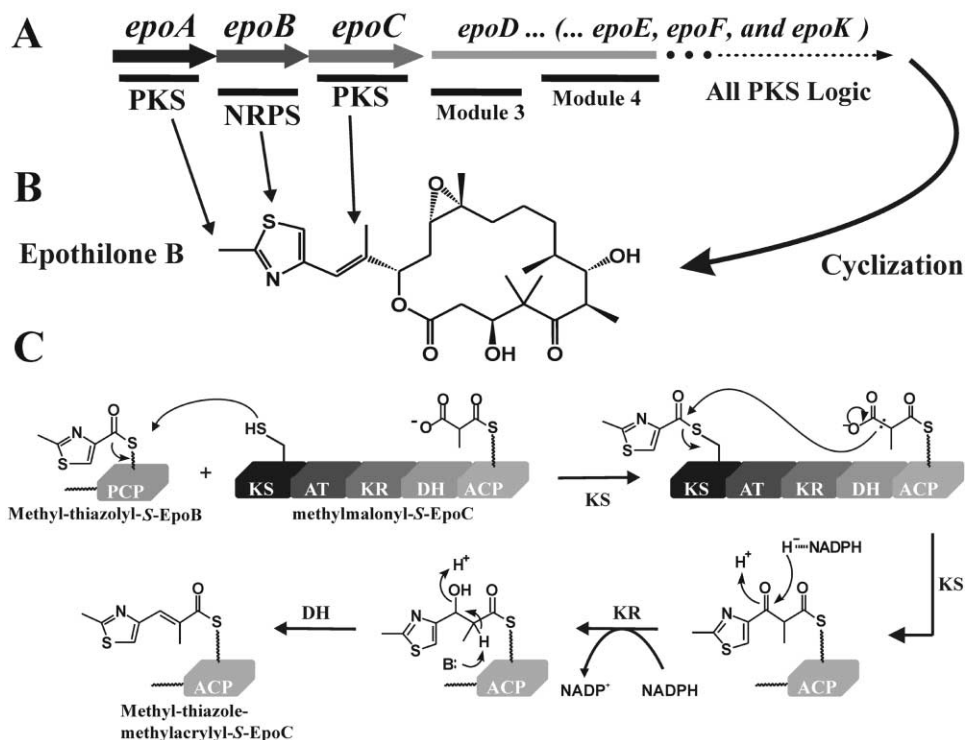


Figure 1. Overview of the Epothilone Biosynthetic Gene Cluster from *Sorangium cellulosum*, Structure of Epothilone B, and Proposed Mechanism of Epothilone Biosynthesis at EpoC, a PKS Module

(A) Epothilone biosynthetic gene cluster from *Sorangium cellulosum*.

(B) Structure of Epothilone B.

(C) Proposed mechanism of epothilone biosynthesis at EpoC, a PKS module.

ponents and verify their identity by high-resolution tandem MS (MS/MS) [15]. The studies presented here demonstrate that ESI/Q-FTMS provides robust information about upstream cofactor requirements and substrate specificities for the EpoC polyketide synthase from epothilone biosynthesis in *Sorangium cellulosum*.

Like the well-known antitumor agent paclitaxel (Taxol), the epothilones have proven to be potent tubulin depolymerization inhibitors in vitro [16], and the anticancer efficacy of several epothilone derivatives is currently being assessed in clinical trials. The gene cluster for the epothilones has been isolated and consists largely of PKS modules, with a single NRPS module, EpoB, located between the first (EpoA) and third (EpoC) PKS modules of the cluster [17, 18]. In the early steps of the epothilone biosynthetic pathway, the methyl-thiazolyl group is formed by acetyl transfer from the first PKS subunit, EpoA, to the cysteine substrate of the NRPS subunit, EpoB [8]. Subsequent cyclization, dehydration, and oxidation by EpoB results in the methyl-thiazolyl species, which then becomes the electrophilic donor to the downstream PKS acceptor subunit EpoC (Figure 1C). The acyltransferase (AT) domain of EpoC intramolecularly primes the ACP domain with a methylmalonyl moiety derived from methylmalonyl-CoA to the holo HS-acyl carrier protein (ACP) in an autoacylation reaction. As illustrated in Figure 1C, the upstream methyl-thiazolyl intermediate is transferred from the PCP carrier site of EpoB to the active-site cysteine of the ketosynthase

(KS) domain of EpoC, and the KS domain of EpoC decarboxylates the methylmalonyl-S-EpoC acyl enzyme to generate the carbon nucleophile that reacts with methyl-thiazolyl substrate. The resulting condensation product can be reduced in the presence of NADPH by the ketoreductase (KR) domain of EpoC and then dehydrated by the dehydratase (DH) domain to produce the methyl-thiazolyl-methylacrylyl-S-EpoC acyl enzyme intermediate that serves as the acyl donor for subsequent elongation of the epothilone chain.

Previous work has reconstituted the early steps of epothilone biosynthesis, using the ACP domain of EpoA (EpoA-ACP), EpoB, and EpoC [19]. Methyl-thiazolyl-S-NAC can also be presented to EpoC to result in formation of the final product. Additionally, unnatural substrates were also shown to be processed by the epothilone machinery [20]. Reaction intermediates in this case were visualized by using a radioactive substrate, hydrolyzing the thioester-bound compounds under basic conditions, and analyzing by HPLC. This approach, although successfully used to characterize many aspects of the epothilone assembly line, poses disadvantages. First, several of the intermediates are labile under the hydrolysis conditions and so cannot be observed directly. For example, malonyl decarboxylates to acetate, while methyl-thiazoline, an intermediate in methyl-thiazole synthesis, opens to N-acetyl-cysteine. Second, it can also be technically challenging to resolve intermediates by TLC or HPLC (i.e., malonyl versus acetyl; cysteine versus N-acetyl-cys-

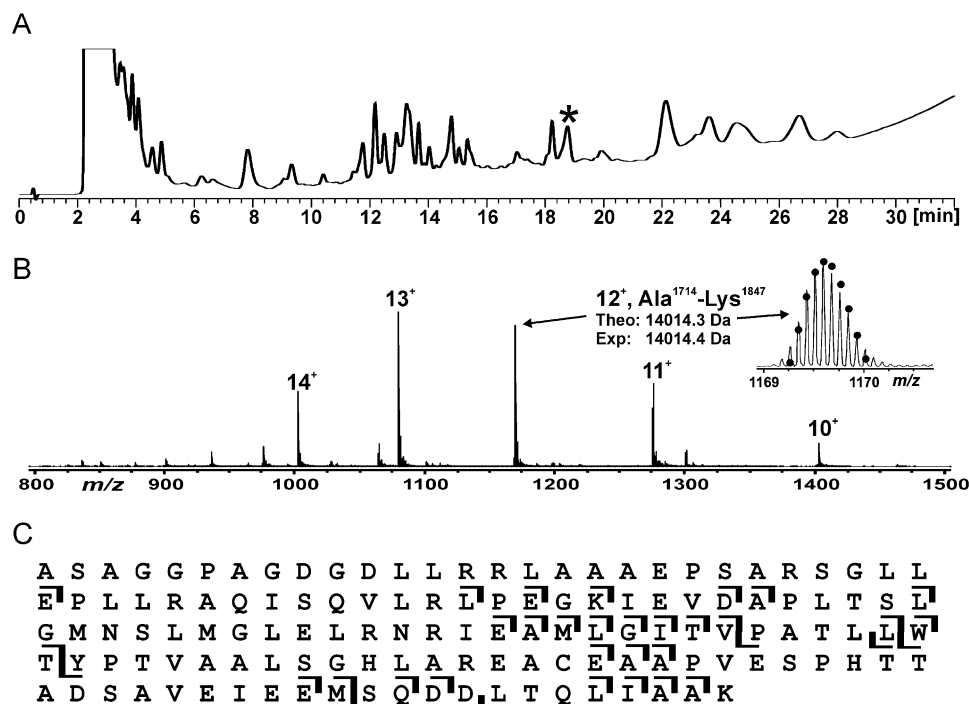


Figure 2. Unambiguous Identification of the Proteolytic Fragment Containing the Active Site Serine of EpoC-ACP

(A) RPLC trace ($\lambda = 220$ nm) of the fractionation of a limited proteolytic digestion of EpoC with trypsin; an asterisk indicates elution of peptide fragment of interest.
 (B) Broadband Fourier-transform mass spectrum (25 scans) of the 18–19 min RPLC fraction; the most abundant species observed corresponds to the proteolytic fragment Ala¹⁷¹⁴-Lys¹⁸⁴⁷, which contains the EpoC-ACP active site serine (Ser¹⁷⁷³), the isotopic distribution of which is shown in the inset.
 (C) The fragment ion map correlating the MS/MS ions observed (data not shown) to the sequence of the ACP carrier peptide (Ala¹⁷¹⁴-Lys¹⁸⁴⁷) generated from EpoC.

teine). Therefore, although the formation of product and the presence of certain reaction intermediates have been previously observed in the epothilone pathway, it has not been possible to quantify the proportions of substrate, reaction intermediates, and products that are bound to the synthase. Here, we present a high-resolution mass spectrometric analysis that provides the positive identification of enzyme starting material, reaction intermediates and products, with semiquantification of their relative ratios.

Results

EpoC

The EpoC gene product was purified from *E. coli* and subjected to limited proteolysis with trypsin followed by HPLC fractionation of the products (Figure 2A). The ESI/FT mass spectrum of the 18–19 min fraction revealed a 14,014.4 Da peptide (Figure 2B), which was subjected to infrared photons for ion dissociation (data not shown). Correlating the intact peptide mass and the observed fragment ions with the DNA-predicted amino acid sequence of EpoC identified this species as Ala¹⁷¹⁴-Lys¹⁸⁴⁷, a peptide fragment that includes the active site serine (Ser¹⁷⁷³) of the ACP domain (Figure 2C); amino acid number refers to EpoC construct including His-tag. The intact M_r value of this ACP carrier peptide indicates that

the majority of ACP is in the apo form, with a satellite peak at +340 Da consistent with ~10% of the protein phosphopantetheinylated during overexpression in vivo (Figure 3A). Although not quantified, autoradiography experiments have indicated that EpoC can be labeled with [³H]-acetyl-CoA (data not shown), a biochemical result consistent with the mass spectrometric data. Treatment of the predominately apo-EpoC with CoA and the phosphopantetheinyl transferase Sfp yielded the spectrum of Figure 3B, revealing nearly total conversion of the ACP domain within EpoC to the holo form (+340 Da), with no trace of apo-enzyme detectable. Also visible in Figures 3A, 3B, and other data sets are unrelated peptides (designated by their respective charge states) and commonly observed “adduct” peaks (designated with asterisks). These arise from oxidation of methionine and/or cysteine residues (+16) and addition of sodium (+22) or potassium (+38) cations instead of protons before and during electrospray ionization.

Detection of Thioester-Bound Intermediates

Treatment of holo-EpoC with methylmalonyl-CoA for 15 min followed by limited proteolysis, HPLC fractionation, and analysis by ESI/Q-FTMS afforded the Figure 3C spectrum. In addition to the holo form of EpoC (14,354.5 Da), species of 56.0 and 100.0 Da higher mass are also visible (in addition to their corresponding adduct peaks).

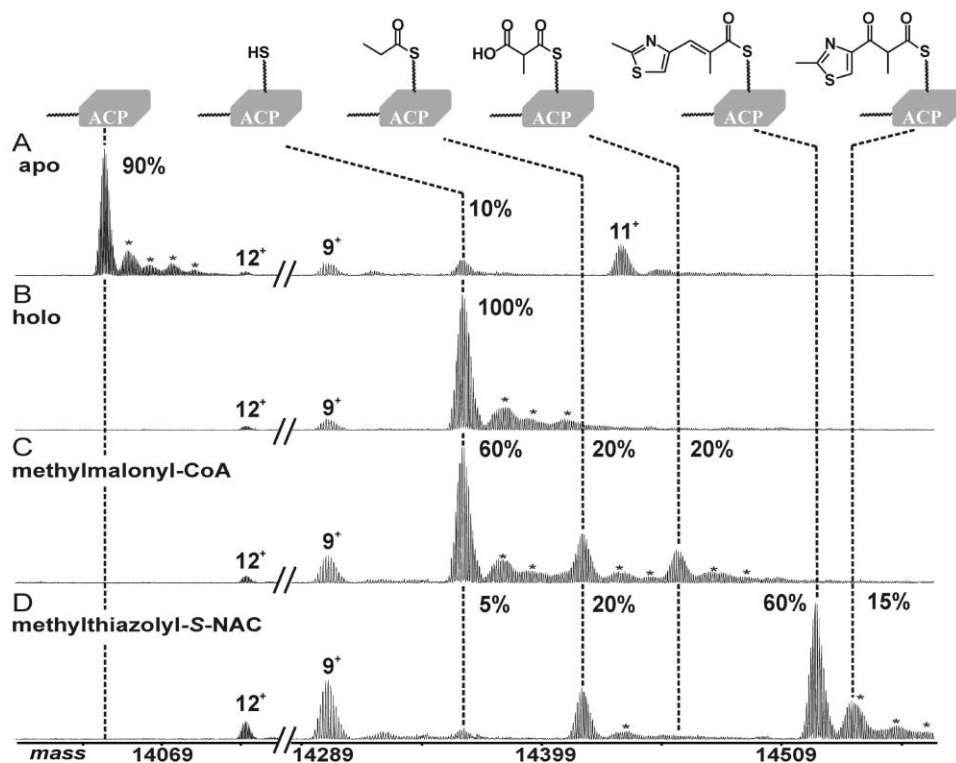


Figure 3. Structures of Various Peptide Forms of Interest Correlated with Loading and Mass Shifts, Including Percent Occupancies, on EpoC-ACP under Various Experimental Conditions

The top of the figure shows structures of various peptide forms of interest (from left to right: apo, holo, decarboxylated methylmalonyl [propionyl] [+56 Da], methylmalonyl [+100 Da], methyl-thiazole-methylacrylyl-S-ACP [+165 Da], and β -ketoacyl intermediate [+181 Da]) and correlation to the equivalent peaks in the mass spectra, as indicated by vertical lines. Shown are the loading and mass shifts, including percent occupancies, on EpoC-ACP (1270–1330 *m/z* region; 11+ ions) of (A) apo-EpoC; (B) holo-EpoC; (C) methylmalonyl loaded; (D) methylmalonyl + methylthiazolyl-S-NAC loaded; asterisks indicate signals arising from artifactual addition of sodium (+22 Da), potassium (+38 Da), or oxidation of Met/Cys residues (+16 Da).

These species are consistent with the thioester intermediates methylmalonyl-S-ACP (+100.0 Da) and propionyl-S-ACP (+56.0 Da), with the latter arising from KS-catalyzed decarboxylation and protonation of the resulting carbanion. The net loss of 44 Da from the +100 Da species is consistent with this chemistry. Quantification of these species gives apparent occupancies of 20% for the methylmalonyl-loaded species (+100 Da) and 20% for the decarboxylated methylmalonyl form (+56 Da) on the ACP domain, while 60% of the carrier protein resides in the holo form.

With the addition of methyl-thiazolyl-S-NAC as an acyl donor for methylmalonyl-S-EpoC to the reaction mixture, 60% of the carrier peptide signal shifted by +165.0 Da, consistent with condensation to form methyl-thiazole-methylacrylyl-S-enzyme (Figure 3D). In this case, no detectable enzyme is bound by methylmalonyl, and the relative ratios of the holo, decarboxylated methylmalonyl, and the β -ketoacyl (+181.0 Da) forms were detected as 5%, 20%, and 15%, respectively (results summarized in Table 1). Careful analysis of the mass spectrum within the adduction region of the methyl-thiazole-methylacrylyl-S-enzyme product has indicated that the reduced β -hydroxyl species prior to dehydration was not observed, consistent with biochemical experiments in which this intermediate was also not observed

by HPLC [19]. An identical experiment with the methyl-thiazolyl-S-NAC donor substrate was carried out in the absence of NADPH. In this case, 10% of the carrier peptide signal shifted by +181 Da, consistent with the nonreduced β -ketoacyl intermediate (Figure 4A). In this case, no detectable enzyme has covalently bound methylmalonyl, ~20% is holo, and ~70% is in the methylmalonyl decarboxylated (propionyl) form (Table 1). Analogous experiments obtained under identical reaction conditions to the one presented in Figure 3D were carried out with the alternate S-NAC acyl donor substrates methylpyridine-S-NAC, phenylthiazolyl-S-NAC, 2-hydroxyphenylthiazolyl-S-NAC, and phenyloxazolyl-S-NAC. Methylmalonyl-S-EpoC catalyzes acyl transfer, C-C bond formation, β -keto reduction, and dehydration to give the indicated methacryl-S-enzyme species (Figures 4B–4E). The relative ratios of thioester-bound EpoC intermediates for all S-NAC experiments are summarized in Table 1.

To study the identity and abundance of acyl-S-EpoC intermediates from the holo forms of the three subunits EpoA, B, and C, reconstitution of acetyl-loaded EpoA-ACP, cysteinyl-loaded EpoB, and methylmalonyl-loaded EpoC (as in [19, 20]) followed by digestion, reverse-phase fractionation, and ESI/FTMS afforded apparent occupancies of 50% for the methyl-thiazole-methyl-

Table 1. Relative Ratios (% Occupancies) of Covalent Acyl Enzyme Intermediates Bound to EpoC-ACP under Various Experimental Conditions

Substrate	Holo	Methylmalonyl	Propionyl	β -keto	Methacryl ^a	Figure
methylthiazole-S-NAC	5%	ND ^b	20%	15%	60%	3D
methylthiazole-S-NAC (no NADPH)	20%	ND ^b	70%	10%	NA ^c	4A
methylpyridine-S-NAC	10%	ND ^b	20%	5%	65%	4B
phenylthiazolyl-S-NAC	35%	25%	10%	10%	20%	4C
2-hydroxyphenylthiazolyl-S-NAC	35%	15%	15%	15%	20%	4D
phenyloxazolyl-S-NAC	50%	20%	15%	5%	10%	4E
reconstitution (cysteine)	15%	ND ^b	20%	15%	50%	5C
reconstitution (serine)	60%	5%	35%	ND ^b	ND ^b	5D

^a Methacryl refers to condensed, reduced, and dehydrated product expected.

^b ND, species not detected.

^c NA, not applicable.

acrylyl-S-enzyme form (+165 Da), 15% β -ketoacyl, no detectable methylmalonyl, 20% propionyl (from methylmalonyl decarboxylation), and 15% holo, respectively (Figure 5C). L-serine can replace L-cysteine as a low V_{max} alternate substrate for EpoB [20]. To determine what level of covalent acyl enzymes accumulate on EpoC during L-serine utilization, reconstitution with acetyl-S-EpoA-ACP, L-seryl-S-EpoB, and methylmalonyl-S-EpoC yielded only the holo, methylmalonyl, and propionyl forms of the carrier peptide (Figure 5D; percent occu-

pancies are summarized in Table 1). Evidence for the accumulation of the methyl-oxazolyl-S-EpoC species was not obtained, with a detection limit of $\sim 0.5\%$ occupancy.

Discussion

Intermediate Quantification

In this ESI-based method, there are two effects that disperse MS signals of differentially modified forms of

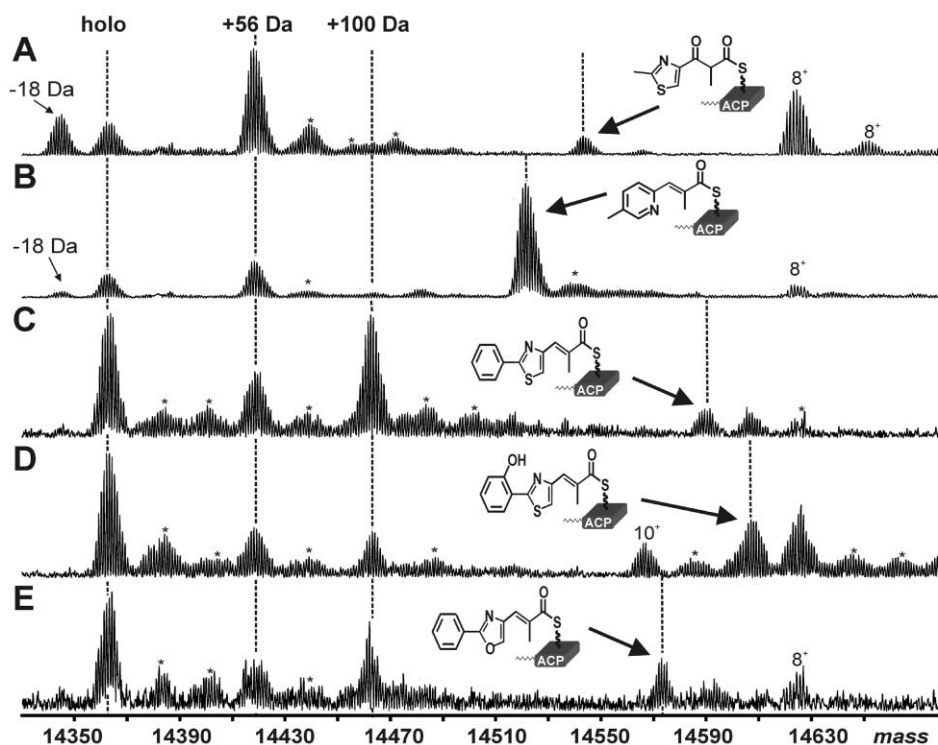


Figure 4. Loading and Mass Shifts on EpoC-ACP under Various Experimental Conditions and Indication of Respective Structures of Final Products and the Corresponding Isotopic Distributions

Loading and mass shifts on EpoC-ACP (1434–1469 m/z region; 10+ ions) under various experimental conditions; asterisks indicate species arising from artifactual addition of sodium (+22 Da), potassium (+38 Da), or oxidation of Met/Cys residues (+16 Da). Shown are incubation of EpoC with (A) methylmalonyl-CoA + methylthiazolyl-S-NAC in the absence of NADPH; (B) methylmalonyl-CoA + methylpyridine-S-NAC; (C) methylmalonyl-CoA + phenylthiazolyl-S-NAC; (D) methylmalonyl-CoA + 2-hydroxyphenylthiazolyl-S-NAC; and (E) methylmalonyl-CoA + phenyloxazolyl-S-NAC. The holo, propionyl (+56 Da), and methylmalonyl (+100 Da) forms are indicated within each panel along with respective structures of final products and corresponding isotopic distributions. Ratio information for intermediates is summarized in Table 1.

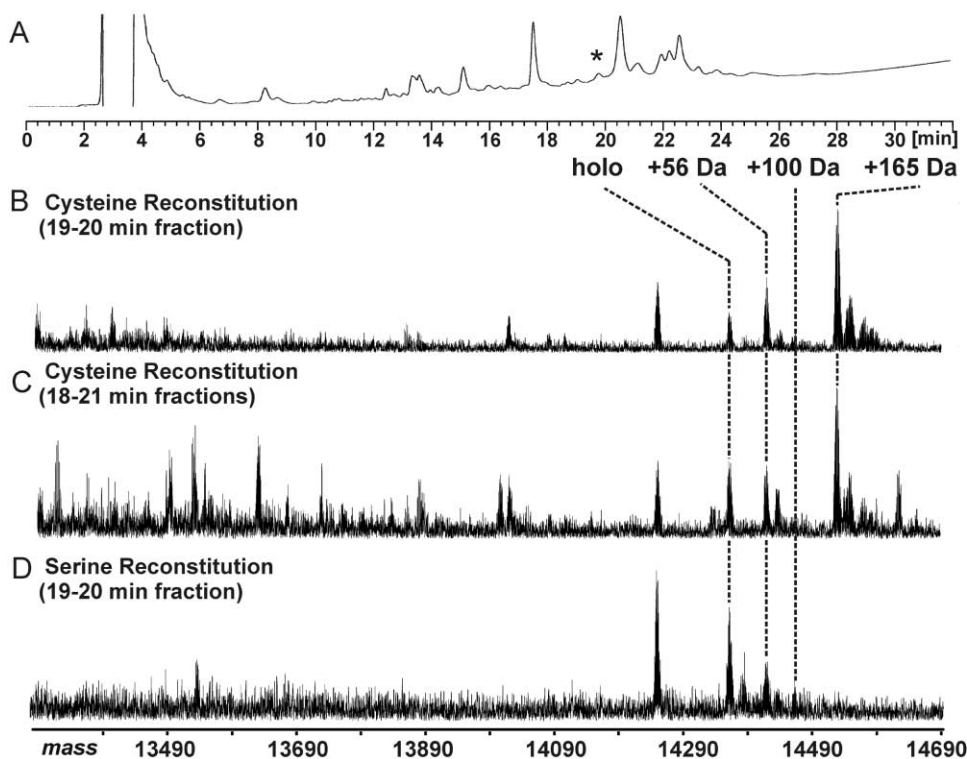


Figure 5. Reconstitution of EpoA-ACP, EpoB, and EpoC and Monitoring ACP of EpoC under Varying Experimental Conditions

Reconstitution of EpoA-ACP, EpoB, and EpoC; monitoring ACP of EpoC (1330–1470 m/z region; 10^+ ions).

(A) RPLC chromatogram ($\lambda = 220$ nm) of the fractionation of a limited proteolytic digestion of the reconstitution experiment with trypsin; asterisk indicates elution of peptide fragment of interest.

(B) FTMS of one individual fraction (19–20 min) from the reconstitution of acetyl-loaded EpoA-ACP, cysteinyl-loaded EpoB, and methylmalonyl-loaded EpoC.

(C) FTMS of physically mixed fractions (18–21 min) from the reconstitution of acetyl-loaded EpoA-ACP, cysteinyl-loaded EpoB, and methylmalonyl-loaded EpoC (note the increase in spectral complexity by combining fractions).

(D) FTMS of one fraction (19–20 min) from the attempted reconstitution of seryl-loaded EpoB, acetyl-loaded EpoA-ACP, and methylmalonyl-loaded EpoC.

carrier peptides. These include (1) fractionation of chemically distinct species by <2 min during chromatography and (2) multiple charge states observed in ESI-MS (e.g., 9^+ , 10^+ , 11^+ , etc.). In general, fractions containing different forms of the same carrier peptide are physically mixed (typically 2–3), and abundance information from multiple charge states is combined in the determination of relative ratios. For example, occupancies of 85% and 15% were obtained in the Figure 3A spectrum for the apo and holo forms of this carrier peptide (11^+ ions). These occupancies were found to be 95%/5% and 95%/5% for the 10^+ and 12^+ charge states of the same peptide, respectively (data not shown). By calculation of the mean over every charge state present, the occupancies determined are $\sim 90\%/10\%$.

The extent to which the occupancies determined by this study reflect the actual ratios on the intact enzyme EpoC depends primarily on the difference in ionization efficiencies between the peptide forms, which is a function of the molecular weight of the peptides upon which the covalent intermediates reside. For the 14 kDa peptide of Figure 3B, 98% of its mass does not change after the various substrates/intermediates are covalently tethered. Although systematic errors in these ratios are

possible, these would not obviate determination of substrate loading or kinetic profiles for NRPS/PKS synthesis due to the systematic reproducibility (within 5%) of the apo/holo ratios observed on different days for the same sample (data not shown). The apo/holo change is the most drastic structural change, with the most mechanistically meaningful ratios determined from species with very subtle structural differences. Hence, as a representative experiment to assess the reproducibility of the methodology presented, the methyl-thiazolyl-S-NAC experiment was performed in triplicate. The relative ratios of holo-enzyme and the thioester-bound intermediates were determined to be 0%, 20% (propionyl-), 20% (methylthiazolyl- β -ketobutyryl-), and 60% (methylthiazolyl-methacrylyl-), with a maximum standard deviation of 5.6% (see Supplemental Figure S1 available with this article online).

Observation of Enzyme-Bound Products and Reaction Intermediates

The distribution of the anticipated condensation products covalently docked on EpoC is comparable when the methyl-thiazole acyl donor is presented on the carrier protein of EpoB (Figure 1C) or if presented as an

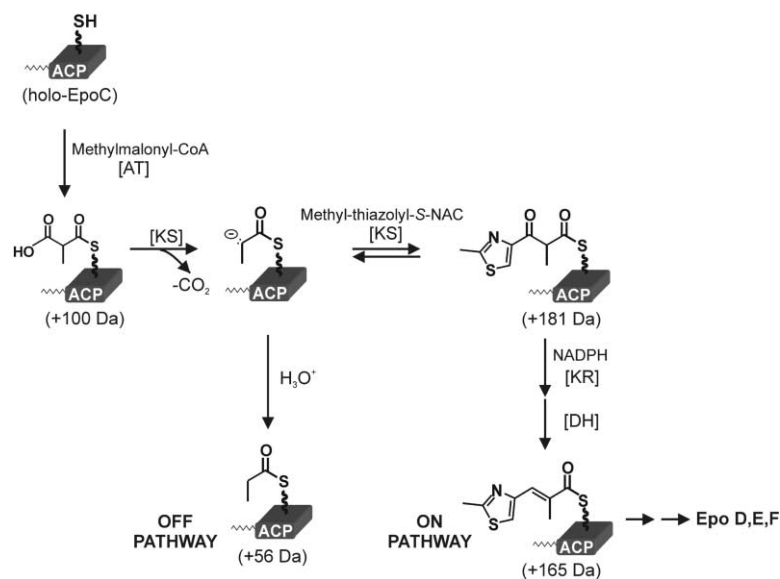


Figure 6. Branchpoint for Partitioning of Intermediates in which Nonproductive Decarboxylation, i.e., Formation of Propionyl-S-EpoC, Is Lowered by Efficient Processing of Immediate Condensation Products on the ACP Domain by the Ketoreductase and Dehydratase Domains

N-acetylcysteamine substrate (rows 1 and 7 of Table 1). Combination of acetyl-EpoA-ACP, cysteinyl-EpoB, and methylmalonyl-EpoC suggests that the two consecutive *trans* transfers of the growing acyl chain to the successive enzymes (EpoB, then EpoC) and subsequent C-C condensation with the appropriate extender unit is efficient. The 50%–60% occupancies of the final product (+165 Da, the methyl-thiazole-methylacrylyl-S-enzyme form) observed in this reconstitution experiment are comparable to results produced by incubating methyl-thiazolyl-S-NAC with methylmalonyl-EpoC, indicating a highly efficient association and transfer between reconstituted enzymes, EpoA, B, and C, as well as the effectiveness of the acceptance of acyl-S-NAC substrates as upstream donor mimics (see Table 1). However, as indicated by comparing the extent of processing (C-C formation, keto reduction, dehydration) in the various alternative S-NAC substrates, the identity of the acyl donor has a profound effect on the efficiency of intermediates accumulating on the ACP domain of the EpoC PKS module. These results begin to provide insights into how alternative substrates may exhibit deleterious effects in the multistep assembly lines. When considering rational reengineering of these systems, results from the S-NAC experiments combined with those from the serine reconstitution experiment must be considered carefully if PKS assembly lines are to be used to make new polyketide derivatives efficiently from unnatural substrates.

Observations from experiments in the presence and absence of acyl-S-NAC substrates (Figure 3C versus Figures 3D and 4A–4E) indicate that with no available upstream donor, substantially smaller amounts of holo-EpoC load the methylmalonyl-extender unit. However, in cases where suitable upstream donors are provided via S-NAC derivatives, a majority of the holo form of the ACP domain is converted to the fully processed methacrylyl-S-EpoC, an observation consistent with the possibility of crosstalk and/or regulation between the KS and AT domains to inhibit nonproductive loading

and decarboxylation of the methylmalonyl extender unit in the absence of an upstream donor.

The nonreduced β -ketoacyl intermediate (+181 Da), the initial product of C-C bond formation, was observed when NADPH was withheld from the reaction mixture (Figure 4A), correlating with previous radio-HPLC results from intermediate hydrolysis [19]. However, when comparing the apparent occupancy of the β -ketoacyl-S-EpoC (10%) relative to the fully processed methacrylyl-S-EpoC (+165) in the identical experiment that included NADPH (60%) (Figure 3D), it is notable that the enzyme exhibits quality control with respect to condensation versus downstream processing of the intermediates on the ACP domain. In the case with no NADPH present, downstream processing after C-C bond formation is not possible, and thus abundantly more enzyme (\sim 20% versus 70%) is in the unproductive decarboxylated/protonated form. Although the reduced amount of β -ketoacid released product observed in the HPLC analysis suggested that product formation might be less efficient in the absence of NADPH (unpublished observations), the relative ratios of reaction intermediates could only be obtained using mass spectrometry. Figure 6 postulates a mechanism to explain this phenomenon in which nonproductive decarboxylation is lowered by efficient processing of immediate condensation products on the ACP domain by the ketoreductase and dehydratase domains.

Uncoupled decarboxylation of a β -keto intermediate has also been observed in mass spectrometry studies of the yersiniabactin pathway. In this case, SAM was withheld to prevent bismethylation of the β -keto group to generate the final product [21]. Perhaps, when a β -keto thioester intermediate is bound to the ACP domain for a substantial period of time (on an enzymatic time scale), the KS domain can catalyze a reverse Claisen condensation. If the C₂ carbanion is then captured by protonation, this will result in formation of a propionyl group on the phosphopantetheinyl group on the ACP domain and the methylthiazole group on the

active-site cysteine of the KS domain. Efforts are currently underway to characterize the KS domain by mass spectrometry to test this hypothesis. In the full epothilone synthase assembly line, EpoA–F [17, 18], there are multiple β -ketoacyl-S-enzyme intermediates formed during chain elongation, and it remains to be seen how many may accumulate long enough for Claisen/retro-Claisen equilibration to occur.

The correlation between results from prior radiolabeling studies on released acids [19] and direct interrogation of peptide-bound intermediates is high. However, in prior work using radiolabels and HPLC, it was difficult to separate acetate, propionate, and malonate without losses. Moreover, basic hydrolysis conditions would likely result in nonenzymatic decarboxylation of the methylmalonyl group perturbing the methylmalonyl/propionyl ratio from that on the enzyme. Direct interrogation of thioester-bound acyl intermediates has provided insight into the ratio of productive and nonproductive intermediates that are either not stable enough to survive hydrolysis from the protein or cannot be separated easily. These studies have indicated that if the activity of a domain is knocked out—in this case, withholding NADPH to prevent action of the KR domain—net non-productive decarboxylation of the methylmalonyl substrate predominates. Furthermore, certain substrates are processed more efficiently through to the on pathway methacryl-S-EpoC acyl enzyme, while nonproductive decarboxylation or failure of EpoC to load the methylmalonyl substrate predominates for other substrates. This study therefore examines the flux of intermediates along the EpoC module of this multi-PKS assembly line and begins to probe how this flux can be perturbed. This information will have important ramifications when troubleshooting engineered PKS assembly lines to construct unnatural products.

Significance

Many drugs used in medicine today emanate from NRPS and PKS enzymes, and there is an ongoing vibrant effort to engineer these enzymes in order to generate novel natural products of significant biomedical import. Data from such MS-based studies will markedly assist this work by enabling efficient read-outs of substrate specificities and editing capabilities of large synthases/synthetases. While the basic “thio-template” mechanism of movement/growth of biosynthetic intermediates is established, the exact description of bound species and their transfer kinetics has not yet been at hand. Illumination of biosynthetic heterogeneity and substrate processing by 100–700 kDa proteins will both deepen our fundamental understanding of NRPS and PKS enzymes and facilitate their rational engineering.

Experimental Procedures

Enzyme Production and Purification

The constructs for EpoB, EpoC, and the ACP domain of EpoA were prepared as described previously [8, 19]. Deviations from the published EpoC purification protocol include the following: (1) cell lysate was allowed to bind in batch to the nickel-NTA resin for 3 hr at 4°C, and (2) resin was washed with 20 column volumes of lysis buffer

before the step gradient with lysis buffer containing increasing amounts of imidazole (10, 15, 30, 60, 100, and 500 mM).

Incubations for Intermediate Loading

A series of experiments was carried out in order to visualize and obtain semiquantifiable ratios of complex intermediates on the active-site-containing peptide of EpoC. Prior to substrate loading, the carrier domains of the enzymes were first phosphopantetheinylated in buffer (50 mM Tris [pH 7], 5 mM MgCl₂, 5 mM tris(2-carboxyethyl)phosphine hydrochloride [TCEP]) by incubating with Sfp and CoA-SH for 4 hr at 30°C in 90 μ l reactions.

Transfer from S-NAC Upstream Donor Mimics

Following priming with the phosphopantetheinyl cofactor, 33 μ M methylmalonyl-CoA, 1 mM NADPH (where applicable), and a large excess of the corresponding S-NAC (1–5 mM) (prepared as described previously [19, 22]) were added to a final volume of 100 μ l. After incubation at 30°C for 15 or 20 min, samples were immediately digested using the protocol described below.

In Vitro Reconstitution

Reconstitution of EpoA-ACP, EpoB, and EpoC required first loading the enzymes individually by incubation with their respective substrates (acetyl-CoA, L-cysteine or L-serine, and methylmalonyl-CoA) and other required substituents/cofactors, such as Sfp, ATP, and NADPH, as described above. Following loading, the enzymes were mixed in a 1.5:1:1 molar ratio, incubated 10 min or 30 min at 30°C, and immediately digested using the protocol described below.

Digestion Conditions

Limited proteolysis was performed by the addition of TPCK-treated trypsin (Promega) to 0.5 nmol of the target protein at protease to substrate ratios ranging from 1:5 to 1:10 in 50 mM NH₄HCO₃ (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 reverse-phase column (Phenomenex) with a linear gradient from 30%–65% ACN (0.1% TFA) for fractionation/desalting prior to FTMS analysis. Samples were lyophilized before resuspension in ESI buffer (49% H₂O, 49% MeOH, 2% formic acid) for FTMS analysis.

ESI/Q-FT Mass Spectrometry

Electrospray ionization (ESI) was used with a custom Q-FTMS instrument operating at 8.5 T [26]. The ions were directed through a heated metal capillary, skimmer, quadrupole, and multiple ion guides into the ion cell ($\sim 10^{-9}$ Torr) of the FTMS. Scans were acquired every 1 s, and data were stored with a MIDAS datastation [23] as 512 K data sets. Spectra were calibrated externally using bovine ubiquitin ($M_r = 8559.62$ 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.

Reported Masses

High-resolution mass spectrometry of large molecules results in isotopic distributions within the mass spectra, explanations of which have been described previously [24, 25]. Briefly, all molecular weights in this manuscript are reported as monoisotopic values, which refers 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 (as displayed in Figures 3–5) involved correlating the experimental monoisotopic molecular weights to the theoretical monoisotopic molecular weights for the enzyme intermediates, with a maximum error tolerance of 15 ppm (see Table S1). For example, the experimental molecular weights, theoretical molecular weights (in parentheses), and corresponding error for the data displayed in Figure 3D are as follows: holo, 14354.5 Da (14354.4 Da) 7 ppm; propionyl, 14410.4 Da (14410.4 Da) 0 ppm; β -keto, 14535.5 Da (14535.4 Da) 7 ppm; and methacryl, 14519.5 Da (14519.4 Da) 7 ppm.

Intermediate Quantification

Two effects disperse MS signals of differentially modified forms of carrier peptides: (1) fractionation of chemically distinct species during chromatography and (2) multiple charge states observed in

ESI-MS (e.g., 9⁺, 10⁺, 11⁺, etc.). In order to account for this dispersion, HPLC fractions containing different forms of the same carrier peptide are physically mixed (typically 2–3) before ESI/FTMS analysis, and abundance information from every charge state observed of a particular species in the broadband spectrum is considered in the determination of relative ratios. The data displayed in Figures 3–5 indicate the mass shifting region of one particular charge state, whereas the relative ratio information reported for a particular experiment has accounted for every charge state present in the mass spectrum. A more detailed explanation and further evaluation of this quantification process is presented at the beginning of the discussion section.

Acknowledgments

We acknowledge the generous support from the University of Illinois and an NSF Fellowship to L.M.H. N.L.K. and C.T.W. acknowledge the National Institutes of Health (GM 067725) and (GM 20011), respectively, and S.E.O. acknowledges an Irving Sigal postdoctoral fellowship (American Chemical Society).

Received: September 29, 2003

Revised: December 9, 2003

Accepted: December 11, 2003

Published: March 19, 2004

References

- Stein, T., Vater, J., Kruff, V., Wittmann-Liebold, B., Franke, P., Panico, M., Dowell, R.M., and Morris, H. (1994). Detection of 4'-phosphopantetheine at the thioester binding site for L-valine of gramicidinS synthetase 2. *FEBS Lett.* **340**, 39–44.
- Stein, T., Vater, J., Kruff, 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**, 15428–15435.
- Mootz, H.D., and Marahiel, M.A. (1997). Biosynthetic systems for nonribosomal peptide antibiotic assembly. *Curr. Opin. Chem. Biol.* **1**, 543–551.
- Khosla, C. (1997). Harnessing the biosynthetic potential of modular polyketide synthases. *Chem. Rev.* **97**, 2577–2590.
- Walsh, C.T., Chen, H., Keating, T.A., Hubbard, B.K., Losey, H.C., Luo, L., Marshall, C.G., Miller, D.A., and Patel, H.M. (2001). Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr. Opin. Chem. Biol.* **5**, 525–534.
- Cane, D.E., Walsh, C.T., and Khosla, C. (1998). Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282**, 63–68.
- Leenders, F., Vater, J., Stein, T., and Franke, P. (1998). Characterization of the binding site of the tripeptide intermediate D-phenylalanyl-L-prolyl-L-valine in gramicidin S biosynthesis. *J. Biol. Chem.* **273**, 18011–18014.
- 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.
- Kallow, W., Neuhof, T., Arezi, B., Jungblut, P., and Döhren, H.V. (1997). Penicillin biosynthesis: intermediates of biosynthesis of δ -L- α -aminoadipyl-L-cysteinyl-D-valine formed by ACV synthetase from *Acremonium chrysogenum*. *FEBS Lett.* **414**, 74–78.
- Carreras, C.W., Gehring, A.M., Walsh, C.T., and Khosla, C. (1997). Utilization of enzymatically phosphopantetheinylated acyl carrier proteins and acetyl-acyl carrier proteins by the actinorhodin polyketide synthase. *Biochemistry* **36**, 11757–11761.
- Cane, D.E., Shim, J.H., Xue, Q., and Fitsimons, B.C. (1995). Trichodiene synthase. Identification of active site residues by site-directed mutagenesis. *Biochemistry* **34**, 2480–2488.
- 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**, 217–224.
- 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**, 385–400.
- Kelleher, N.L., Costello, C.A., Begley, T.P., and McLafferty, F.W. (1995). Thiaminase I (42 kDa) heterogeneity, sequence refinement, and active site location from high-resolution tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **6**, 981–984.
- Meng, F., Cargile, B.J., Patrie, S.M., Johnson, J.R., McLoughlin, S., and Kelleher, N.L. (2002). Processing complex mixtures of intact proteins for direct analysis of intact proteins by mass spectrometry. *Anal. Chem.* **74**, 2923–2929.
- Bollag, D.M., McQueney, P.A., Zhu, J., Hensens, O., Koupal, L., Liesch, J., Goetz, M., Lazarides, E., and Woods, C.M. (1995). Epothilones, a new class of microtubule-stabilizing agents with a taxol-like mechanism of action. *Cancer Res.* **55**, 2325–2333.
- Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C., and Julien, B. (2000). Cloning and heterologous expression of the epothilone gene cluster. *Science* **287**, 640–642.
- Molnár, I., Schupp, T., Ono, M., Zirkle, R.E., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D.D., et al. (2000). The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. *Chem. Biol.* **7**, 97–109.
- 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.
- Schneider, T.L., Walsh, C.T., and O'Connor, S.E. (2002). Utilization of alternate substrates by the first three modules of the epothilone synthetase assembly line. *J. Am. Chem. Soc.* **124**, 11272–11273.
- 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.
- Schneider, T.L., Shen, B., and Walsh, C.T. (2003). Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry* **42**, 9722–9730.
- 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.
- 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.
- McLafferty, F.W. (1994). High-resolution tandem FT mass spectrometry above 10 kDa. *Acc. Chem. Res.* **27**, 379–386.
- Patrie, S.M., Quinn, J.P., Whipple, D., Blakney, G., Marshall, A.G., Hendrickson, C.L., and Kelleher, N.L. (2001). Poster *WPM263*, 49th Amer. Soc. Mass Spectrom. Conf. on Mass Spectrom. & Allied Topics, Chicago, IL, May 2001.