ELUCIDATION OF CYANOBACTIN BIOSYNTHESIS

by

John Andrew McIntosh

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Medicinal Chemistry

The University of Utah

December 2011

STATEMENT OF DISSERTATION APPROVAL

The dissertation of John Andrew McIntosh

has been approved by the following supervisory committee members:

Eric Schmidt	, Chair	29-Aug-2011 Date Approved
Grzegorz Bulaj	, Member	29-Aug-2011 Date Approved
Darrell Davis	, Member	29-Aug-2011 Date Approved
Chris Ireland	, Member	29-Aug-2011 Date Approved
Michael Kay	, Member	29-Aug-2011

and by Darrell Davis , Chair of

the Department of Medicinal Chemistry

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

Cyanobactins are N-to-C macrocyclic peptides that contain diverse modifications such as heterocyclization of Cys, Ser, or Thr, and isoprenylation of Ser, Thr, and Tyr. Although the above can be inferred to occur based on the final products, none of the enzymatic steps en route to cyanobactin biosynthesis had been characterized prior to this work. Indeed, until very recently, nothing at all was known about cyanobactin biosynthesis. Only after cloning and sequencing of the genetic elements required for cyanobactin biosynthesis was their biosynthetic origin deduced. Surprisingly, these complex natural products are biosynthesized by extensive posttranslational modification of ribosomally synthesized precursor peptides. As noted, these precursors are initially synthesized on the ribosome. Following ribosomal synthesis, various modifying enzymes carry out posttranslational modification of the aforementioned amino acids, as well as proteolysis of the precursor peptide to liberate 6-12 amino acid peptidyl groups, which are then macrocyclized. However, the manner in which the disparate genetic components required for cyanobactin biosynthesis functioned enzymatically to create these highly diverse and medicinally interesting compounds was unknown prior to this work. Herein the results of several studies that elucidate the steps en route to cyanobactin biosynthesis are described. The characterized steps include: N- and C-terminal proteolysis, macrocyclization, heterocyclization of Cys, Ser, and Thr, and prenylation of Ser, Thr and Tyr.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

ACN	acetonitrile
ADP	adenosine diphosphate
AEBSF	4-(2-aminoethyl) benzenesulfonyl fluoride
AIP	autoinducing peptide
AMP	adenosine monophosphate
ATP	adenosine triphosphate
BME	beta-mercaptoethanol
C domain	condensation domain
CD	circular dichroism
CHCA	α-cyano-4-hydroxycinnamic acid
COSY	Correlation Spectroscopy
Cy domain	cyclization domain
DAD	diode array detection
DIPEA	N,N-diisopropylethylamine
DMAPP	dimethylallyl pyrophosphate
DMF	dimethylformamide
DNFB	1-fluro-2,4-dinitrobenzene
DNP	dinitrophenyl
DTT	dithiothreitol

EDTA	ethylenediaminetetraaceticacid
FAD	Flavin adenine dinucleotide
fMet	N-formylmethionine
FMN	Flavin mononucleotide
FPLC	Fast protein liquid chromatography
FT-ICR	Fourier-transform ion cyclotron resonance
GTP	guanosine triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high-pressure liquid chromatography
HSQC	heteronuclear quantum coherence
ICP-OES	Inductively coupled plasma optical emission spectrometry
IPTG	isopropyl-b-D-1-thiogalactopyranoside
LB	Luria-Bertani
LC	liquid chromatography
MALDI	matrix-assisted laser desorption ionization
MBP	maltose binding protein
μL	microliter
μM	micromolar
mM	millimolar
MS	mass spectrometry

MTBE	methyl tert-butyl ether
NMP	N-methylpyrrolidone
NMR	nuclear magnetic resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
NRP	nonribosomal peptide
PCR	polymerase chain reaction
PDI	protein disulfide isomerase
PMSF	phenylmethanesulfonyl fluoride
PQQ	pyrroloquinoline quinone
PT	prenyltransferase
PyBOP	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
RP	ribosomal peptide natural product
SAM	S-adenosylmethionine
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TCEP	tris(2-carboxyethyl) phosphine
TE	thioesterase
TOF	time-of-flight
UHPLC	ultra high-pressure liquid chromatography

CHAPTER 1

INTRODUCTION

1.1 Literature Review

1.1.1 Natural Products

Natural products constitute a very large fraction of approved drugs and drug leads. The great utility of natural products stems from the diversity of structures and the incredible range of bioactivities found in these molecules. Because natural products are so incredibly vast in scope, many different classes have been delineated, which include but are not limited to: alkaloids, isoprenoids, polyketides, nonribosomal peptides, and ribosomal peptides. Of particular relevance here are the nonribosomal and ribosomal peptide natural products. The chemical logic that underlies molecules of both nonribosomal and ribosomal classes is identical. To wit, both classes are polymers of amino acid monomers that have been joined to form peptide chains linked through amide bonds. Although at a chemical level there are many similarities between ribosomal and nonribosomal peptides, biosynthetically they are quite different.

1.1.2 Nonribosomal and Ribosomal Peptide Natural Products

Nonribosomal peptides (NRPs) derive from extraordinarily large biosynthetic assembly lines, in which roughly 75 kDa of enzyme are required to activate and incorporate a single amino acid. Among NRPs, much of the chemical diversity is created at the level of the monomers that are incorporated. Notably, these monomers are not limited by the 20 canonical amino acids. Added diversity can be introduced by tailoring domains, which may or may not be localized within the main assembly line.

In contrast, ribosomal peptide natural products (RPs), as the name would suggest, derive from ribosomal biosynthesis. Unlike NRPs, which encompass all peptides

biosynthesized on nonribosomal peptide synthetases, defining RPs is not so straightforward. That is, nearly all proteins and peptides in nature are synthesized on the ribosome, and yet the vast majority of ribosomally-synthesized polypeptides are not considered RPs. Here RPs will be defined as posttranslationally modified peptides that function as secondary metabolites. In turn, secondary metabolites have been defined as molecules that are nonessential for the survival of an organism, but which nevertheless confer some benefit.(*1*) As noted, RPs share a common biosynthetic framework which centers on the posttranslational modification of a ribosomally derived precursor peptide. Interestingly, although the precursor peptides themselves could be seen as passive players in the course of biosynthesis, many conserved features can frequently be observed among precursor peptides from diverse groups (Figure 1.1). In particular, many precursors frequently contain N-terminal signal sequences to direct export of posttranslationally modified natural products. Export is especially crucial for larger, more sparsely modified RPs, which would otherwise be unable to cross the membrane of the producing cell, and thus be prevented from reaching their molecular targets. Other conserved features include a so-called 'leader sequence', which is a highly conserved N-terminal peptide sequence that is thought to serve as a binding handle for modifying enzymes, but which itself is not a substrate for modification. Moving from the N- to the C-terminus, precursor peptides will frequently contain a proteolytic recognition site, which in several studied examples takes the form of a double-glycine motif. Proteolytic cleavage of the signal sequence and leader peptide from the core natural product peptide sequence is presumably nearly always required for activity, and indeed among studied examples, this hypothesis has been supported. Next is the so-called 'core peptide', whose amino acid

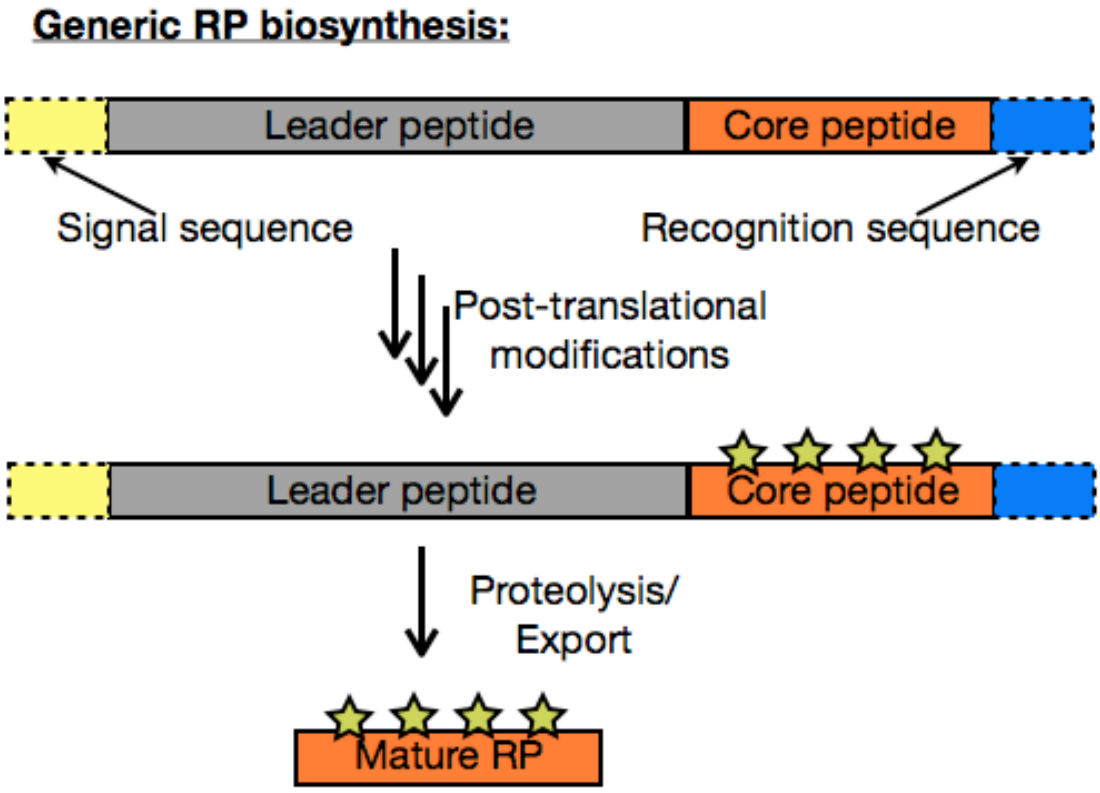


Figure 1.1 A generic scheme depicting ribosomal peptide biosynthesis. Stars denote posttranslational modifications.

residues compose the bulk of the final natural product. Lastly, and somewhat infrequently, precursor peptides may contain a C-terminal sequence, which is also cleaved from the final natural product. Interestingly, RPs that do contain C-terminal recognition sequences are frequently found to be macrocyclic in nature.

Early work on RPs was largely limited to a few closely related molecules. In particular, some of the earliest work was done on the lantibiotics, which are named for the presence of lanthionine crosslinks that constrain the peptide backbone. Chemically, lanthionine bonds are simply thioethers derived from the condensation of a cysteine thiol with a dehydrated serine or threonine residue. Of the lantibiotics, the oldest and most well-characterized is nisin, whose antibacterial activity was discovered in 1928.(*2, 3*) Despite this early discovery, the chemical structure, and the genetic basis for nisin were not revealed until many years later.(*3, 4*) Overall, lantibiotic biosynthesis proceeds first via enzymatically catalyzed dehydration of serine and threonine residues, and then by regio- and stereoselective attack of cysteine thiols on the dehydrated residues. Although the aforementioned sequence was hypothesized to occur for many years,(*3, 5*) a biochemical demonstration of this process was not achieved until relatively recently.(*6*) As noted, lantibiotics are chiefly modified via introduction of lanthionine crosslinks. However, other modifications among the lantibiotics were known for several years, including the presence of standalone dehydrated serine and threonine (i.e., dehydrations that do not proceed to form lanthionine bonds), as well as thioethers that cross link cysteine thiols to α-carbons,(*7*) and other unusual modifications such as aminovinyl cysteine.(*8-10*)

Other pioneering work on RPs focused on microcins, which are defined as low-molecular weight antimicrobial peptides biosynthesized by *Escherichia coli*. Of these only microcins B17 and C7 were long known to be posttranslationally modified.(*11-13*) Microcin B17 is predominantly modified by heterocyclization of cysteine and serine residues to yield oxazoles and thiazoles, some of which occur in tandem. Microcin C7 is a mimick of aspartyl adenylate, and as such the mature form contains only a single amino acid, aspartate, which is conjugated to a modified adenosine nucleotide at its carboxy terminus.

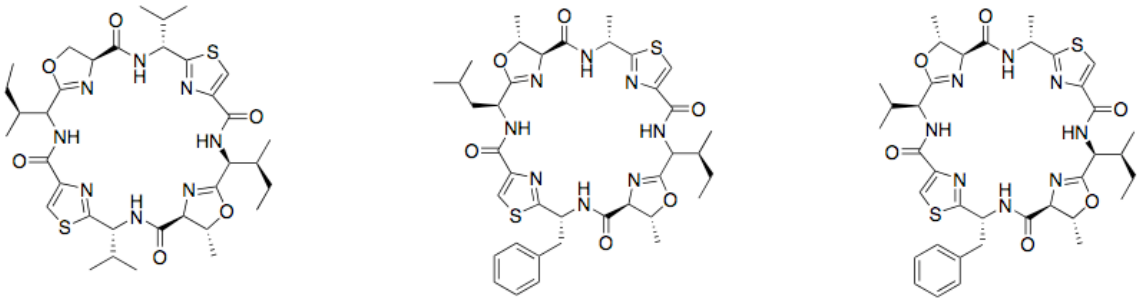
Of the above mentioned RPs, only microcin C7 could plausibly be synthesized via a nonribosomal pathway. In contrast, the lantibiotics and microcin B17 are such large molecules (30-40 residues) that their synthesis by nonribosomal peptide synthetases (which typically require ~75 kDa of enzyme per amino acid incorporated) would be grossly inefficient from an energetic perspective, and contribute relatively little to the overall structural diversity (since the majority of the residues in each case are derived from simple proteinogenic L-amino acids).

1.1.3 Biosynthetic Origin of the Patellamides

Overall, in the 1990s and early 2000s, the possibilities for structural modification of ribosomally derived peptides were underestimated. Likewise, it was far better appreciated how nonribosomal peptide synthetases could produce highly derived peptide-like molecules. In light of this intellectual context, it is unsurprising that most complex peptide-like molecules were reflexively assigned as deriving from nonribosomal biosynthesis.(*14*) In particular, one class whose biosynthetic origin was frequently

assigned as nonribosomal were the patellamides (Figure 1.2). The patellamides were initially isolated from marine animals known as ascidians, more commonly referred to as 'sea squirts.'(*15-18*) These molecules possessed not only intruiging bioactivities, but also interesting chemical features as well. To wit, patellamides are small, (6-8 amino acids) circular peptides that contain heterocycles derived from cysteine, serine and threonine, as well as D-amino acids.(*19*) Although there was much speculation as to their origin, one hypothesis held that they derived from a nonribosomal biosynthesis.(*20*) This hypothesis was based upon (1) the presence of a macrocyclic framework, which is common among NRPs, but had not been previously observed in RPs (2) the presence of multiple heterocycles, of varying oxidation states (3) the presence of several D-amino acids residues in known cyanobactins. With regard to (2) above, although microcin B17 contains heterocycles, many more NRPs were known containing this modification. Moreover, the variability in oxidation states observed in patellamides was not observed in microcin B17, and itself suggests a domain-swapping mode of evolution, which is thought to obtain among NRPs.

Another mystery surrounding the patellamide family of compounds had to do with the organism responsible for their biosynthesis. Given that these compounds were not found to be localized within any symbiont, it was natural to hypothesize that the animals themselves might have been the relevant producers.(*21*) Had this hypothesis been supported, these would doubtless have been some of the most complex natural products known to be produced by higher eukaryotes. An alternative hypothesis held that a bacterial symbiont within the ascidian might biosynthesize the patellamides, perhaps to



Patellamide A Patellamide B Patellamide C

Figure 1.2 Patellamides A-C

the benefit of their animal host. Intriguingly, several patellamide-containing ascidians were already well-known to contain an abundance of photosynthetic microorganisms—cyanobacteria. In particular, one species of cyanobacteria, *Prochloron didemni* had been found to be particularly abundant in several ascidians. Additionally, there was evidence for a symbiotic relationship between the *Prochloron* and its ascidian hosts, as the animals live in nutrient limited conditions, and *Prochloron* had been shown to supply them with much of their carbon requirements. Thus, given the already extant symbiotic relationship between *Prochloron* and their ascidian hosts, it was not unreasonable to hypothesize that an added dimension to this symbiotic relationship might revolve around the production of natural products, such as the patellamides, which given their cytotoxicity,(*17*) could potentially serve to deter predation (though it must be said that to impose purpose on natural products is to venture into epistemologically murky waters; that is, although the production of natural products is doubtless favorable under certain conditions, it is difficult to establish with a high degree of certainty exactly which selection pressures promote specific natural products).

Remarkably, the *Prochloron* hypothesis was confirmed in 2005, with the isolation of a gene cluster that was shown via heterologous expression experiments to be sufficient for patellamide production in *E. coli*.(*22*) This work also demonstrated, again to great surprise, that the patellamides derived from a ribosomal mode of biosynthesis, and were thus RPs (Figure 1.3). That the patellamides were RPs could be deduced by inspection of the gene cluster alone. Out of six open-reading frames (*A*-*G*), one—*patE*—contained near its C-terminus two 8-amino acid sequences that (after accounting for posttranslational modification) were identical with the sequences found in two natural

products, patellamides A and C (Figure 1.3). In light of the work done on the microcins and lantibiotics, the following biosynthetic hypothesis could be proposed: (1) PatE, a 70 amino acid peptide, is synthesized ribosomally (2) posttranslational modifications such as heterocyclization of cysteine, serine and threonine occur, yielding a heterocycle-containing PatE (3) oxidation of thiazoline and/or oxazoline affords thiazoles and/or oxazoles (4) proteolytic cleavage frees the natural product encoding sequences (5) N-to-C terminal macrocyclization yields the circular structures (6) nonenzymatic epimerization of α-carbons next to thiazole or oxazole (which is known to be facile) affords the final isolated structures (Figure 1.4). Supporting this hypothesis was the presence of several open reading frames clustered with *patE*, some of which had bioinformatically predictable functions consistent with the above biosynthetic steps (*vide infra*).

Further work then showed that several other families of macrocyclic peptides isolated from marine animals, such as the patellins and lissoclinamides, were also RPs (Figure 1.5). Indeed patellamide-like pathways were found to be so widespread within cyanobacteria (not just *Prochloron*) and encompassed so many previously isolated compounds, all with divergent compound family names, that in 2008, a overall group nomenclature was proposed, leading to the umbrella term ‘cyanobactins’. which describes any homologous pathway to macrocyclic RPs.(*23, 24*) Quite interestingly, in several cases the gene clusters responsible for cyanobactin biosynthesis were virtually identical to that initially isolated for the patellamide family. Consequently, it became apparent that several forms of molecular evolution were occurring in *Prochloron* to

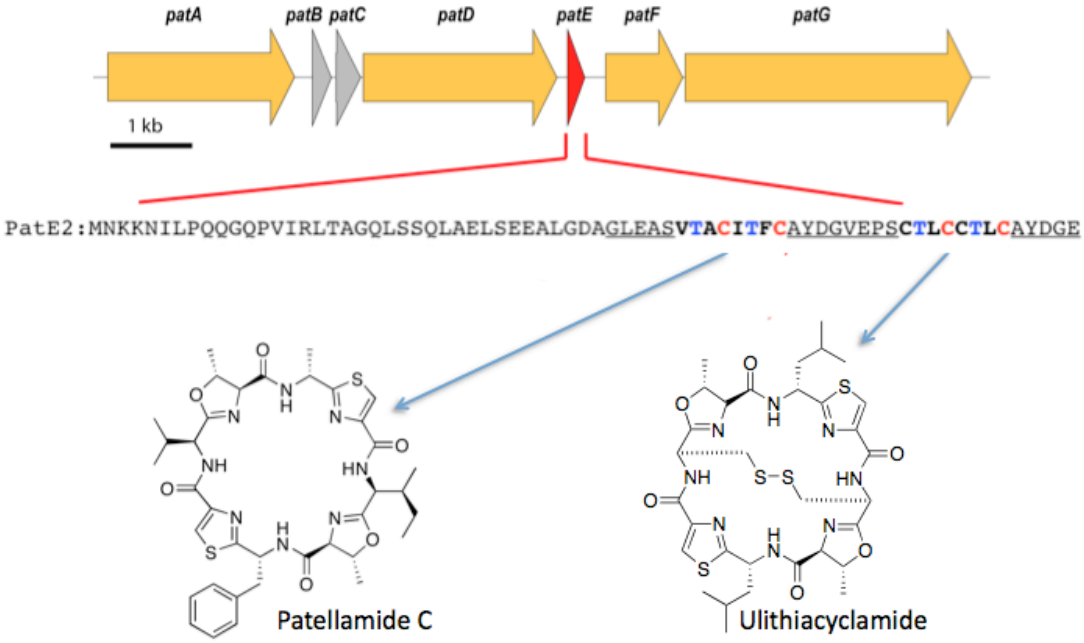


Figure 1.3 The patellamide pathway. Top: gene cluster isolated in 2005 to the patellamides. Middle: patellamide precursor peptide, highlighted in bold and with modified residues colored blue or red are the modified 'cassettes' that encode compounds patellamide C and ulithiacyclamide. Bottom: patellamide C and ulithiacyclamide.

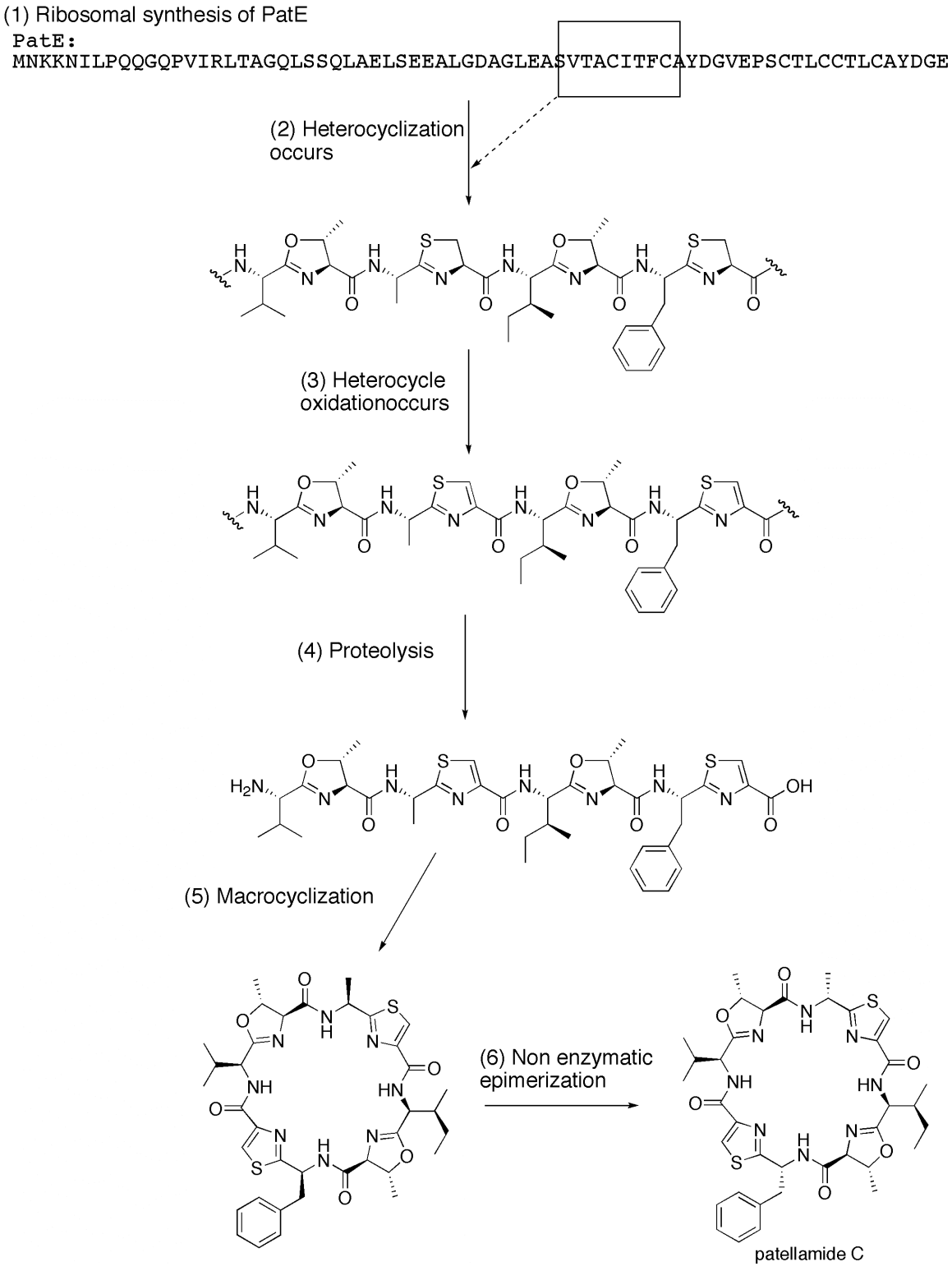


Figure 1.4 Initial proposed biosynthetic scheme for patellamide C

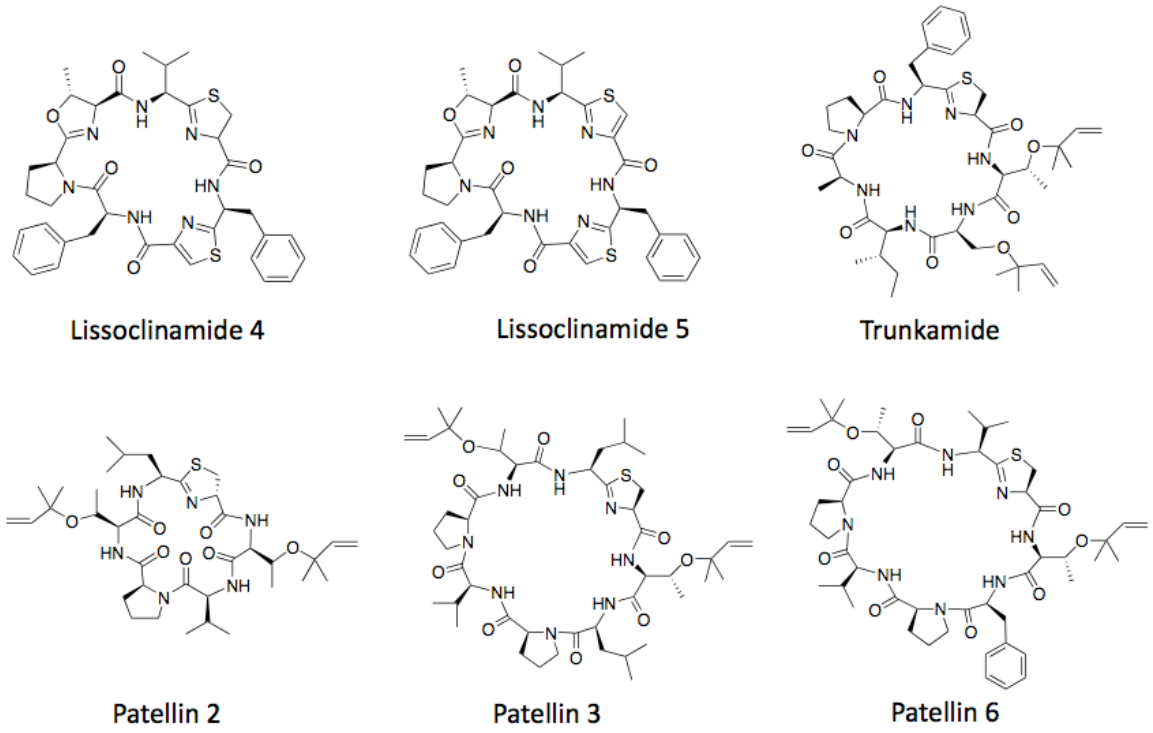


Figure 1.5 Lissoclinamides and patellins.

afford these different metabolites. First, without significant change to any predicted biosynthetic gene, point mutagenesis of the natural product encoding 'cassettes' within *patE* was occurring. The occurrence of large numbers of point mutations within the natural product encoding cassettes leads to the hypothesis that the enzymes responsible for posttranslational modification of the precursor peptide are broadly substrate tolerant.(*25*) Second, whole pathways were recombining to yield natural products with both altered cassette sequences and novel posttranslational modifications (Figure 1.6). In particular, comparison of the patellamide and trunkamide pathways provided a very clear example whereby recombination could afford a new set of posttranslational modification.(*24*)

As noted, the patellamide family of compounds contain heterocycles derived from cysteine, serine and threonine. In patellamide family metabolites, between three and four heterocycles per compound of variable oxidation states are frequently observed. In contrast, the trunkamide family of compounds typically contains only a single cysteine-derived heterocycle per compound, and only in the unoxidized state (thiazoline). At serine and threonine, in lieu of the oxygen-based heterocycles found in the patellamide family, trunkamide (*tru*) pathway metabolites contain 5-carbon alkylations of oxygen on these residues. Based upon a structural inspection, these alkylations almost certainly derive from dimethylallyl pyrophosphate (DMAPP), representing a nearly unprecedented intersection of RPs and isoprene natural products.(*26*) What is most interesting about the chemical differences in the trunkamide and patellamide pathway metabolites is that these chemical differences are obviously correlated with a recombination event that occurred between the *D* and *G* genes of the pathway.(*24*) This recombinational alteration leads to

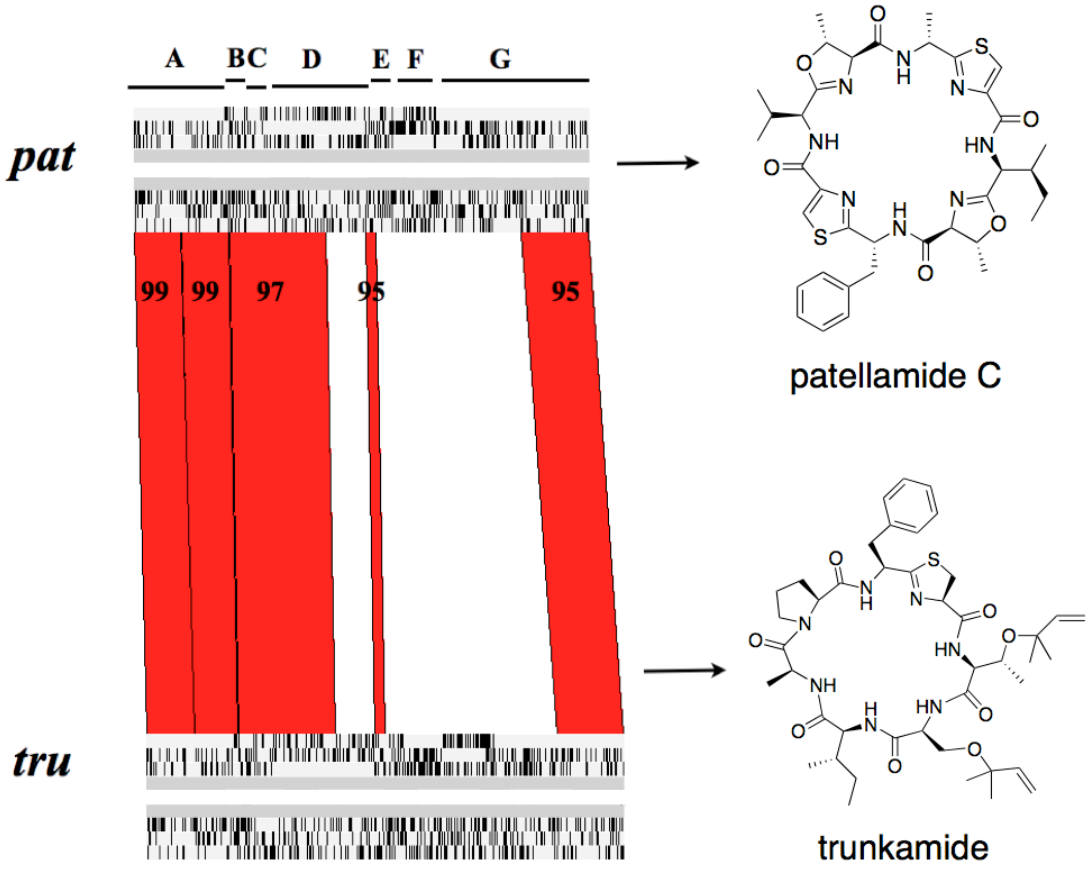


Figure 1.6 Alignment of *pat* and *tru* pathways. Red portions denote regions of high homology. Discontinuity suggests recombination event. Adapted from (24).

several differences between these two pathways, which are 95-99% identical otherwise. In particular, the C-termini of PatD and TruD and N-termini of PatG and TruG are highly divergent, while PatF is duplicated in the *tru* pathway to give TruF1 and TruF2, all of which are only about 50% identical to one another. Observation of this large discontinuity in the conservation of the pathways can readily be invoked to explain the chemical differences in the product. Further, this allowed for the narrowing of biosynthetic hypothesis concerning the differences in serine and threonine modification between *pat* and *tru* pathways. That is, whatever the enzymatic cause for the chemical changes in products, it most likely to be contained within PatD/TruD, PatF/TruF1/F2, or PatG/TruG.

1.2 Research Objectives

Overall our goals were first to define the enzymatic causes of posttranslational modification and, second to elucidate the order of enzymatic reactions *en route* to the cyanobactins. Scientifically, the pursuit of these goals could be justified given the widespread occurrence and structural novelty of cyanobactin natural products. Practically, the heterologous synthesis of extant and novel cyanobactin molecules in *E. coli* constituted a holy grail for natural products based drug discovery, and a more complete biochemical understanding of cyanobactin biosynthesis could be expected to aid these efforts. Additional background to the fascinating world of RPs are presented in chapter 2 along with a rich set of comparisons to NRPs.(*26*)

1.2.1 Specific Aim I

Perhaps the most clear-cut roles to be proposed based on bioinformatic analyses for any enzymes within the *pat* or *tru* pathways were for PatA/TruA and PatG/TruG.(*22*) Both enzymes could be related bioinformatically to subtilisin-like proteases. Since the peptide sequences that are modified to produce cyanobactins are flanked at their N- and C-termini by peptide recognition sequences, it was clear that two proteolytic events would be required to liberate them from that context. Consequently, the presence of two predicted proteases within cyanobactin gene clusters was highly suggestive of a role for these genes in proteolysis. Thus, early biochemical experiments focused on *in vitro* characterization of the proteolytic activities of PatA/TruA and PatG/TruG.(*27, 28*) The results of these experiments are presented in Chapters 3 and 4, and showed firstly that PatA/TruA were carrying out proteolysis N-terminal to cassette sequences, and secondly that PatG/TruG were carrying out proteolysis C-terminal to cassette sequences in tandem with macrocyclization. Additionally, we found that PatG possesses an astonishingly relaxed substrate selectivity and can macrocyclize a wide variety of analogs.

1.2.2 Specific Aim II

Although experiments on the proteolytic enzymes were conducted first given the relative obviousness of the hypotheses, in terms of biosynthesis, one might hypothesize based upon earlier work on the lantibiotics and microcin B17 that many of the posttranslational modifications, such as heterocyclization and prenylation would occur on the full-length intact precursor peptide.(*29-33*) Drawing further parallels to the groundbreaking microcin B17 work, it was possible to transitively relate PatD/TruD to

two enzymes McbD and McbB of the microcin B17 pathway, which were shown to participate in heterocyclization of cysteine and serine residues.(*29*) Consequently, our primary hypothesis as to the enzymatic cause of heterocyclization held that PatD/TruD were most likely to be responsible. Alternatively, another possibility was that PatD/TruD might only be responsible for cysteine heterocyclization, while in the patellamide pathway, PatF was responsible for heterocyclization of serine and threonine. This alternative hypothesis was based on the differences in heterocyclization patterns between the *pat* and *tru* pathways, and the dissimilarity of PatF and TruF1/F2 genes.(*24*) Still further support for this hypothesis was found in a cyanobactin pathway from the free-living cyanobacterium, *Trichodesmium erythraeum*, which led only to Cys heterocycles, and lacked a PatF homolog.(*34*) The results of these experiments are presented in Chapters 5 and 6, and showed that PatD/TruD are solely responsible for the differences in heterocyclization between the *pat* and *tru* pathways.

1.2.3 Specific Aim III

Finally, having defined all other posttranslational modifications leading to the cyanobactins, we set out to reveal the enzymatic cause of isoprenylation of serine and threonine in the trunkamide pathway. As noted above, only PatD/TruD, PatF/TruF1/F2, and PatG/TruG were significantly different between the *pat* and *tru* pathways. However, since by this time we had already assigned enzymatic roles to PatD/TruD and PatG/TruG,(*28, 35-37*) the most reasonable hypothesis left to us was that TruF1/F2 were the relevant enzymes for prenyltransfer, notwithstanding the fact that neither TruF1 nor TruF2 were sequence similar to any characterized prenyltransferase, nor even to any

other characterized enzyme. However, in our attempts to test this hypothesis, we were stymied by the insolubility of TruF1 and TruF2 upon overexpression in *E. coli*. A further difficulty was that although we recognized that earlier work on RPs generally suggested that the precursor peptide would be the substrate for posttranslational modification, given the complexity of cyanobactin biosynthesis it was nevertheless ineluctable that some other biosynthetic intermediate could be the relevant substrate for prenyltransfer. To sail between the proverbial Scylla (which here can be taken to represent protein insolubility) and Charybdis (which here represents our profound ignorance of the prenyltransferase substrate), we turned to a wholly uncharacterized cyanobactin pathway from *Lyngbya aestuarii*, known as the *lyn* pathway. This pathway had previously been predicted to carry out isoprenylation of tyrosine residues within cyanobactins. Thus, possible substrates for each biosynthetic step were synthesized and tested for acceptability as substrates for prenyltransfer. The results of these studies are presented in chapter 7, and showed that the TruF1 homologue in the *lyn* pathway, LynF carried out prenylation of tyrosine.

1.3 References

1. Williams, D. H., Stone, M. J., Hauck, P. R., and Rahman, S. K. (1989) Why are secondary metabolites (natural products) biosynthesized?, *J. Nat. Prod. 52*, 1189-1208.

2. Rodgers, L. A. (1928) The inhibiting effect of *Streptococcus lactis* on *Lactobacillus bulgaris.*, *J. Bacteriol. 16*.

3. Hansen, J. N. (1993) Antibiotics synthesized by posttranslational modification, *Annu. Rev. Microbiol. 47*, 535-564.

4. Gross, E., and Morell, J. L. (1971) Structure of nisin, *J. Am. Chem. Soc. 93*, 4634-4635.

5. Ingram, L. (1970) A ribosomal mechanism for synthesis of peptides related to nisin, *Biochimica et biophysica acta 224*, 263-265.

6. Xie, L., Miller, L. M., Chatterjee, C., Averin, O., Kelleher, N. L., and van der Donk, W. A. (2004) Lacticin 481: in vitro reconstitution of lantibiotic sythetase activity. , *Science 303*, 679-681.

7. Kawulka, K. E., Sprules, T., Diaper, C. M., Whittal, R. M., McKay, R. T., Mercier, P., Zuber, P., and Vederas, J. C. (2004) Structure of subtilosin A, a cyclic antimicrobial peptide from Bacillus subtilis with unusual sulfur to alpha-carbon cross-links: formation and reduction of alpha-thio-alpha-amino acid derivatives, *Biochemistry 43*, 3385-3395.

8. Kupke, T., Kempter, C., Gnau, V., Jung, G., and Gotz, F. (1994) Mass spectroscopic analysis of a novel enzymatic reaction. Oxidative decarboxylation of the lantibiotic precursor peptide EpiA catalyzed by the flavoprotein EpiD, *J. Biol. Chem. 269*, 5653-5659.

9. Kupke, T., Stevanovic, S., Sahl, H. G., and Gotz, F. (1992) Purification and characterization of EpiD, a flavoprotein involved in the biosynthesis of the lantibiotic epidermin, *J. Bacteriol. 174*, 5354-5361.

10. Allgaier, H., Jung, G., Werner, R. G., Schneider, U., and Zähner, H. (1986) Epidermin: sequencing of a heterodetic 21-peptide amide antibiotic, *Eur. J. Biochem. 160*, 9-22.

11. Yorgey, P., Lee, J., Kordel, J., Vivas, E., Warner, P., Jebaratnam, D., and Kolter, R. (1994) Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor, *Proc. Natl. Acad. Sci. U. S. A. 91*, 4519-4523.

12. Gonzalez-Pastor, J. E., San Millan, J. L., Castilla, M. A., and Moreno, F. (1995) Structure and organization of plasmid genes required to produce the translation inhibitor microcin C7, *J. Bacteriol. 177*, 7131-7140.

13. Guijarro, J. I., Gonzalez-Pastor, J. E., Baleux, F., San Millan, J. L., Castilla, M. A., Rico, M., Moreno, F., and Delepierre, M. (1995) Chemical structure and translation inhibition studies of the antibiotic microcin C7, *J. Biol. Chem. 270*, 23520-23532.

14. Bodansky, M., and Perlman, D. (1964) Are peptide antibiotics small proteins?, *Nature 204*, 840-844.

15. Degnan, B. M., Hawkins, C. J., Lavin, M. F., McCaffrey, E. J., Parry, D. L., and Watters, D. J. (1989) Novel cytotoxic compounds from the ascidian Lissoclinum bistratum, *J. Med. Chem. 32*, 1354-1359.

16. Carroll, A. R., Coll, J. C., Bourne, D. J., MacLeod, J. K., Zabriskie, T., Ireland, C. M., and Bowden, B. F. (1996) Patellins 1-6 and trunkamide A: novel cyclic hexa-, hepta- and octa-peptides from colonial ascidians, Lissoclinum sp. , *Aust. J. Chem. 49*, 659-667.

17. Ireland, C. M., Durso, A. R., Newman, R. A., and Hacker, M. P. (1982) Antineoplastic cyclic peptides from the marine tunicate Lissoclinum patella, *J. Org. Chem. 47*, 1807-1811.

18. Ireland, C., and Scheuer, P. J. (1980) Ulicyclamide and ulithiacyclamide, two new small peptides from a marine tunicate, *J. Am. Chem. Soc. 102*, 5688-5691.

19. Rashid, M. A., Gustafson, K. R., Cardellina, J. H., and Boyd, M. R. (1995) Patellamide F, a new cytotoxic cyclic peptide from the colonial ascidian *Lissoclinum patella*, *J. Nat. Prod. 58*, 594-597.

20. Schmidt, E. W., Sudek, S., and Haygood, M. G. (2004) Genetic evidence supports secondary metabolic diversity in *Prochloron* spp., the cyanobacterial symbiont of a tropical ascidian, *J. Nat. Prod. 67*, 1341-1345.

21. Salomon, C. E., and Faulkner, D. J. (2002) Localization studies of bioactive cyclic peptides in the ascidian *Lissoclinum patella*, *J. Nat. Prod. 65*, 689-692.

22. Schmidt, E. W., Nelson, J. T., Rasko, D. A., Sudek, S., Eisen, J. A., Haygood, M. G., and Ravel, J. (2005) Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella, *Proc. Natl. Acad. Sci. U. S. A. 102*, 7315-7320

23. Donia, M. S., Schmidt, E. W., Lew, M., and Hung-Wen, L. (2010) Cyanobactins - ubiquitous cyanobacterial ribosomal peptide metabolites, in *Comprehensive Natural Products II*, pp 539-558, Elsevier, Oxford.

24. Donia, M. S., Ravel, J., and Schmidt, E. W. (2008) A global assembly line for cyanobactins, *Nat. Chem. Biol. 4*, 341-343.

25. Donia, M. S., Hathaway, B. J., Sudek, S., Haygood, M. G., Rosovitz, M. J., Ravel, J., and Schmidt, E. W. (2006) Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians, *Nat. Chem. Biol. 2*, 729-735.

26. McIntosh, J. A., Donia, M. S., and Schmidt, E. W. (2009) Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds, *Nat. Prod. Rep. 26*, 537-559.

27. Lee, J., McIntosh, J. A., Hathaway, B. J., and Schmidt, E. W. (2009) Using marine natural products to discover a protease that catalyzes peptide macrocyclization of diverse substrates, *J. Am. Chem. Soc. 131*, 2122-2124.

28. McIntosh, J. A., Robertson, C. R., Vinayak, A., Satish, N. K., Bulaj, G. W., and Schmidt, E. W. (2010) Circular Logic: Nonribosomal Peptide-like Macrocyclization with a Ribosomal Peptide Catalyst, *J. Am. Chem. Soc. 132*, 15499-15501.

29. Milne, J. C., Roy, R. S., Eliot, A. C., Kelleher, N. L., Wokhlu, A., Nickels, B., and Walsh, C. T. (1999) Cofactor requirements and reconstitution of microcin B17 synthetase: a multienzyme complex that catalyzes the formation of oxazoles and thiazoles in the antibiotic microcin B17, *Biochemistry 38*, 4768-4781.

30. Roy, R. S., Kim, S., Baleja, J. D., and Walsh, C. T. (1998) Role of the microcin B17 propeptide in substrate recognition: solution structure and mutational analysis of McbA1-26, *Chem. Biol. 5*, 217-228.

31. Milne, J. C., Eliot, A. C., Kelleher, N. L., and Walsh, C. T. (1998) ATP/GTP hydrolysis is required for oxazole and thiazole biosynthesis in the peptide antibiotic microcin B17, *Biochemistry 37*, 13250-13261.

32. Madison, L. L., Vivas, E. I., Li, Y. M., Walsh, C. T., and Kolter, R. (1997) The leader peptide is essential for the posttranslational modification of the DNA-gyrase inhibitor microcin B17, *Mol. Microbiol. 23*, 161-168.

33. Li, Y. M., Milne, J. C., Madison, L. L., Kolter, R., and Walsh, C. T. (1996) From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase, *Science 274*, 1188-1193.

34. Sudek, S., Haygood, M. G., Youssef, D. T., and Schmidt, E. W. (2006) Structure of trichamide, a cyclic peptide from the bloom-forming cyanobacterium Trichodesmium erythraeum, predicted from the genome sequence, *Appl. Environ. Microbiol. 72*, 4382-4387.

35. McIntosh, J. A., Donia, M. S., and Schmidt, E. W. (2010) Insights into Heterocyclization from Two Highly Similar Enzymes, *J. Am. Chem. Soc. 132*, 4089-4091.

36. McIntosh, J. A., and Schmidt, E. W. (2010) Marine Molecular Machines: Heterocyclization in Cyanobactin Biosynthesis, *ChemBioChem 11*, 1413-1421.

37. McIntosh, J. A., Donia, M. S., and Schmidt, E. W. (2011) Enzymatic basis of ribosomal peptide prenylation in cyanobacteria, *J. Am. Chem. Soc. Epub* July 18, 2011.

CHAPTER 2

RIBOSOMAL PEPTIDE NATURAL PRODUCTS: BRIDGING THE RIBOSOMAL AND NONRIBOSOMAL WORLDS

McIntosh, J. A., Donia, M. S., Schmidt, E. W. (2009) Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds, *Nat. Prod. Rep. 26*, 537-559.

Note: my contribution to this paper was in writing and referencing the majority of the document.

CHAPTER 3

USING MARINE NATURAL PRODUCTS TO DISCOVER A PROTEASE THAT CATALYZES PEPTIDE MACROCYCLIZATION

Manuscript reproduced with permission from:

Lee, J. H., McIntosh, J. A., Hathaway, B. J., Schmidt, E. W. (2009) Using marine natural products to discover a protease that catalyzes peptide macrocyclization, *J. Am. Chem. Soc. 131*, 2122-2124.

Note: my contribution to this paper was in helping to plan experiments and in performing the mass spectrometry-based characterization of several substrates presented here.

J|A|C|S COMMUNICATIONS

Published on Web 01/23/2009

Using Marine Natural Products to Discover a Protease that Catalyzes Peptide Macrocyclization of Diverse Substrates

Jaeheon Lee,[†] John McIntosh,[‡] Brian J. Hathaway,[‡] and Eric W. Schmidt*[,†,‡]

Departments of Biology and Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112

Received November 25, 2008; E-mail: ews1@utah.edu

Numerous N−C terminally cyclized ribosomal peptides have been isolated from natural sources (Figure 1), yet enzymatic routes to these peptides remain largely unknown.[1−3] Cyclic peptides have structural features that make them good drug candidates, and in fact synthetic cyclic peptides are used clinically.[3,4] Ribosomal peptides can be artificially cyclized using intein technology, which although groundbreaking suffers from certain sequence and mechanistic requirements.[5] In principle, proteases could also catalyze this process, as exemplified by the proposed pilin cyclization catalyst, which was examined *in vivo* by mutagenesis,[6] and an artificial cyclization of a trypsin inhibitor.[7] The potential synthesis of cyanobactins and cyclotides by proteases is also supported by *in vivo* analysis.[2,8−11] However, no proteins have yet been characterized *in vitro* whose native function is to catalyze N−C cyclization.

Cyanobactins provide excellent models for the study of ribosomal N−C terminal cyclization. These compounds are synthesized as ribosomal precursor peptides that are extensively modified by enzymes to yield cyclic natural products.[2] Cyanobactins are found in diverse cyanobacteria, including symbionts of marine animals. In the first report of this pathway type, a cyanobactin gene cluster, *pat*, was cloned from the metagenome of a coral reef animal and transferred to *E. coli*, leading to production in the laboratory of a cyclic peptide marine natural product previously attributed to an animal.[2] Subsequently, each gene in the *pat* cluster (*patA-G*) was independently expressed under control of a T7 promoter to determine that *patA,D,E,F,G* were necessary and sufficient for cyanobactin biosynthesis and *patB,C* were not required.[10] However, the enzymatic basis of each transformation could not be discerned because removing any essential enzyme led to a loss of detectable products.

Using this initial discovery, we proceeded to assign the function of genes in the pathway by comparing over 30 closely related pathways from the cyanobacterial symbionts that we showed earlier to be the true producers of the cyclic peptides—a novel approach to functional analysis, which we term "deep metagenome mining".[12] This method was applied to show that cyanobactins are made by enzymes with broad substrate selectivity (Figure 1 and Table S1).[10,12] We then showed that wholly artificial cyclic peptides could be generated by this system *in vivo*.[10] The putative cyanobactin peptide cyclase enzymes were demonstrated by these methods to be broadly substrate tolerant.

Most cyanobactin precursor peptides contain two product "cassettes" that are flanked by putative modifying enzyme recognition sequences (Figures 1 and 2).[2,10] Thus, the cassettes are cleaved out of the full-length precursor at both their C- and N-termini and then cyclized to form the cyanobactin products.

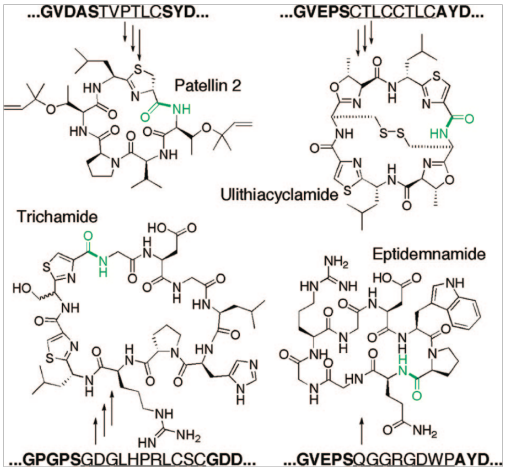


Figure 1. N−C cyclic peptides of the cyanobactin group. The partial precursor peptide sequences are shown that give rise to the cyclic products. Amide bonds formed by PatG and its relatives are shown in green, while recognition sequences are shown in bold and cassette sequences leading to products are underlined. Patellin 2, ulithiacyclamide, and trichamide are natural products that require multiple enzymatic processing steps in their synthesis, while the eptidemnamide precursor was engineered and requires only cleavage and cyclization.

It seemed probable that proteases PatA and PatG could participate in these steps, but several other enzymes are also required to synthesize cyanobactin natural products.[13] Although much experimental evidence indicated that PatA and PatG were involved in this process, their precise roles were not defined. Here, we report on the *in vitro* activity of these enzymes, defining their roles in cyclization of natural and artificial peptides and providing the first *in vitro* study of natural ribosomal peptide N−C cyclization.

Previously, we genetically engineered an artificial precursor peptide, PatEdm, encoding the new compound eptidemnamide.[10] PatEdm encodes the sequence for the natural cyclic peptide patellamide C in cassette 1 and the sequence for the synthetic cyclic peptide eptidemnamide in cassette 2 (Figure 2). Eptidemnamide was produced by the full complement of cyanobactin biosynthetic proteins using heterologous *in vivo* experiments.[10] Because eptidemnamide requires only cyclization, and not the other enzymatic transformations required to produce the natural products, it was used for biochemical experiments described below.

Both PatA and PatG proteases are clearly paralogous and contain homologous domains, including a subtilisin-like serine

[†] Department of Biology.
[‡] Department of Medicinal Chemistry.

10.1021/ja8092168 CCC: $40.75

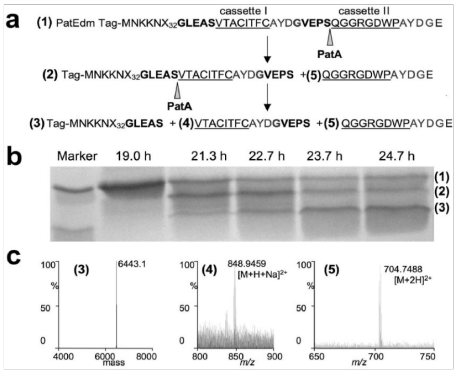


Figure 2. PatA catalyzes two N-terminal cleavage reactions. (a) The precursor peptide PatEdm is cleaved by PatA at a recognition sequence to release first **5** and then **4**, with free N-termini for cyclization. (b) SDS-PAGE of a time-course experiment with PatA and PatEdm shows that the first site is slowly cleaved, followed by relatively rapid proteolysis of the second site. (c) All reaction products were observed by mass spectrometry.

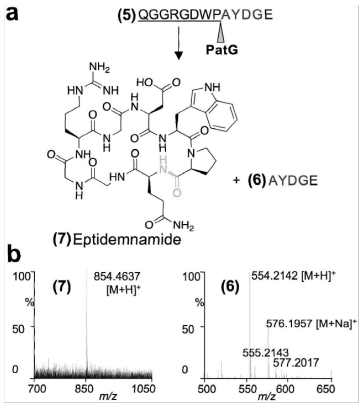


Figure 3. PatG catalyzes cyclization by a transamidation mechanism. (a) Starting from product **5** of PatA, PatG cleaves this peptide into a cyclic compound, eptidemnamide, and a small linear fragment representing the recognition sequence. (b) High resolution mass spectrometry and comparison to authentic standards confirmed that this cyclic peptide was formed from compound **5**.

protease domain and a domain of unknown function. PatG also contains a N-terminal oxidase domain that is not involved in peptide cyclization as shown by metagenomic analysis.[13] The PatA and PatG proteases contain the Asp-Ser-His catalytic triad,[14] as shown by sequence alignment and protein threading (Figure S1).

To determine the potential roles of PatA and PatG in the cyclization of ribosomal peptide, we expressed and purified several constructs of PatA, PatG, and the artificial precursor peptide, PatEdm. Subsequent biochemical analysis defined the roles of each of these proteins. Expressed PatA was isolated as the full-length product with a mixture of fragments, some of which contained only the subtilisin-like protease domain (Figure S2). This domain was neatly cleaved at the linker region between the C-terminal protease and the domain of unknown function, as determined by tryptic-digest mass spectrometry. The protease domain could be purified to homogeneity. PatEdm was purified at high yield despite the inherent difficulties often encountered in expression of precursor peptides.

Both whole PatA, containing a mixture of fragments, and the purified protease domain of PatA were used for biochemical experiments. When incubated overnight with purified PatA or PatA protease domain, PatEdm was slowly cleaved to release first the eptidemnamide-encoding sequence and then the internally located patellamide-C sequence (Figure 2 and Table S3). To further confirm that this protease activity originated in PatA, a Ser-Ala active-site mutant was purified and shown to be inactive in this assay. In time-course experiments, the larger fragment was clearly produced first, followed by accumulation of the shorter fragment (Figure 2). By high-resolution mass spectrometry, we could observe the N-terminal leader sequence, the patellamide-C coding sequence, and the eptidemnamide-coding sequence (Figure 2 and S4). A synthetic standard of the eptidemnamide-coding sequence was used to further confirm this analysis (Figure S3). The enzyme reaction was very slow, but it led to eventual quantitative cleavage of the precursor peptide no matter what enzyme/substrate ratio was employed. This enzyme reaction is not optimal as is discussed further in the paragraphs concerning the PatG protein.

The precursor peptide, PatEdm, was cleaved neatly at the predicted N-terminal recognition sequence, G(L/V)E(A/P)S in both instances (Figure 2 and Figure S4). Thus, the cyanobactin cassettes fused to their conserved C-terminal recognition sequences, AYDG(E), were released upon incubation with PatA. By contrast, when incubated with PatEdm, PatG did not catalyze any detectable proteolysis in numerous different experimental conditions (Table S3). We therefore predicted that PatG would catalyze cleavage and cyclization of the C-terminal recognition sequence and that the products of PatA would be the substrates of PatG.

To assess the PatG reaction, we used synthetic peptide substrate QGGRGDWP-AYDGE. The substrate was identical to the product released by PatA from cassette II, as shown by high resolution LC−MS (Figures 2, 3, S3, and S4).

Several PatG variants were used with the synthetic substrate. Wild-type PatG was expressed, and in addition we constructed a variant in which the N-terminal oxidase domain was absent. Finally, TruG is a related protein that naturally lacks the oxidase domain (Figure S2). These three proteins were separately incubated with the synthetic substrate, either with or without the addition of PatA protease. All of these combinations led to production of cyclic eptidemnamide (Figure 3 and S5, Table S3). The protease recognition sequence, AYDGE, was also clearly discernible in these reaction mixtures as a discrete cleavage product. Cofactors and added metals were not required to catalyze the reaction. In the absence of PatG, with PatA only or without enzyme for example, only starting synthetic peptide was observed.

The presence of eptidemnamide was confirmed by comparing diode array-HPLC and high resolution MS/MS data with an authentic standard (Figures 3, S3, and S5). After 41 h, approximately 1.5 μg of eptidemnamide were synthesized from 50 μg of precursor peptide, representing a ~5% yield. No linear eptidemnamide analogue was detected in these experiments, although unreacted synthetic precursor was still clearly observed. This yield could be optimized by addition of more enzyme. Doubling the enzyme led to double the amount of product (~10% yield), while an increase to a 3:1 substrate/enzyme ratio led to a 75% ± 10% yield.

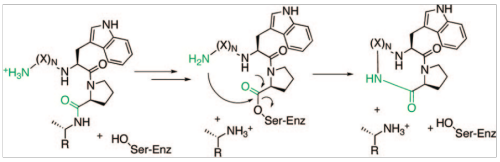


Figure 4. Proposed transamidation mechanism of peptide cyclization. In this simplified scheme, the active-site Ser of PatG is shown forming an enzyme−substrate tetrahedral intermediate with loss of the AYDGE recognition sequence. Subsequently, the N-terminus of the substrate attacks the activated ester bond to release a cyclic peptide and the free enzyme.

This process is clearly far from optimal, since only ~1 turnover is observed per enzyme molecule per day. Despite this slow rate, each monomeric enzyme unit turns over substrate between >2 to 3 times during the course of the experiment under current conditions. Therefore, PatG is shown to be the cyclization catalyst, as this turnover exceeds a stoichiometric reaction. There are several possibilities that could explain low turnover. The substrate is artificial, and the reaction proceeds *in vitro*. Many *in vitro* enzyme reactions require other enzyme partners or buffer optimization to achieve a maximum rate, since they are highly artificial in nature. As one of many possibilities, perhaps the proteases require partnering with an export protein, as is common in lantibiotics. Finally, it should be noted that a single 1 L expression of PatG provides enough protease to synthesize several milligrams of peptide even without optimization.

Previously, it was proposed that cyanobactin cyclization could proceed spontaneously.[15] In this event, PatG would act merely as a protease, and the small linear product would cyclize on its own or via PatG catalysis at a nonproteolytic site. To rule out these possibilities, we obtained the linear precursor, QG-GRGDWP, which lacked the AYDGE sequence. In enzyme reactions with PatG in various combinations and in comparison to a full set of control reactions, this peptide was neither modified nor cyclized (Table S3). Thus, cyclization is not spontaneous, and several lines of evidence favor a one-step transamidation mechanism to cyclization. The data currently support the mechanism shown in Figure 4, although further experiments are required for confirmation.

From these experiments, it is clear that PatA catalyzes the proteolytic cleavage of the N-terminal recognition sequence, while PatG catalyzes cleavage of the C-terminal sequence in tandem with cyclization. No addition of energy (in the form of ATP, for example) is required to drive this process, making PatG the first natural peptide-cyclizing enzyme to be characterized as a pure protein. Aside from the lack of an energy requirement, PatG has another potential technological advantage in that it is highly tolerant of diverse substrate sequences as long as the C-terminal recognition domain is present. So far, we have identified >30 natural peptide sequences that appear to be cleaved and cyclized by PatG (Table S1).[10,12] These data resemble results of alanine-scanning mutagenesis experiments, in that every amino acid position is substituted at least once by a nonconserved amino acid residue. Six, seven, and eight amino acid peptides are cyclized by PatG *in vivo* based upon this same analysis. In addition, close relatives of PatG catalyze the synthesis of >60 additional known natural products (Table S1). Finally, the wholly synthetic peptide, eptidemnamide, is cyclized by PatG *in vivo*[10] and *in vitro*. The exact limitation of product size and sequence content are not yet known. If this enzyme process can be improved, and if broad substrate activity can be demonstrated *in vitro*, the enzyme could serve as a useful general catalyst for the cyclization of peptides.

Acknowledgment. This work was funded by NIH GM071425. We thank T. Bugni and J. Reppart for help with mass spectrometry and G. Bulaj, C. Robertson, and B. Green for synthesis of the PatG substrate peptide.

Supporting Information Available: Experimental details, alignment data, tables, and MS data. The material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Trabi, M.; Craik, D. J. *Trends Biochem. Sci.* **2002**, *27*, 132–138.
(2) Schmidt, E. W.; Nelson, J. T.; Rasko, D. A.; Sudek, S.; Eisen, J. A.; Haygood, M. G.; Ravel, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7315–7320.
(3) Wang, C. K.; Kaas, Q.; Chiche, L.; Craik, D. J. *Nucleic Acids Res.* **2008**, *36*, 210.
(4) Scarborough, R. M. *Am. Heart J.* **1999**, *138*, 1093–1104.
(5) Scott, C. P.; Abel-Santos, E.; Wall, M.; Wahnon, D. C.; Benkovic, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13638–13643.
(6) Eisenbrandt, R.; Kalkum, M.; Lurz, R.; Lanka, E. *J. Bacteriol.* **2000**, *182*, 6751–6761.
(7) Marx, U. C.; Korsinczky, M. L.; Schirra, H. J.; Jones, A.; Condie, B.; Otvos, L., Jr.; Craik, D. J. *J. Biol. Chem.* **2003**, *278*, 21782–21789.
(8) Long, P. F.; Dunlap, W. C.; Battershill, C. N.; Jaspars, M. *ChemBioChem*, **2005**, *6*, 1760–1765.
(9) Saska, I.; Gillon, A. D.; Hatsugai, N.; Didetzgen, R. G.; Hara-Nishimura, I.; Anderson, M. A.; Craik, D. J. *J. Biol. Chem.* **2007**, *282*, 29721–29728.
(10) Donia, M. S.; Hathaway, B. J.; Sudek, S.; Haygood, M. G.; Rosovitz, M. J.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2006**, *2*, 729–735.
(11) Gillon, A. D.; Saska, I.; Jennings, C. V.; Guarino, R. F.; Craik, D. J.; Anderson, M. A. *Plant J.* **2008**, *53*, 505–515.
(12) Donia, M. S.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2008**, *4*, 341–343.
(13) Sudek, S.; Haygood, M. G.; Youssef, D. T.; Schmidt, E. W. *Appl. Environ. Microbiol.* **2006**, *72*, 4382–4387.
(14) Siezen, R. J.; Renckens, B.; Boekhorst, J. *Proteins* **2007**, *67*, 681–694.
(15) Milne, B. F.; Long, P. F.; Starcevic, A.; Hranueli, D.; Jaspars, M. *Org. Biomol. Chem.* **2006**, *4*, 631–638.

JA8092168

Supporting Online Material for

Using Marine Natural Products to Discover a Protease that Catalyzes Peptide Macrocyclization of Diverse Substrates

Jaeheon Lee,[1] John McIntosh,[2] Brian J. Hathaway,[2] and Eric W. Schmidt[1,2*]

1. Department of Biology and 2. Department of Medicinal Chemistry, University of Utah, Salt Lake City UT 84112 USA

ews1@utah.edu

This file contains:

Materials and Methods

General Methods

Chemicals were purchased from Sigma-Aldrich. Chemically competent DH5α and the TOPO-TA vector were obtained from Invitrogen. Chemically competent BL21(DE3) *E. coli* cells were purchased from Stratagene. pCDFDuet-1, pACYCDuet-1, pRSFDuet-1, pET-28b vectors were obtained from Novagen. All restriction endonucleases were purchased from New England Biolabs and the DNA ligation kit was purchased from Takara. PCR amplications were performed using Platinum *Taq* DNA Polymerase High Fidelity from Invitrogen. *E. coli* strain DH5α was used for routine subcloning and transformed by electroporation according to Sambrook *et al* (*1*). BL21 Star™ (DE3) (Invitrogen) was used for protein expression. Primers were synthesized by the University of Utah DNA / Peptide Core Facility and are listed in Table S2. An authentic standard of synthetic eptidemnamide was purchased from New England Peptide LLC.

Sequence Alignment and Threading

FASTA sequences for PatA domain 1 and PatG domain 2 were compared to available sequences by BLAST analysis (www.ncbi.nlm.nih.gov/blast/) (*2*). Hits with known crystal structures were chosen to serve as threading models (Pat A: 1DBH6a; Pat G D1THM). Local alignment with ClustalX to the respective sequences confirmed the accuracy of these comparisons, allowing identification of not only active site residues, but also conserved motifs for the length of the protein (*3*). Once the appropriateness of the model was determined, sequences were threaded to

the model using the Phyre program,(*4*) which generated PDB files with the residues of the PatA and PatG proteases fitted to the relevant structure. Proposed active site residues in the model were determined by comparison with biochemically validated active site residues in the parent structure. Finally, the subtilisin BPN' sequence was used as a structural guide. Because it is very distant from the Pat proteases, it was aligned to the ClustalX output by hand using conserved sequence motifs identified by threading (Fig. S1).

Preparation of Expression Vectors

Protein expression vectors used in this study were based upon Duet vectors N-terminally fused with either maltose binding protein (MBP) from the pMAL-c4 vector (New England Biolabs) or a 6x-Histidine tag from pET28b (Novagen). The *patA* gene was amplified from the previously reported patApCDF vector (*5*) using primers PatAf and PatAr and cloned into pACYCDuet-1. Both MBP- and 6x-His-tagged PatA constructs were cloned. The *patG* gene in the previously reported patGpCDF vector (*5*) was fused with a 6x-His tag. PatG contains an oxidase domain, which was removed by PCR using primers Gprof and Gpror (Table S2). The resulting fragment was cloned into the pCDFDuet-1 vector containing a 6x-His tag. The *truG* gene was cloned into pCDFDuet-1 vector and tagged with 6x-His. Finally, the previously reported patEdmRSF vector (*5*) was also N-terminally tagged with 6x-His.

To introduce an active-site serine to alanine mutation into *patA*, a single round of PCR mutagenesis was performed using the QuickChange Multi Site-directed Mutagenesis kit,

following the manufacturer's protocol, with primers PatAmf and PatAmr. The cloned PCR product was sequenced and ligated into the pACYCDuet-1 vector with an N-terminal MBP tag.

Peptide Synthesis

Cyclic eptidemnamide was previously synthesized and characterized by 2D NMR and mass spectrometry. In this study, it was necessary to obtain an MS/MS spectrum of eptidemnamide (Fig. S3). In addition, a synthetic substrate for PatG, QGGRGDWPAYDGE, was synthesized by standard solid phase methodology using Fmoc (*N*-(9-fluorenyl)methoxycarbonyl) chemistry (*6*), deprotected, and purified to homogeneity by HPLC. High-resolution mass spectrometry confirmed that the expected sequence was obtained from synthesis (Fig. S3). This sequence encoded the cyclic peptide eptidemnamide fused to the PatG recognition sequence AYDGE. Finally, a control peptide, QGGRGDWP, which lacked the PatG recognition sequence, was purchased from Elm Biophamaceuticals Inc. The sequence was confirmed by mass spectrometry (Fig. S3).

Protein Expression and Purification

For all protein expression experiments, seed cultures were grown overnight and a 1:1000 dilution was used to inoculate Luria-Bertani broth (LB) containing appropriate antibiotics. For MBP fusion proteins, LB broth containing glucose (2g per liter) was used. Cultures were grown at 37°C until mid log phase (OD_{600} ~0.5), when protein expression was induced with IPTG (isopropyl β-D-1-thiogalactopyranoside).

To express PatA constructs, cultures were induced with 0.5 mM IPTG and incubated for an additional 24-28 h at 15°C in a shaker. After harvesting the cells, each sample was resuspended in 40 ml of lysis buffer (20 mM Tris pH 7.5, 250 mM NaCl, and 2 mM BME) with pepstatin (2.5 μM) and lysed by sonication followed by centrifugation at 14,000×g for 45 minutes. For MBP fusion proteins, the cleared lysates were applied to an amylose column, washed with 12 column volumes of lysis buffer, and eluted by gravity flow with lysis buffer containing 10 mM maltose. For His-tagged fusion proteins, the cleared cell lysates were applied to an Ni-NTA column, and protein fractions were eluted by gravity flow under increasing concentrations of imidazole. The eluates were analyzed by SDS-PAGE and the protein fractions were dialyzed overnight against buffer containing 20 mM Tris pH 7.5, 50 mM NaCl, 2 mM BME, 10% (v/v) glycerol. The dialyzed proteins were stored at -80°C. PatA-His was further purified by size exclusion chromatography. The dialyzed fraction was concentrated using a Vivaspin 20 concentrator and loaded into a SD-200 column connected to a AKTA prime FPLC system. The eluted fractions were analyzed by SDS-PAGE (Fig. S2) and the appropriate fractions were stored at -80°C.

PatG, PatG protease, and TruG were obtained as 6x-His fusion proteins as described above. A mutant of PatA containing an active-site Ser-Ala mutation was purified as an MBP fusion in the manner described above.

PatEdm is a substrate for the above enzymes and was optimally purified from inclusion bodies. 6x-His tagged PatEdm was purified as expressed and purified as follows. An overnight culture was inoculated into 1 liter of LB broth and grown at 37°C until OD_{600} ~ 0.5. The culture was induced with 0.1 mM ITPG and incubated at 30°C for 18 h in a shaker. The cells were harvested

by centrifugation, suspended in 80 ml of Buffer B (100 mM NaH_2PO_4, 10 mM Tris, and 8 M urea adjusted to pH 8.0), and purified under denaturing conditions using the QIAexpressionist protocol from Qiagen. The purified protein was dialyzed three times over two days against 20 mM Tris pH 7.5, 50 mM NaCl, 2 mM BME. The dialyzed protein was stored at -80°C.

LC-MS Analysis

LC-MS analyses were performed using a Waters Alliance 2795 HPLC with a analytical C18 column (Phenomenex, Luna) connected to a Micromass Q-TOF electropsray mass spectrometer using a concurrently running LockSpray with Leucine-Enkephalin (m/z = 556.2771 for $[M+H]^+$) as the mass standard. MS/MS analysis was performed to compare the fragmentation pattern of enzyme-synthesized eptidemnamide with the one of authentic standard of synthetic eptidemnamide with collision-induced dissociation energy at 55 eV. Direct infusion experiments were run without an internal standard.

Cleavage of PatEdm

Purified PatEdm was incubated with purified PatA and PatG constructs under various conditions. PatEdm (40 µg, 42 µM) was added to a reaction mixture (50 mM Tris pH 7.5, 10 mM $CaCl_2$, 3 mM DTT, 10 µM flavin mononucleotide (FMN)) in a water bath at 34°C. Reactions were run without enzyme or with 14 µM enzyme (Table S3). The following enzymes were used in various assays: 6x-His-PatA, MBP-PatA, 6x-His-PatA protease domain, 6x-His-PatG, 6x-His-PatG protease, 6x-His-TruG, and combinations thereof (Table S3). Time points were taken at 1 h, 3 h, 6 h, 10 h, and following overnight incubation. For time-course experiments with PatA, in

addition to the time points described above, time points were taken at 19 h, 21 h 15 min, 22 h 43 min 23 h 40 min, and 24h 40 min. Time points were analyzed by SDS-PAGE, and reaction products were verified by mass spectrometry (Fig. S4). Only the PatA protease domain was required to catalyze cleavage of PatEdm. No cleavage products were observed in numerous experiments unless the PatA protease domain was present in the reaction mixture.

To further verify the protease activity of PatA, MBP-PatA and MBP-PatA Ser-Ala mutant were incubated separately with PatEdm as described above. The wild-type enzyme catalyzed cleavage of PatEdm, while the Ser-Ala mutant did not (Table S3).

Cyclization of Synthetic Precursor Peptide

Synthetic precursor peptides, either QGGRGDWPAYDGE or QGGRGDWP at 710 μM, were dissolved in a 50 μl reaction mixture (50 mM Tris pH 7.5, 10 mM $CaCl_2$, 3 mM DTT, and 10 μM FMN) in a 34°C water bath. Enzymes were added to final concentrations of 14 μM, and the reaction was allowed to proceed for 41 h. The following enzymes were used: 6x-His-PatG, 6x-His-PatG protease, and 6x-His-TruG (Table S3). These enzymes were also co-incubated with 6x-His-PatA. As control reactions, 6x-His-PatA alone or no enzyme was used. Reactions were quenched by freezing at -80°C, and reaction products were resolved by HPLC (Hitachi L6200 intelligent pump and L3000 photo diode array detector) using a Phenomenex CN column. At a flow rate of 1 ml min^{-1} initial conditions of 99 % water, 1 % acetonitrile were maintained for 5 minutes, followed by a linear gradient up to 70 % water, 30 % acetonitrile over 30 minutes. Standards of eptidemnamide and synthetic linear peptides were used to determine elution times

of reaction products and to estimate the concentration of eptidemnamide obtained from enzyme reactions. To verify that the eluted fractions corresponded to the expected compounds, every fraction was collected during the HPLC run and further analyzed by direct infusion and LC-mass spectrometry. PatG protease or TruG were sufficient to catalyze cyclization, while other proteins or cofactors did not contribute to activity (Table S3).

Determination of Cofactor Requirements

FMN and Ca^{2+} were added to enzyme reactions because PatG contains a putative FMN-binding oxidase domain and potential Ca^{2+}-binding residues were identified by threading. However, when $CaCl_2$ or FMN were left out of reaction mixtures, the reaction products appeared to be identical to those obtained with those cofactors. ATP (1 mM) was also added to select reactions as described above, though it did not have a noticeable effect on eptidemnamide synthesis or on the cleavage of PatEdm (Table S3).

Yield Improvement by Varying PatG Concentration

Doubling the amount of PatG in enzyme reactions as described above led to doubling of the yield in the same time period. To further explore this phenomenon, synthetic precursor peptide, QGGRGDWPAYDGE was mixed to a final concentration of 284 μM in a 1 ml reaction volume (50 mM Tris pH 7.5, 10 mM $CaCl_2$, 3 mM DTT, and 10 μM FMN) in a 35 °C water bath. Enzymes were added to final concentrations of 95 μM of PatG protease and 8 μM of PatA, and the reaction was allowed to proceed for 67 h. The reaction was quenched by freezing at –80 °C,

and 400 µl of reaction products were resolved by HPLC using a Phenomenex CN column as described above. A standard of eptidemnamide was used to determine elution times and estimate the concentration of eptidemnamide obtained from the enzyme reaction. We obtained eptidemnamide (90 µg, 75 ± 10% *isolated* yield) from 200 µg of precursor peptide in this 1 ml reaction volume.

TABLES (Supporting Material)

Supporting Table 1. Deep metagenome analysis reveals that cyanobactin cyclization enzymes accept broadly different substrates. Cyanobactin sequences encoded within larger, PatEdm-like precursor peptides are shown. Posttranslational modifications include heterocycles (yellow) and prenyl groups (green). Underlined residues indicate oxidized heterocycles. Italics indicate that oxidation is variable.

All sequences shown here are N-C cyclized. Table headers indicate the level of experimental evidence. “Processed” indicates that both chemicals and biosynthetic genes are known. Since these genes are clustered and biochemical and *in vivo* heterologous expression data exists for many representatives, the level of evidence is high. “Probably processed” indicates that, while the structures are known, genes have not been cloned. “Identified by deep genome mining” indicates that, while the genes are known, the resulting structures have not been chemically isolated. All of the genes in the mining category are related to the PatA / PatG family.

Importantly, every position is varied at least once in sequences that have been experimentally verified to be modified by PatG / TruG. The other categories provide a sense of the known diversity of this group, all of which is modified by similar or identical enzymes.

Processed by PatG / TruG	Probably processed by PatG / TruG	Identified by deep genome mining (PatG)	Processed by relatives of PatG	Probably processed by relatives of PatG
QGGRGDWP	ITVCITVC	VSSCITFC	HCATIC	FTACAC
VTACITFC	VTACFC	LTTCITFC	ATVSIC	ATVCAC
ITVCISVC	VSACVC	LTACVTFC	FTGCMC	ATACMC
LTACITFC	VSVTVC	ITVCITVC	ITGCIC	GCVPLQC
ACFPTIC	VCVTVC	LAACITFC	ATGCMC	GCIPLQC
FCFPTVC	ITFTAC	LTACITLC	ATGCAC	ILYCNPSLC
TVPTLC	FTVPCVC	ITVCISAC	GDGLHPRLCSC	ATGCVC
TLPVPTLC	ITVPCLC	STVCFTVC		YGTGEFFNP
TFPVPTVC	ATFCAC	VTACIAFC	From genome	YLYPINP
TSIAPFC	LSGPIC	VTACITSC	sequence:	ISASFC
CTLCCTLC	ICFPTVC	VTACITLC	VCMPCYP	ATACVC
CTLCCTFC	ICFPTIC	VTTCITFC	ACMPCYP	TTACVC
	VCFPTIC	VTACTTFC		VTVTVT
	ICPVCMC	CTLCCTLR		FPISAPPGVTFS
	VCPVCMC	CILCCTLC		FPISFPC
	IPISFPC	CTLCCALC		PASYPTIP
	VTPFVC	CTLCCTVC		PASYPTIP
	LSGPIC	CTVCCAVC		PASYPTIP
	VTIPIIP	CTLCYTLC		FPMPPMC
	ITIPIIP	LCFPTVC		VPLSATC
	ITACITFC	FCVPTVC		FPMPPMC
	VTVCITFC	FCFPAVC		ASACIC
	VTVCVTFC	FCLPTVC		
	ITACLTFC			
	TLPVPTVC			
	TVPVPSFC			
	VPVCFCIC			
	VPVCLCIC			

Supporting Table 2. Primers used in this study.

	Primer Name	Primer Sequence (5' – 3')
1	PatAf	CCCATATGAATAGAGATATTTTGCGAAC
2	PatAr	GGCTCGAGTTAGTAAGAAGAAGACCAAGAA
3	PatAmf	CGCCTCAGCGGCACCGCCTTTGCCACTCCG
4	PatAmr	CGGAGTGGCAAAGGCGGTGCCGCTGAGGCG
5	Gprof	CCCATATGGGTGAGGATGAAATCGAATCTG
6	Gpror	TTCTTTACCAGACTCGAGGGTA
7	TruGf	TTCATATGAGTCGTCATCCTTTTAATATTTGCC
8	TruGr	CGGTACCCCAATAACTACTTTGAGACGGTG

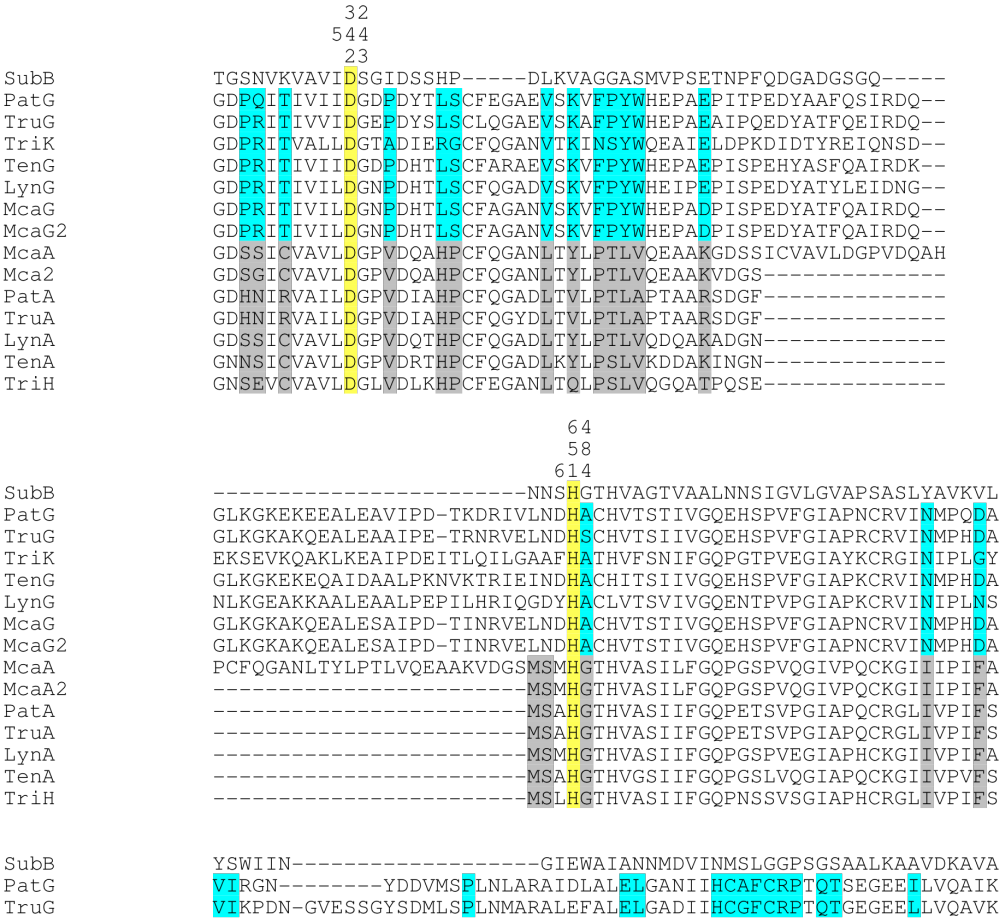
Supporting Table 3. Summary of Studies on Proteases. Most samples and controls were run at least in triplicate or more. All important controls were run at least in triplicate.

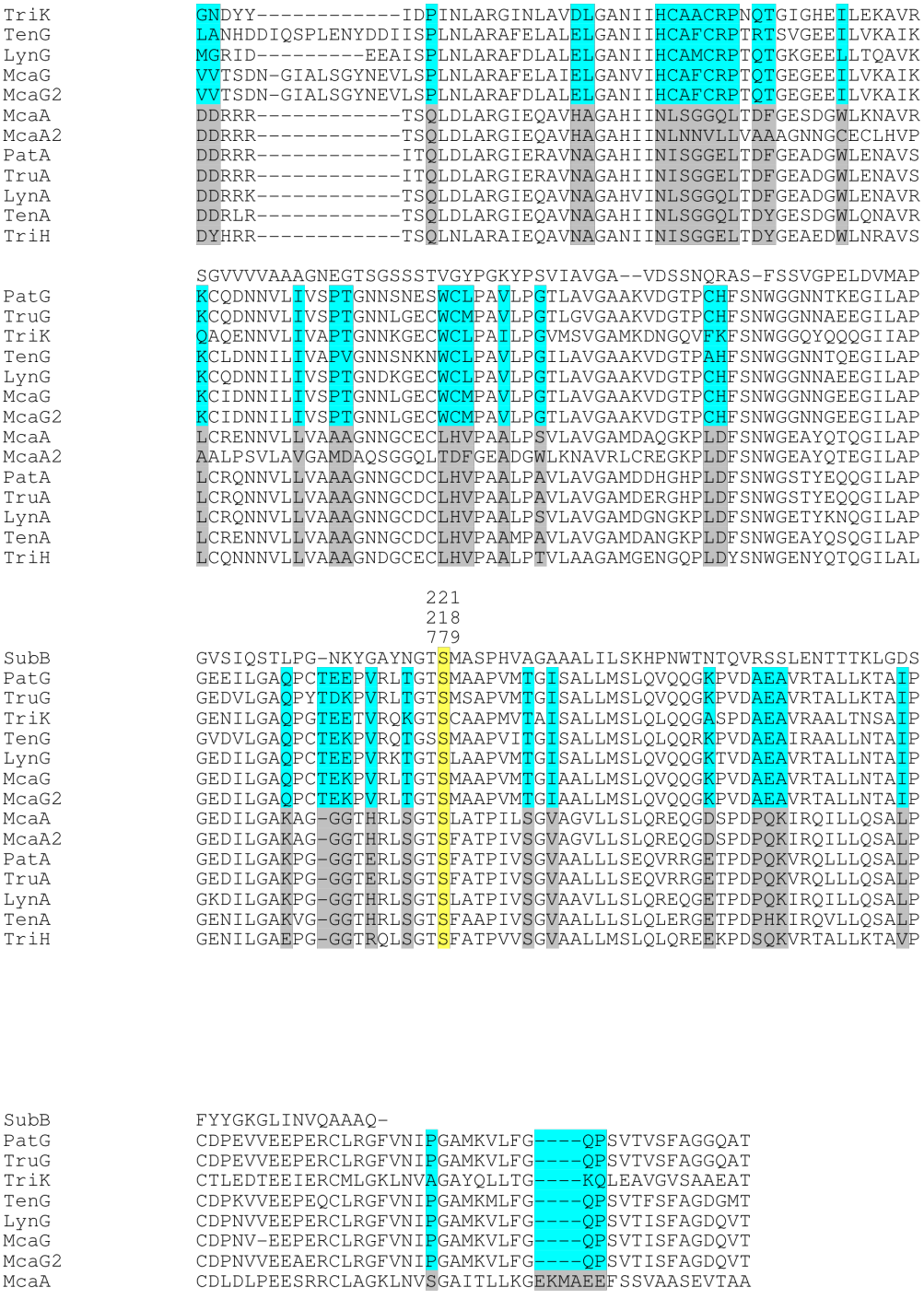
Protein	Substrate	Cofactor	Product
PatA	PatEdm	FMN, ATP	Three cleaved products; (3), (4), (5) (Fig. 2)
PatA	PatEdm	FMN	Three cleaved products; (3), (4), (5) (Fig. 2)
PatA	PatEdm	none	Three cleaved products; (3), (4), (5) (Fig. 2)
PatG	PatEdm	FMN	none
ΔPatA (a Ser-Ala mutant)	PatEdm	FMN	none
PatG protease	PatEdm	FMN	none
TruG	PatEdm	FMN	none
PatA+PatG	PatEdm	FMN	Three cleaved products; (3), (4), (5) (Fig. 2)
PatA+PatGprotease	PatEdm	FMN	Three cleaved products; (3), (4), (5) (Fig. 2)
PatA+TruG	PatEdm	FMN	Three cleaved products; (3), (4), (5) (Fig. 2)
PatG	Synthetic peptide (QGGRGDWPAYDGE)	FMN	Eptidemnamide
PatG protease	Synthetic peptide (QGGRGDWPAYDGE)	FMN	Eptidemnamide
PatG protease	Synthetic peptide (QGGRGDWP)	FMN	none
TruG	Synthetic peptide (QGGRGDWPAYDGE)	FMN	Eptidemnamide
PatA+PatG	Synthetic peptide (QGGRGDWPAYDGE)	FMN	Eptidemnamide
PatA+PatGprotease	Synthetic peptide (QGGRGDWPAYDGE)	FMN	Eptidemnamide
PatA+PatGprotease	Synthetic peptide (QGGRGDWPAYDGE)	FMN	Eptidemnamide
PatA+PatGprotease	Synthetic peptide (QGGRGDWPAYDGE)	none	Eptidemnamide
PatA+PatGprotease	Synthetic peptide (QGGRGDWP)	FMN	none
PatA+TruG	Synthetic peptide (QGGRGDWPAYDGE)	FMN, ATP	Eptidemnamide
PatA+TruG	Synthetic peptide (QGGRGDWPAYDGE)	FMN	Eptidemnamide
No enzyme	Synthetic peptide (QGGRGDWPAYDGE)	FMN	none
No enzyme	Synthetic peptide (QGGRGDWP)	FMN	none
PatG protease	Synthetic peptides (QGGRGDWPAYDGE)+(QGGRGDWP)	FMN	Eptidemnamide
PatA+PatGprotease	Synthetic peptides (QGGRGDWPAYDGE)+(QGGRGDWP)	FMN	Eptidemnamide

FIGURES (Supporting Material)

Figure S1. Sequence comparison of protease domains from known PatA and PatG homologs. A) An alignment of subtilisin BPN' at top (SubB), followed by PatG family proteins (mid) and PatA family proteins (bottom). Catalytic triad residues are in yellow, with numbering corresponding to SubB (top), PatG (mid), and PatA (bottom). In blue, residues that are conserved in the PatG group are shown, while those conserved in the PatA group are in gray. Additionally, there is a large loop containing numerous Lys and Glu residues at approximately residue 575 (PatG numbering) that may be important to PatG activity. B) Threading model of PatA showing location of important residues. At left, structural elements are shown as colors while the catalytic triad is white. At right, the catalytic triad is yellow and residues specific to PatA or PatG are shown in red. Threading model for PatG was nearly identical within the limitations of these experiments. Note that the residues specific to PatA or PatG are widely interspersed throughout the protein. We are currently working to obtain structural data to better define the importance of these residues.

A




TriK GNDYY----------IDPINLARGINLAVDLGANIIHCAACRPNQTGIGHEILEKAVR
TenG LANHDDIQSPLENYDDIISPLNLARAFELALELGANIIHCAFCRPTRTSVGEEILVKAIK
LynG MGRID---------EEAISPLNLARAFDLALELGANIIHCAMCRPTQTGKGEELLTQAVK
McaG VVTSDN-GIALSGYNEVLSPLNLARAFELAIELGANVIHCAFCRPTQTGEGEEILVKAIK
McaG2 VVTSDN-GIALSGYNEVLSPLNLARAFDLALELGANIIHCAFCRPTQTGEGEEILVKAIK
McaA DDRRR----------TSQLDLARGIEQAVHAGAHIINLSGGQLTDFGESDGWLKNAVR
McaA2 DDRRR----------TSQLDLARGIEQAVHAGAHIINLNNVLLVAAAGNNGCECLHVP
PatA DDRRR----------ITQLDLARGIERAVNAGAHIINISGGELTDFGEADGWLENAVS
TruA DDRRR----------ITQLDLARGIERAVNAGAHIINISGGELTDFGEADGWLENAVS
LynA DDRRK----------TSQLDLARGIEQAVNAGAHVINLSGGQLTDFGEADGWLENAVR
TenA DDRLR----------TSQLDLARGIEQAVNAGAHIINLSGGQLTDYGESDGWLQNAVR
TriH DYHRR----------TSQLNLARAIEQAVNAGANIINISGGELTDYGEAEDWLNRAVS
SGVVVVAAAGNEGTSGSSSTVGYPGKYPSVIAVGA--VDSSNQRAS-FSSVGPELDVMAP
PatG KCQDNNVLIVSPTGNNSNESWCLPAVLPGTLAVGAAKVDGTPCHFSNWGGNNTKEGILAP
TruG KCQDNNVLIVSPTGNNLGECWCMPAVLPGTLGVGAAKVDGTPCHFSNWGGNNAEEGILAP
TriK QAQENNVLIVAPTGNNKGECWCLPAILPGVMSVGAMKDNGQVFKFSNWGGQYQQQGIIAP
TenG KCLDNNILIVAPVGNNSNKNWCLPAVLPGILAVGAAKVDGTPAHFSNWGGNNTQEGILAP
LynG KCQDNNILIVSPTGNDKGECWCLPAVLPGTLAVGAAKVDGTPCHFSNWGGNNAEEGILAP
McaG KCIDNNILIVSPTGNNLGECWCMPAVLPGTLAVGAAKVDGTPCHFSNWGGNNGEEGILAP
McaG2 KCIDNNILIVSPTGNNLGECWCMPAVLPGTLAVGAAKVDGTPCHFSNWGGNNGEEGILAP
McaA LCRENNVLLVAAAGNNGCECLHVPAALPSVLAVGAMDAQGKPLDFSNWGEAYQTQGILAP
McaA2 AALPSVLAVGAMDAQSGGQLTDFGEADGWLKNAVRLCREGKPLDFSNWGEAYQTEGILAP
PatA LCRQNNVLLVAAAGNNGCDCLHVPAALPAVLAVGAMDDHGHPLDFSNWGSTYEQQGILAP
TruA LCRQNNVLLVAAAGNNGCDCLHVPAALPAVLAVGAMDERGHPLDFSNWGSTYEQQGILAP
LynA LCRQNNVLLVAAAGNNGCDCLHVPAALPSVLAVGAMDGNGKPLDFSNWGETYKNQGILAP
TenA LCRENNVLLVAAAGNNGCDCLHVPAAMPAVLAVGAMDANGKPLDFSNWGEAYQSQGILAP
TriH LCQNNNVLLVAAAGNDGCECLHVPAALPTVLAAGAMGENGQPLDYSNWGENYQTQGILAL
221
218
779
SubB GVSIQSTLPG-NKYGAYNGTSMASPHVAGAAALILSKHPNWTNTQVRSSLENTTTKLGDS
PatG GEEILGAQPCTEEPVRLTGTSMAAPVMTGISALLMSLQVQQGKPVDAEAVRTALLKTAIP
TruG GEDVLGAQPYTDKPVRLTGTSMSAPVMTGISALLMSLQVQQGKPVDAEAVRTALLKTAIP
TriK GENILGAQPGTEETVRQKGTSCAAPMVTAISALLMSLQLQQGASPDAEAVRAALTNSAIP
TenG GVDVLGAQPCTEKPVRQIGSSMAAPVITGISALLMSLQLQQRKPVDAEAIRAALLNTAIP
LynG GEDILGAQPCTEEPVRKTGTSLAAPVMTGISALLMSLQVQQGKTVDAEAVRTALLNTAIP
McaG GEDILGAQPCTEKPVRLTGTSMAAPVMTGIAALLMSLQVQQGKPVDAEAVRTALLNTAIP
McaG2 GEDILGAQPCTEKPVRLTGTSMAAPVMTGIAALLMSLQVQQGKPVDAEAVRTALLNTAIP
McaA GEDILGAKAG-GGTHRLSGTSLATPILSGVAGVLLSLQREQGDSPDPQKIRQILLQSALP
McaA2 GEDILGAKAG-GGTHRLSGTSFATPIVSGVAGVLLSLQREQGDSPDPQKIRQILLQSALP
PatA GEDILGAKPG-GGTERLSGTSFATPIVSGVAALLLSEQVRRGETPDPQKVRQLLLQSALP
TruA GEDILGAKPG-GGTERLSGTSFATPIVSGVAALLLSEQVRRGETPDPQKVRQLLLQSALP
LynA GKDILGAKPG-GGTHRLSGTSLATPIVSGVAAVLLSLQREQGETPDPQKIRQILLQSALP
TenA GENILGAKVG-GGTHRLSGTSFAAPIVSGVAALLLSLQLERGETPDPHKIRQVLLQSALP
TriH GENILGAEPG-GGTRQLSGTSFATPVVSGVAALLMSLQLQREEKPDSQKVRTALLKTAVP
SubB FYYGKGLINVQAAAQ-
PatG CDPEVVEEPERCLRGFVNIPGAMKVLFG----QPSVTVSFAGGQAT
TruG CDPEVVEEPERCLRGFVNIPGAMKVLFG----QPSVTVSFAGGQAT
TriK CTLEDTEEIERCMLGKLNVAGAYQLLTG----KQLEAVGVSAAEAT
TenG CDPKVVEEPEQCLRGFVNIPGAMKMLFG----QPSVTFSFAGDGMT
LynG CDPNVVEEPERCLRGFVNIPGAMKVLFG----QPSVTISFAGDQVT
McaG CDPNV-EEPERCLRGFVNIPGAMKVLFG----QPSVTISFAGDQVT
McaG2 CDPNVVEEAERCLRGFVNIPGAMKVLFG----QPSVTISFAGDQVT
McaA CDLDLPEESRRCLAGKLNVSGAITLLKGEKMAEEFSSVAASEVTAA

```
McaA2           CELDLPEESRRCLAGKLNVSGAITLLKGEKMSEEFSSVAASEVTAA
PatA            CDDDAPEQARRCLAGRLNVSGAFTLLKGGNMSEELATASFPSVEAS
TruA            CDDDAPEQARRCLAGRLNVSGAFTLLKGGNMSEELATASFPSVEAS
LynA            CDPDLPEETRRCLAGKLNVSGAITLLKGGKMAEEFPSVEVSEVQAA
TenA            CDADLPEDAKRCLAGKLNVSGAITLLKGGKMAEEFTSVSASEVEAA
TriH            CHAQ---EKRRCLVGQMNISGAIAHITGETMSESEQDNSN-GIEAS
```

B

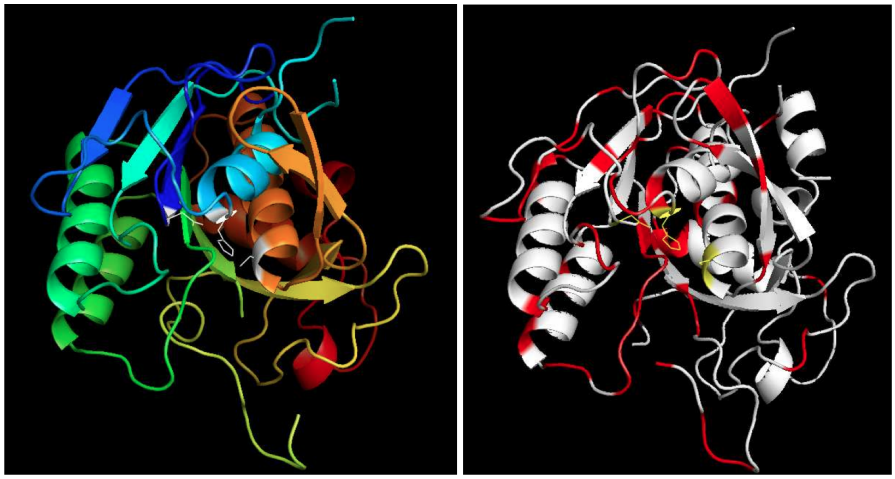
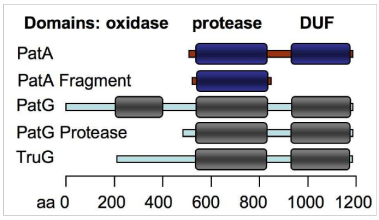
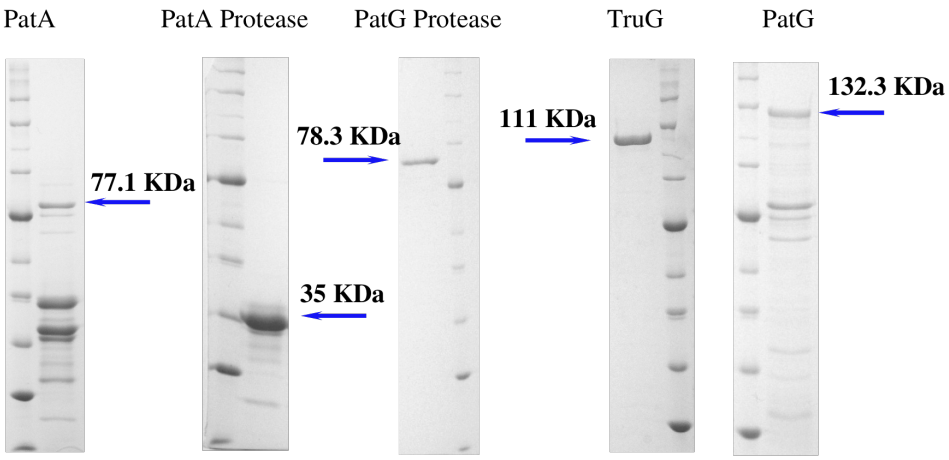


Figure S2. Recombinant protein expression and purification. A) Schematic of purified protease enzymes. B) Purified enzymes used in this study. Blue arrows point at purified proteins named at top of gel. Marker lanes (Broad Range Protein Marker, New England Biolabs) are included. C) Purified PatEdm (left), its direct infusion ESI-MS spectrum (right), and deconvoluted mass data (bottom). Mass corresponds to the expected peptide missing its N-terminal Met, which was calculated to be 9486.5. The deconvolution was obtained using the program MagTran1.02.

A

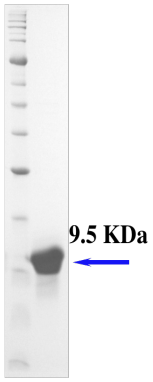


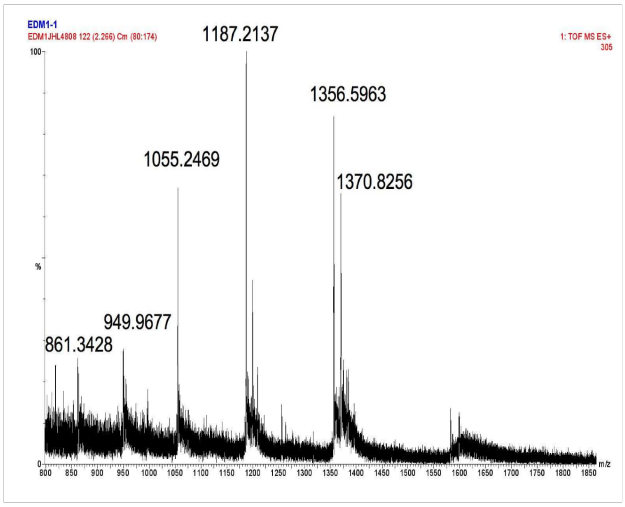
B



C

PatEdm


9.5 KDa


EDM1-1
EDM1JHL4808 122 (2.266) Cm (80:174)
1: TOF MS ES+
305
1187.2137
1356.5963
1055.2469
1370.8256
949.9677
861.3428
%
m/z

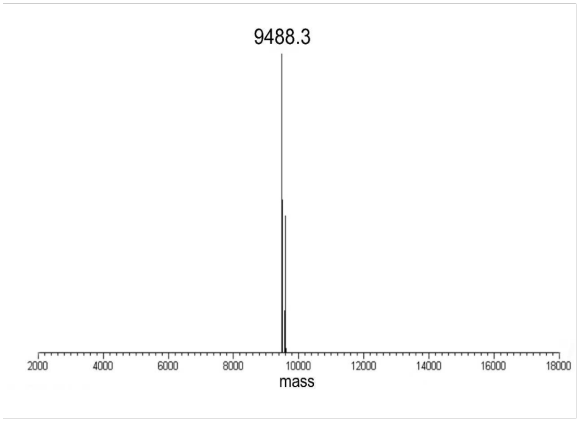
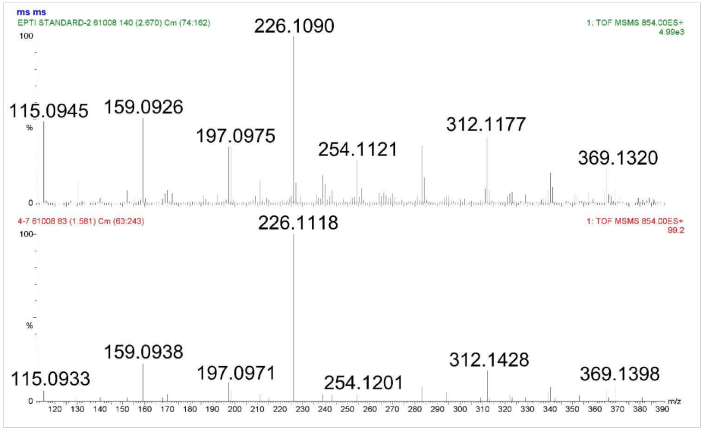
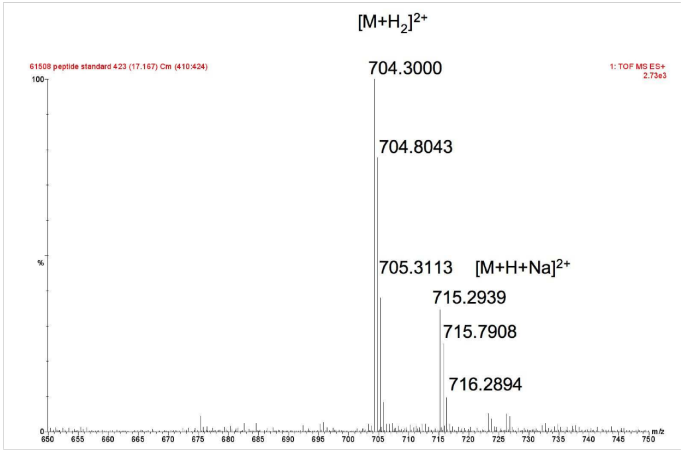

9488.3
mass

Figure S3. Characterization of synthetic standards used in this study. Eptidemnamide was fully characterized previously. A) new MS/MS data for the molecule are shown and compared with an enzyme reaction product. B) High-resolution MS data for a synthetic substrate for PatG, QGGRGDWPAYDGE. The calculated mass for $[M+2H]^{2+}$ is 704.3000. C) Low-resolution MS data for a synthetic peptide, QGGRGDWP. The calculated mass for $[M+H]^{+}$ is 872.9162.

A



B



C

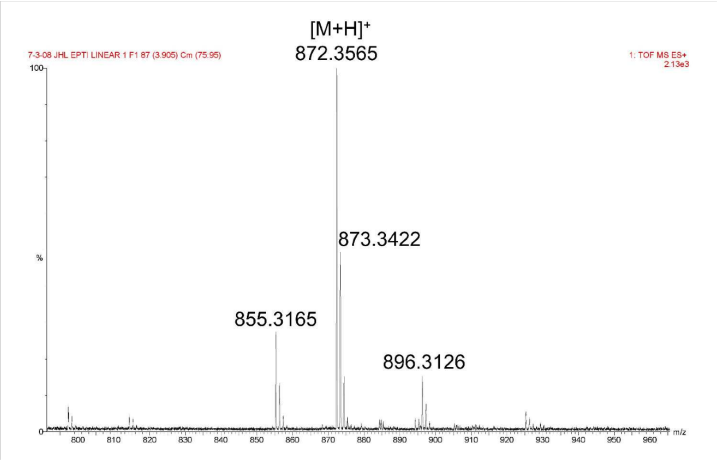
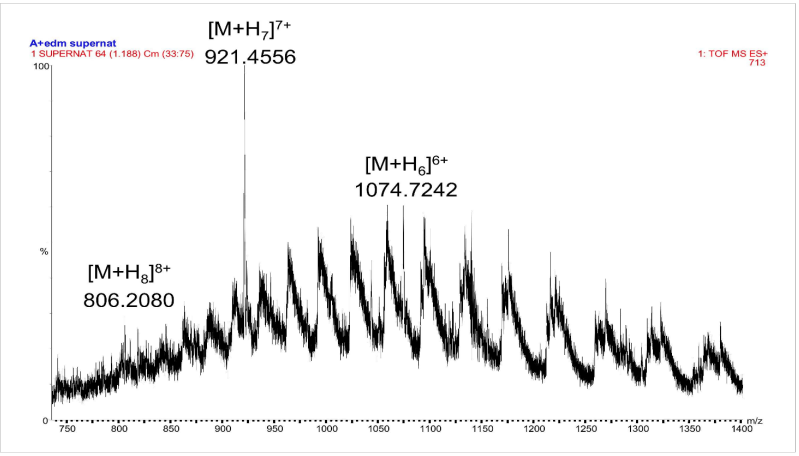
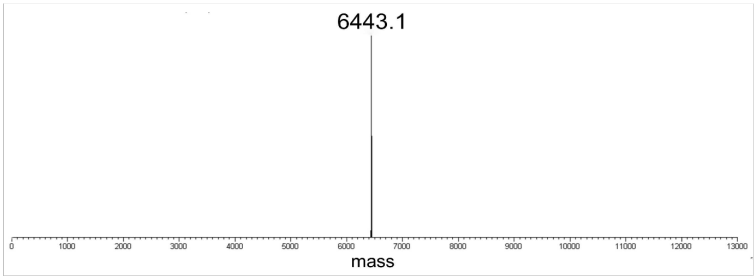

[M+H]+
872.3565
7-3-08 JHL EPTI LINEAR 1 F1 87 (3.905) Cm (75:95)
1: TOF MS ES+
2.13e3
100
%
0
873.3422
855.3165
896.3126
800
810
820
830
840
850
860
870
880
890
900
910
920
930
940
950
960
m/z

Figure S4. Mass spectrometric analysis of PatA products. Following reaction with PatA, PatEdm reaction products were analyzed by mass spectrometry. A) Low-resolution MS data of the N-terminal leader sequence (denoted as (3) in Fig. 2). B) Deconvoluted spectrum resulting from A. This deconvolution was performed with automatic processing software, MagTran 1.02. The predicted mass for peptide (3) is 6439.1. C) Low-resolution MS data of the patellamide-C coding sequence (peptide (4) in Fig. 2) and the eptidemnamide-coding sequence (peptide (5) in Fig. 2). The calculated $[M+2H]^{2+}$ for these peptides is 704.3000 and 849.3620, respectively. Since a standard was available to verify peptide (5), it could be used as an internal standard for peptide (4) to show that it is within 10 ppm of the expected value.

A



B



C

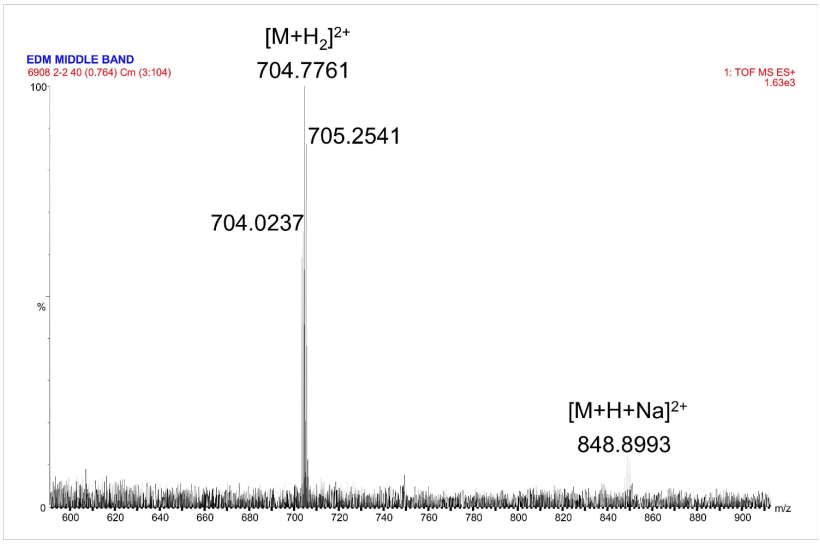
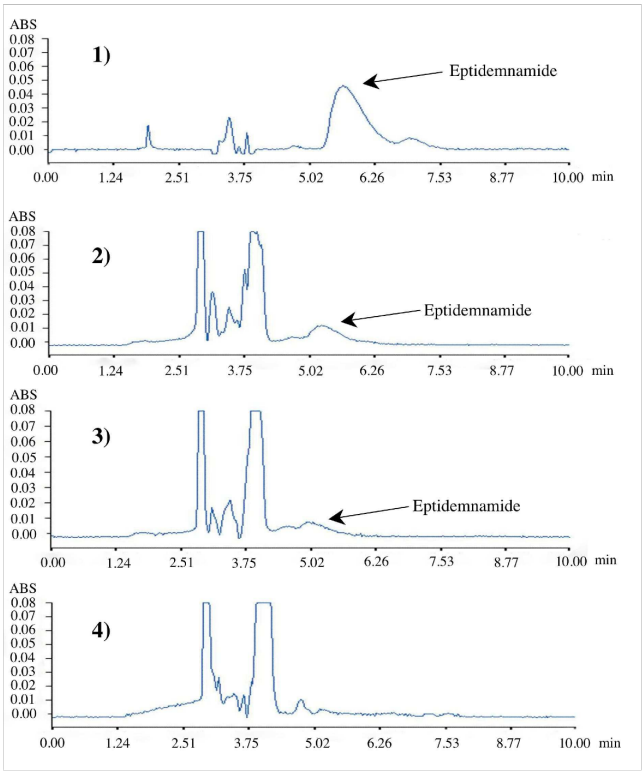
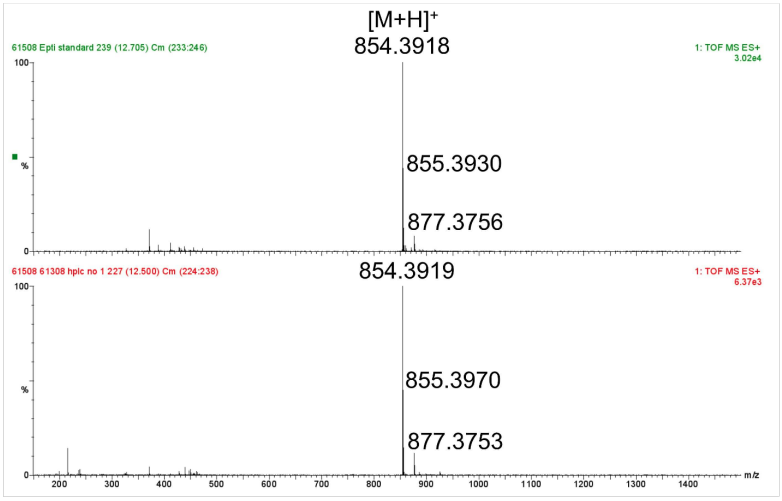

EDM MIDDLE BAND
6908 2-2 40 (0.764) Cm (3:104)
1: TOF MS ES+
1.63e3
100
%
0
[M+H2]2+
704.7761
705.2541
704.0237
[M+H+Na]2+
848.8993
600
620
640
660
680
700
720
740
760
780
800
820
840
860
880
900
m/z

Figure S5. DAD-HPLC and high resolution LC-MS analysis of PatG reactions. PatG variant was treated with synthetic peptide QGGRGDWPAYDGE for 41 h as described in the experimental. The products were analyzed as follows. A) DAD-HPLC. The y-axis represents signal intensity at 280 nm and is identical for all three spectra. 1) 10 µg synthetic standard of eptidemnamide; 2) PatA + PatG reaction with QGGRGDWPAYDGE; integration indicates that 1.5 µg eptidemnamide was synthesized; 3) PatG reaction with QGGRGDWPAYDGE; 4) QGGRGDWPAYDGE incubated with PatA, showing no synthesis of eptidemnamide. B) High resolution mass spectrometry analysis of synthetic standard (Top) and reaction product from PatA +PatG (Bottom). C) 1) Top: Data filtered for mass of 704.2 +/- 0.2 Da. $[M+H_2]^2+$ (calcd *m/z* = 1407.5923 $[M+H]^+$ for QGGRGDWPAYDGE ; Mid: Data filtered for mass of 554.2 2 +/- 0.2 Da $[M+H]^+$ (calcd *m/z* = 554.2092 $[M+H]^+$ for AYDGE; Bottom: Base peak intensity chromatogram. 2) MS data from top panel in 1), 3) MS data from middle panel in 1).

A

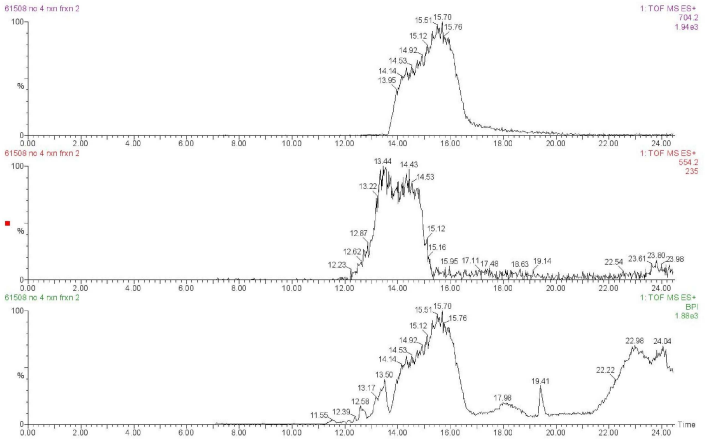


B

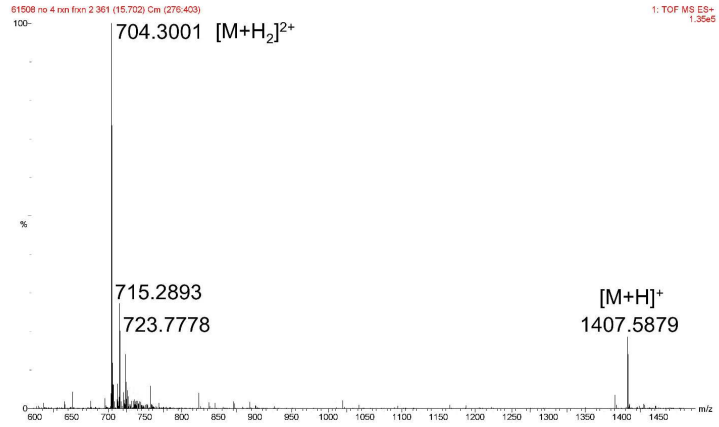

[M+H]+
854.3918
61508 Epti standard 239 (12.705) Cm (233:246)
1: TOF MS ES+
3.02e4
100
%
0
855.3930
877.3756
854.3919
61508 61308 hplc no 1 227 (12.500) Cm (224:238)
1: TOF MS ES+
6.37e3
100
%
0
855.3970
877.3753
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
m/z

C

1)


61508 no 4 rxn frxn 2
1: TOF MS ES+
704.2
1.94e3
15.51
15.70
15.12
15.76
14.92
14.53
14.14
13.95
61508 no 4 rxn frxn 2
1: TOF MS ES+
594.2
235
13.44
14.43
14.53
13.22
12.67
12.62
12.23
15.12
15.16
15.95
17.11
17.48
18.63
19.14
22.54
23.61
23.80
23.98
61508 no 4 rxn frxn 2
1: TOF MS ES+
BPI
1.88e3
15.51
15.70
15.12
15.76
14.92
14.53
14.14
13.90
13.17
12.68
11.55
12.39
17.98
19.41
22.22
22.98
24.04
Time

2)


61508 no 4 rxn frxn 2 361 (15.702) Cm (276:403)
1: TOF MS ES+
1.35e5
704.3001 [M+H2]2+
715.2893
723.7778
[M+H]+
1407.5879
m/z

3)

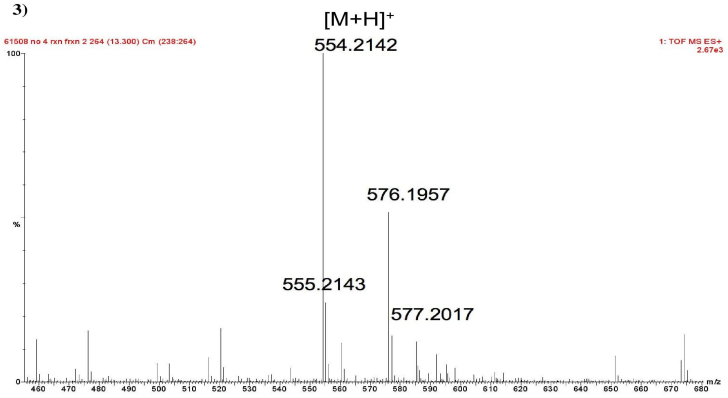
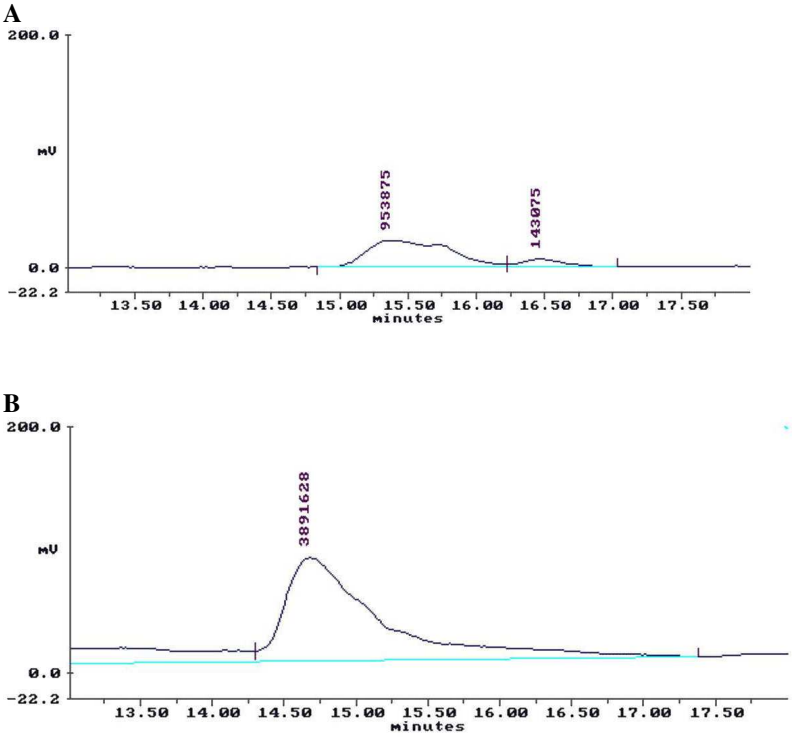

[M+H]+
554.2142
61508 no 4 rxn frxn 2 264 (13.300) Cm (239:264)
1: TOF MS ES+
2.67e3
100
%
0
576.1957
555.2143
577.2017
460 470 480 490 500 510 520 530 540 550 560 570 580 590 600 610 620 630 640 650 660 670 680
m/z

Figure S6. DAD-HPLC analysis of PatG scale-up reactions. PatG protease was treated with synthetic peptide QGGRGDWPAYDGE in a 1:3 enzyme:substrate ratio. A) DAD-HPLC of a 10 μg synthetic standard of eptidemnamide. B) DAD HPLC of a PatG reaction with QGGRGDWPAYDGE. 400 μL of a 1 mL reaction volume was injected. Integration indicates that 90 μg eptidemnamide was synthesized in this reaction from 200 μg of starting peptide, representing a 75% yield. C) A photograph of the purified, lyophilized product. (Note: 2 peaks are present in the synthetic standard because there is a Pro *cis-trans* isomerization, as defined by 2D NMR experiments and previously reported.[5] Only one isomer is present in the enzymatic product. Slight time differences are due to buffer differences; these peaks co-elute when co-injected and have been validated by MS.)



C

1. J. Sambrook, D. W. Russell, *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2001).
2. A. A. Schaffer *et al.*, *Nucleic Acids Res.* **29,** 2994 (2001).
3. C. A. Smith, H. S. Toogood, H. M. Baker, R. M. Daniel, E. N. Baker, *J. Mol. Biol.* **294**, 1027 (1999).
4. R. M. Bennett-Lovsey, A. D. Herbert, M. J. Sternberg, L. A. Kelley, *Proteins* **70**, 611 (2008).
5. M. S. Donia *et al.*, *Nat. Chem. Biol.* **2**, 729 (2006).
6. P. B. G. B. Jacob S. Nielsen, *J. Peptide Sci.* **10**, 249 (2004).

CHAPTER 4

CIRCULAR LOGIC: NONRIBOSOMAL PEPTIDE-LIKE MACROCYCLIZATION WITH A RIBOSOMAL PEPTIDE CATALYST

Manuscript reproduced with permission from:

McIntosh, J. A., Robertson, C. R, Agarwal, V., Nair, S. K., Bulaj, G. W. Schmidt, E. W. (2010) Circular logic: nonribosomal peptide-like macrocyclization with a ribosomal peptide catalyst, *J. Am. Chem. Soc. 132 (44)*, 15499-15501.

Note: my contribution to this paper was in planning, performing, and analyzing the results of all biochemical experiments with the exception of mass spectrometry data generated by the University of Utah Mass Spectrometry and Proteomic Core facility.

J|A|C|S COMMUNICATIONS

Published on Web 10/20/2010

Circular Logic: Nonribosomal Peptide-like Macrocyclization with a Ribosomal Peptide Catalyst

John A. McIntosh,[†] Charles R. Robertson,[†] Vinayak Agarwal,[‡] Satish K. Nair,[‡,§] Grzegorz W. Bulaj,[†] and Eric W. Schmidt*[,†]

Department of Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112, and Center for Biophysics and Computational Biology, Department of Biochemistry, University of Illinois, Urbana−Champaign, Illinois 61801

Received July 29, 2010; E-mail: ews1@utah.edu

Abstract: A protease from ribosomal peptide biosynthesis macrocyclizes diverse substrates, including those resembling nonribosomal peptide and hybrid polyketide−peptide products. The proposed mechanism is analogous to thioesterase-catalyzed chemistry, but the substrates are amide bonds rather than thioesters.

Macrocyclization is a common strategy to improve the rigidity and stability of bioactive metabolites.[1,2] In polyketide and nonribosomal peptide biosynthesis, macrocyclization via lactones or lactams is typically catalyzed by thioesterase (TE) domains, which contain a serine protease-like Asp-His-Ser catalytic triad (Figure 1).[1] The TE domain transfers the peptide/polyketide chain from a carrier protein to the active site serine, which is then displaced by a nucleophile, generating either a linear product or more commonly a macrocycle, as in tyrocidine A (**1**).

Quite interestingly, one of the major groups of macrocyclic ribosomal peptides, the cyanobactins,[3] is cyclized in a similar way: a subtilisin-like serine protease catalyzes cleavage of a C-terminal peptide sequence in tandem with N−C macrolactamization, leading to compounds such as patellamide C (**2**).[4−6] Among ribosomal peptides, both the cyanobactins and cyclotides are N−C cyclic, while other ribosomal peptides, such as capistruin,[7] microcin J25,[8] and the microviridins,[9] are cyclized via side-chain residues using ATP via wholly different biochemical mechanisms. There is indirect evidence that cyclotides are circularized in a similar fashion to cyanobactins,[10] and circular ribosomal peptides are common in diverse organisms. However, no definitive enzymatic or genetic studies of N−C macrocyclization have been performed on any ribosomal peptide system other than the cyanobactins.

Previously, we have shown that the subtilisin-like protease, PatG, is solely responsible for catalyzing macrocyclization in the patellamide pathway.[6] Metagenomic and biochemical analyses of the patellamide pathway showed that PatG is a broad-substrate enzyme, processing 29 known precursor peptide sequences encoding macrocycles of 7−8 amino acids.[4,5] Every natural product residue is mutated at least once in this series, and PatG could also produce the unnatural compound eptidemnamide (**3**) both *in vivo* and *in vitro* (Figure 2). Consequently, the cyanobactin macrocyclases exhibit exceptionally relaxed substrate specificity. Furthermore, unlike other ribosomal peptide natural product pathways,[11,12] the cyanobactin macrocyclases require only a C-terminal 4−5 amino acid recognition sequence, AYDG(E), which allowed us to employ short synthetic peptides as substrates.

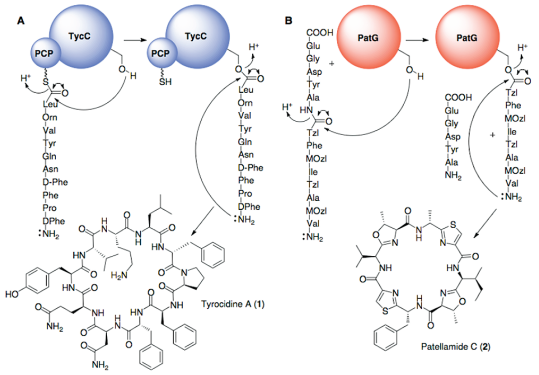


Figure 1. Macrocyclization in ribosomal and nonribosomal synthesis. (A) The nonribosomal TycC TE domain circularizes tyrocidines. PCP = peptidyl carrier protein. (B) The ribosomally acting PatG protease circularizes patellamides and many other compounds. The proposed catalytic mechanism is indicated here. Tzl = thiazol(ine); MOzl = methyloxazoline.

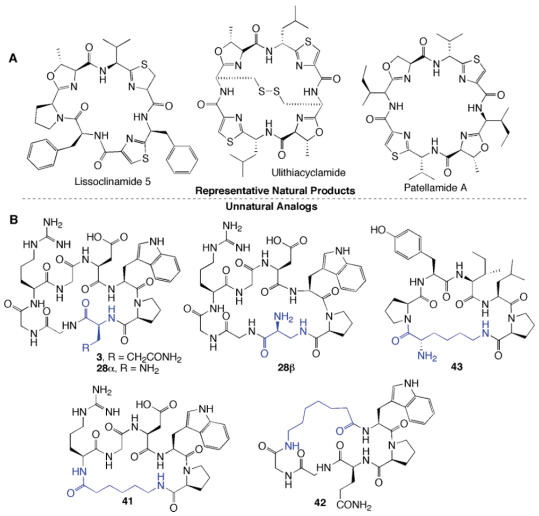


Figure 2. Products of PatG macrocyclization. (A) Representative known natural products cyclized by PatG, out of a total of 39 known natural products in this series. (B) Macrocyclization products from this study, showing side-chain circularization (**28**, **43**) and polyketide-like cyclization (**41**, **42**).

We proposed that PatG performs macrocyclization in a manner that is mechanistically analogous to TE domains, although the proteins and the substrates are quite different. In particular, while

[†] University of Utah.
[‡] Center for Biophysics and Computational Biology, University of Illinois.
[§] Department of Biochemistry, University of Illinois.

10.1021/ja1067806

Table 1. Synthetic Peptide Substrates and Macrocyclic PatG Products[a]

Compd	Sequence	Compd	Products
4	QGGRGDWPAYDGE	**3**	C
5	QGGRGDWAAYDGE	–	NR
6	(**Dap**)GGRGDWPAYDGE	**28**	C $\alpha > \beta$
7	(**Orn**)GGRGDWPAYDGE	**29**	C $\alpha{:}\delta \sim 1{:}1$
8	**K**GGRGDWPAYDGE	**30**	C $\alpha \gg \epsilon$
9	(**Glyc**)GGRGDWPAYDGE	**31**	L
10	GGRGDWPAYDGE	**32**	C
11	**q**GGRGDWPAYDGE	**33**	L
12	QGG**r**GDWPAYDGE	**34**	C
13	QGGRG**d**WPAYDGE	**35**	C
14	QGGRGD**w**PAYDGE	–	NR
15	QGGRGDW**p**AYDGE	–	NR
16	QGGWPAYDGE	–	NR
17	Q**G**GGRGDWPAYDGE	**36**	C
18	Q**GG**GGRGDWPAYDGE	**37**	C
19	Q**GGG**GGRGDWPAYDGE	**38**	C
20	**G**GGRGDWPAYDGE	**39**	C
21	**GG**GGRGDWPAYDGE	**40**	C
22	(**Ahx**)RGDWPAYDGE	**41**	C:L 1:2
23	QGG(**Ahp**)WPAYDGE	**42**	C:L 1:2
24	KPYILPAYDGE	**43**	C:L 1:2, $\alpha \ll \epsilon$
25	**K**KPYILPAYDGE	**44**	C:L 1:2
26	**K**KPY**ll**PAYDGE	–	NR
27	GWTLNSAGYLLGPAYDGE	–	NR

[a] The PatG recognition element, AYDGE, is underlined. Residues in bold indicate differences from canonical type sequences **4**, **24**, and **27**, while lower-case residues are in the D-configuration. C = circular; NR = no reaction; L = linear with predicted cleavage between P and A. If only C or L is indicated, >90% of products were in that form only.

as their name implies TEs require activated thioesters (or esters)[13] to catalyze circularization, the PatG substrates are simple amides readily accessible through standard solid phase peptide synthesis. Based on this mechanistic hypothesis, in this study we probed the capacity of PatG to circularize peptides containing nonproteinogenic amino acids and polyketide-like linkers. We show that PatG synthesizes macrocycles that are similar to those from the thiotemplate-based pathways,[13] lending weight to the mechanistic hypothesis and providing a significant step toward bridging ribosomal and nonribosomal worlds for the synthesis of complex peptide metabolites.

In this study we tested 23 analogues (**5**−**27**) to define the substrate tolerance of PatG. The analogues are of variable length and amino acid composition, but nearly all contain Pro followed by the macrocyclase recognition sequence AYDGE (Table 1). Pro was used because all of the natural compounds contain heterocycles immediately prior to AYDGE. These heterocycles are either Pro or thiazol(in)e/oxazol(in)e derived from Cys, Ser, or Thr. One exception (**5**) contained Ala in place of Pro to test this putative heterocycle requirement.

Analogues **5**−**23** are based on the previously reported PatG substrate **4**, which leads to eptidemnamide (**3**). **6**−**9** contain N-terminal residue substitutions that explore the potential for cyclization via side-chain nucleophiles, including OH in glycolate (Glyc) and NH_2 in diaminopropionate (Dap), Lys, and ornithine (Orn). **11**−**15** explore the tolerance for D-amino acids. In **10**, **16**−**21**, **24**, and **27** cyclizable sequence lengths from 5 to 11 amino acids are explored. **22** and **23** resemble polyketide−peptide hybrids, containing alkyl spacers aminohexanoic acid (Ahx) and aminoheptanoic acid (Ahp) that replace portions of the eptidemnamide sequence. Finally, to further explore substrate selectivity several substrates with wholly different sequences were attempted. **24**−**26** are based on the hormone neurotensin, while **27** mimics the neuropeptide galanin. The peptides were synthesized via standard solid phase synthesis (Supporting Information).

To assay macrocyclization, analogues were incubated with the PatG protease domain (Supporting Information). As controls, substrates were also incubated in an equivalent manner using a site-directed mutant in which the active-site Ser of PatG was replaced with Ala. Products were analyzed via matrix-assisted laser desorption ionization-mass spectrometry (MALDI-MS), which allowed us to readily observe the loss of starting material and the accumulation of products of the expected mass; for the most part these products are the only reasonable chemical entities that fit the mass. Representative products from each compound family were confirmed by high-resolution MS and MS-MS on a Fourier transform ion cyclotron resonance (FT-ICR) instrument. This method unambiguously establishes the position of circularization because numerous ions are present that only could arise if the amide bond was synthesized as shown in Figure 2. For full details, see Supporting Information. The identity of **3** was previously further confirmed by 2D NMR.[5]

For example, for compound **41** containing the polyketide-like Ahx, the predicted linear *m*/*z* would be 743 while the cyclic product would be *m*/*z* = 725, indicating a peptide with a mass decreased by 18 Da ($-H_2O$). Both ions were observed as the major nonmatrix peaks in the MALDI spectrum. This sample was then applied to LC-MS using an FT-ICR instrument, providing parent ions for both molecules that confirmed their molecular formulas, deviating by δ 0.5 ppm for the linear and δ 0.91 ppm for the cyclic variants. The MS/MS data for the cyclic product provided a very complex series of ions that overlapped the new cyclizing amide bond (Pro-Ahx) and that were inconsistent with possible alternative structures. For the linear product these ions arising from cyclization were absent, while the peaks for the linear portion were much more intense, and the resulting spectrum was much simpler.

Results are summarized in Table 1 and Figure 2b. As seen from precursor **5**, Ala is not accepted in place of Pro, supporting a requirement for a heterocyclic motif at the last position. All of the side-chain nucleophile peptides (**6**−**9**, **24**, and **25**) are substrates for the enzyme, although the OH-containing glycolate peptide **9** is linearized. Surprisingly, D-amino acids can be tolerated in some central positions (**12** and **13**), but apparently not too close to the C- or N-termini (**11**, **14**, **15**, and **26**). However, even when epimerization could be tolerated, MALDI-MS indicated that the reactions did not approach completion. PatG synthesized products in lengths from 6 to 11 amino acids (**10**, **17**−**21**, **24**), although there was some substrate-dependence as **27** was not a substrate. Peptide **16**, which could only form a five amino acid cycle, was not circularized.

Most significantly, peptides **22** and **23** containing polyketide-like linker regions were also substrates for circularization by PatG. Previously, it was shown that the tyrocidine TycC TE domain could cyclize hybrid polyketide−peptide esters, indicating a fundamental similarity in the biochemistry of these two enzyme classes despite wholly different sequences and substrates.[13]

TE domains often catalyze hydrolysis rather than macrocyclization − unnatural substrate analogues often trigger the hydrolysis reaction in place of a natural cyclization − and PatG was similar in this regard. Depending upon the substrate, PatG synthesized cyclic/linear products in ratios between 1:2 (for **22**) and >10:1 (for **8**) as determined by fluorescence HPLC and mass spectrometry. These cyclic or linear (-AYDGE) products were not formed in the active site Ser-Ala mutant of PatG. Products closely related to eptidemnamide (**3**) were always either completely (>90%) circularized or linearized, or else they were nonreactive. In short, they did not exhibit combinations of linear and cyclic products. The only

exceptions in this group were **22** and **23**, with highly flexible linkers, which were hydrolyzed in ratios reminiscent of nonribosomal TE domains.

The regioselectivity of macrocyclization for **6**−**8** and **24** was determined by derivatizing the HPLC purified cyclic peptide with 1-fluoro-2,4-dinitrobenzene, followed by acid hydrolysis. The hydrolysates were compared using HPLC and/or LC-MS with authentic standards of α- and β-amino dinitrophenyl (DNP) Dap, α- and δ-amino DNP-Orn, or α- and ε-amino DNP-Lys. The results show that side-chain cyclization via Dap, Orn, or Lys did occur in some substrates (**6**−**7**, **24**). Substrates with N-terminal Dap or Orn residues yielded a mixture of regioisomers, while substrates with N-terminal Lys showed nearly total selectivity for either backbone- (as in **8**) or side-chain-cyclized (as in **24**) peptides.

Taken together, these results show that PatG circularizes a broad array of substrates, including those with nonproteinogenic and D-amino acids and those containing polyketide-like linkers. It should be remarked that, based upon known natural products, PatG is also known to circularize an additional 29 natural substrates that encapsulate many extremely different amino acid sequences than those examined in this study. There are some limitations to substrate selectivity that are apparent in unreacted substrates, but the substrate specificity is remarkably broad and comparable to TE domains, especially given that lengths of 6−11 amino acids are effectively circularized through the terminal amino acid. It is also a remarkably broad-substrate enzyme, with a relatively short (4−5 amino acid) recognition sequence directing reactions with many different substrates and precluding water from the active site despite extreme differences in substrate length and constitution.

In this and previous *in vitro* studies using PatG, a limitation is its exceptionally slow rate, with some reactions complete in ∼24 h using a 50% catalyst load. In previous work with PatG, lower catalyst loads led to complete conversion in longer reaction periods, demonstrating that PatG acts catalytically, albeit slowly *in vitro*.[6] Moreover, *in vivo* other cyanobactin macrocyclases seem relatively efficient, where yields of up to 2.5 mg of compound per L of culture have been observed,[14] indicating that *in vitro* reactions could likely be improved. To the best of our knowledge, this report represents the first time that a ribosomal natural product catalyst has been shown to accept such extremely diverse, unnatural substrates *in vitro*.

In summary, although PatG is a catalyst from ribosomal peptide natural product synthesis that operates on amide bonds, its behavior is reminiscent of TE domains from nonribosomal and polyketide synthesis. Although this enzyme is not optimal for *in vitro* use, ultimately, it is hoped that this and other studies will provide a toolkit for genetic engineering of diverse small molecules *in vivo* and will help to bridge nonribosomal and ribosomal biosynthesis.[15] It is relatively straightforward to engineer peptide production *in vivo* using ribosomal synthesis, but nonribosomal machinery leads to much more chemically diverse products. Nonproteinogenic amino acids, such as those found in nonribosomal peptides, can already be ribosomally encoded using existing technology.[16] Ultimately, a combination of tools such as the enzymatic methods described here in concert with the ability to add unusual functions using the ribosome itself will enable the production of chemically diverse products *in vivo*. By bridging the biochemistry of these two worlds, the goal is to take advantage of the engineering simplicity of the ribosome while synthesizing the elaborate products more typical of complex nonribosomal peptides.

Acknowledgment. At U. Utah, we thank S. Endicott and R. Schackmann for synthesizing many of the analogues used in this study; C. Hubbard and A. Barrios for peptide purification assistance; D. Davis for helpful discussions; J. Muller, K. Parsawar, and C. Nelson for mass spectrometry assistance; and J. Olsen and D. Tianero for help with NMR. This work was partly funded by an ACS Medicinal Chemistry predoctoral fellowship funded by Sanofi Aventis to J.A.M. as well as NIH Grant GM071425 to E.W.S.

Supporting Information Available: Full methods and data are available. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kopp, F.; Marahiel, M. A. *Nat. Prod. Rep.* **2007**, *24*, 735–49.
(2) Burton, P. S.; Conradi, R. A.; Ho, N. F.; Hilgers, A. R.; Borchardt, R. T. *J. Pharm. Sci.* **1996**, *85*, 1336–40.
(3) Donia, M. S.; Schmidt, E. W.; Lew, M.; Hung-Wen, L. In *Comprehensive Natural Products II*; Elsevier: Oxford, 2010; pp 539−558.
(4) Donia, M. S.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2008**, *4*, 341–3.
(5) Donia, M. S.; Hathaway, B. J.; Sudek, S.; Haygood, M. G.; Rosovitz, M. J.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2006**, *2*, 729–35.
(6) Lee, J.; McIntosh, J. A.; Hathaway, B. J.; Schmidt, E. W. *J. Am. Chem. Soc.* **2009**, *131*, 2122–4.
(7) Knappe, T. A.; Linne, U.; Robbel, L.; Marahiel, M. A. *Chem. Biol.* **2009**, *16*, 1290–8.
(8) Duquesne, S.; Destoumieux-Garzon, D.; Zirah, S.; Goulard, C.; Peduzzi, J.; Rebuffat, S. *Chem. Biol.* **2007**, *14*, 793–803.
(9) Ziemert, N.; Ishida, K.; Liaimer, A.; Hertweck, C.; Dittmann, E. *Angew. Chem., Int. Ed.* **2008**, *47*, 7756–9.
(10) Henriques, S. T.; Craik, D. J. *Drug Discov. Today* **2010**, *15*, 57–64.
(11) Cheung, W. L.; Pan, S. J.; Link, A. J. *J. Am. Chem. Soc.* **2010**, *132*, 2514–5.
(12) Philmus, B.; Guerrette, J. P.; Hemscheidt, T. K. *ACS Chem. Biol.* **2009**, *4*, 429–34.
(13) Kohli, R. M.; Burke, M. D.; Tao, J.; Walsh, C. T. *J. Am. Chem. Soc.* **2003**, *125*, 7160–1.
(14) Donia, M. S.; Ruffner, D. E.; Cao, S.; Schmidt, E. W. **2010**, *Appl. Environ. Microbiol.* Submitted.
(15) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *Nat. Prod. Rep.* **2009**, *26*, 537–559.
(16) Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem.* **2010**, *79*, 413–44.

JA1067806

Supporting online material for

Circular logic: Nonribosomal peptide-like macrocyclization with a ribosomal peptide catalyst

John A. McIntosh,[1] Charles R. Robertson,[1] Vinayak Agarwal,[2] Satish K. Nair,[2,3] Grzegorz W. Bulaj,[1] and Eric W. Schmidt[1]*

1. Department of Medicinal Chemistry, University of Utah, Salt Lake City, UT 84112 USA

2. Center for Biophysics and Computational Biology 3. Department of Biochemistry, University of Illinois at Urbana-Champaign, IL 61801 USA

ews1@utah.edu

This file contains:

Experimental methods

Tables S1-S3

Figures S1-S5

Experimental methods

General methods

All HPLC separations were performed on a Hitachi LaChrom Elite system. MALDI-MS analysis of peptide cyclization reactions was performed on a Micromass MALDI micro MX (Waters) instrument using an automated targeting protocol. LC-FT-ICR analysis was performed on a LTQ FT Ultra Hybrid Mass Spectrometer (Thermo Scientific). MALDI-MS of DNP-Lys derivatives was performed on a DE-STR (PerSeptive Biosystems/ABI) instrument. Low-resolution LC-MS of DNP-Dap derivatives was performed on a Micromass ZQ (Waters) instrument equipped with a Waters 2487 HPLC system. ^{13}C and ^{1}H NMR were performed on a 400 MHz NMR spectrometer (Varian).

Peptide synthesis

Peptides were synthesized using Fmoc-based solid-phase peptide synthesis. Fmoc-Amino acids, piperidine, *N,N*-diisopropylethylamine (DIPEA) and preloaded Rink-amide glutamate-resin (substitution 0.58 mM/g) were purchased from Chem-Impex, Intl. Coupling reagent benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) was purchased from ChemPep, Inc. All other chemicals were purchased from Sigma-Aldrich. Synthesis was carried out on a Protein Technologies, Inc. Symphony multichannel synthesizer. Briefly, amino acids (200 mM in NMP) were mixed with resin (100 mM scale, 58 mg), PyBOP (200 mM), and DIPEA (400 mM) in a double coupling fashion (45 min and 30 min), washing with 70/30 DMF/NMP between couplings. Fmoc was removed by 10 min treatments (x2) of 20% piperidine. Peptides were cleaved from resin by treatment with reagent K (1 mL per 100 mg resin; 82.5% trifluoroacetic acid, 5%

nH_2O, 5% ethanedithiol, 2.5% thioanisole v/v, 75 mg/mL phenol) and filtered. The flow-through was treated with MTBE (3x volume) and chilled to -20 °C for 1 h. The precipitated peptides were then purified on a 219TP1010 semi-prep diphenyl column (Vydac) on a linear gradient from 95% buffer A (H_2O with 0.1% TFA), 5% buffer B (90% acetonitrile, 10% nH_2O, 0.1% TFA) to 60% buffer A, 40% buffer B. Synthesis of the correct products was confirmed by MALDI-MS (see Table S1). Initially, the neurotensin analogue **24** (KPYILPAYDGE) was synthesized as KKPYILPAYDGE **25** (see Table S1). During the course of enzyme reactions, however, a hydrolytic removal of the N-terminal Lys residue gave predominantly **24**. When the PatG Ser→Ala mutant was incubated with **25**, **24** was produced, indicating that PatG did not catalyze removal of the N-terminal Lys. Although both **24** and **25** were circularized to yield **43** and **44,** respectively, owing to the small amount of **44** that was produced, its products were not subjected to regioselectivity analysis.

Protein expression

The PatG clone used in this study was a 6X-His tagged truncation of PatG spanning residues 513-866. An active site mutant was also cloned using the QuikChange™ Site-Directed Mutagenesis Kit (Stratagene). Clones were verified by DNA sequencing. PatG 513-866 and the active site mutant, PatG 513-866 S783A, were expressed in BL21 DE3 cells (Novagen) grown in 8 L of LB broth. The outgrowth was performed at 30 ºC. Once the cells reached an OD_{600} of 0.4, the temperature was reduced to 18 ºC, and the synthetic inducer isopropyl β-D-1-thiogalactopyranoside (IPTG) was added (0.2 mM). Growth was allowed to continue overnight, and cells were harvested by centrifugation. Osmotic

shock lysis was performed according to previously established procedures.[1] Cell debris was then pelleted by centrifugation for 20 minutes at 13,000 RPM in a JA-20 rotor. Cleared lysates were then filtered with 5, 0.45, and 0.2 μm filters, and then applied to a gravity column containing 10 mL of Ni-NTA resin (Qiagen). The column was washed with 10 column volumes of wash buffer (500 mM NaCl, 25 mM imidazole pH 8.0), followed by 3 x 15 mL of elution buffer (750 mM NaCl, 250 mM imidazole pH 8.0). The eluents were analyzed by SDS-PAGE and dialyzed three times against 50 mM NaCl, 20mM tris pH 8.0. Finally, the protein was stocked in 50 mM NaCl, 20 mM tris pH 8.0, 10% glycerol, and frozen at -80 ºC.

Enzyme reactions

Peptides (100 μM) were incubated with PatG (50 μM) along with tris pH 8.0 (20 mM), DTT (3 mM), $CaCl_2$ (1 mM), and FMN (10 μM). Reactions were incubated for 24 h at 34 ºC, and then frozen at -80 ºC until analysis. Control reactions were performed with a PatG active site mutant. Samples analyzed by MALDI-MS were desalted and concentrated using C_{18} ZipTips (Millipore) according to the manufacturer's instructions. After MALDI analysis, reactions were submitted to the University of Utah Mass Spectrometry and Proteomics core facility for FT-ICR analysis. The yield of reactions containing **22** and **24** were calculated via comparison of enzyme reaction products to a calibration curve of fluorescence intensity of Tyr, Trp, and/or to starting material. The relative ratios of cyclic to linear products for **6** and **23** were assessed via ion abundance in LC FT-ICR. Lastly, the relative ratios of cyclic and linear products in reactions containing **7** and **8** was assessed by comparing the integrated peak areas of the fluorescence chromatograms.

Determination of regioselectivity

Products were purified from reaction mixtures by reversed phase HPLC (**29** and **30** were purified using a 214TP1010 Vydac C4 semi-preparative column on a gradient from 100% buffer A (H_2O with 0.05% TFA) to 50% buffer A / 50% buffer B (Acetonitrile); **28** and **43** were purified using a 219TP54 Vydac Diphenyl column on a gradient from 100% buffer A (H_2O with 0.07% TFA) to 40% buffer A / 60% buffer B (Acetonitrile with 0.07% TFA). Purity of cyclic peptides was assessed by MALDI-MS. The DNFB derivitizations were conducted according to previously established procedures.[2] Briefly, cyclic peptide fractions were resuspended in buffer in 0.65 mL PCR tubes (40 μL; 0.1 M borate pH 9.0). To the dissolved cyclic peptides a DNFB solution was added (60 μL; 10 mM in Acetonitrile). The combined solutions were then mixed by vortexing and then placed in a thermocycler at 65 °C for 30 minutes, after which time they were frozen at -80 °C. Subsequently, the reactions were transferred to 1 mL glass reaction vessels and dried by speedvac. The dried material was resuspended in a solution of HCl (0.1 mL; 6 M in H_2O), capped with silicone lids, and placed in a sand bath at 110°C for approximately 12 h, when the temperature of the sand bath was lowered and the reactions were blown dry under a stream of argon. Products were then directly analyzed by HPLC, LC-MS, or UHPLC-MS in comparison with authentic standards (Figure S4).

Synthesis of DNP-Dap, -Orn and -Lys derivatives

Standards of 3-amino-2-[(2,4-dinitrophenyl)amino]propanoic acid (α-DNP-Dap), 2-amino-3-[(2,4-dinitrophenyl)amino]propanoic acid (β-DNP Dap), 5-amino-2[(2,4-dinitrophenyl)amino]pentanoic acid (α-DNP Orn), 2-amino-5[(2,4-

dinitrophenyl)amino]pentanoic acid (δ-DNP Orn) as well as 6-amino-2-[(2,4-dinitrophenyl)amino]hexanoic acid (α-DNP Lys) and 2-amino-6-[(2,4-dinitrophenyl)amino]hexanoic acid (ε-DNP Lys) were prepared via previously established procedures.[3,4] Surprisingly, although DNP-amino acid derivatives have been used routinely in peptide chemistry for 50 years, we could not find NMR data for the above compounds. We therefore subjected our derivatives to NMR analysis, the results of which are presented below. DNP-Dap regioisomers were separated using a 214TP104 C_4 column (Vydac) on an isocratic gradient of aqueous ammonium sulfate (1 M). DNP-Lys and DNP-Orn regioisomers were preparatively separated using a 100 x 10 mm Onyx monolithic semi-prep C_{18} column (Phenomenex) on a gradient from 99% H_2O/1% acetonitrile (AcN) to 100% AcN over 2h.

3-amino-2-[(2,4-dinitrophenyl)amino]propanoic acid (α-DNP-Dap)

ESI-MS-(+): *m*/*z*=271.0 [M+H$^+$]; expected=271.1. By NMR, this derivative was observed as a mixture of two stable conformers: ^{1}H NMR (D_2O, 400 MHz) δ 9.12 (1H, t, *J* = 2.7 Hz), 8.28 (1H, dd, *J* = 2.7, 7.0 Hz), 8.26 (1H, dd, *J* = 2.7, 7.4 Hz), 6.98 (1H, d, *J* = 9.8 Hz), 6.94 (1H, d, *J* = 9.3 Hz), 4.32 (1H, m), 3.70 (1H, dd, *J* = 3.9, 14.5 Hz), 3.47 (1H, dd, *J* = 7.0, 14.5 Hz), 3.16 (1H, dd, *J* = 4.3, 13.7 Hz), 3.10 (1H, 5.9, 13.7 Hz). ^{13}C NMR (D_2O, 100 MHz) δ 176.8, 176.6, 171.5, 148.2, 135.8, 130.7, 130.5, 124.8, 115.2, 60.5, 59.6, 42.8, 43.6.

2-amino-3-[(2,4-dinitrophenyl)amino]propanoic acid (β-DNP Dap)

ESI-MS-(+): *m*/*z*=271.0 [M+H$^+$]; expected=271.1. ^{1}H NMR (D_2O, 400 MHz) δ 8.97 (1H,

s), 8.20 (1H, d, *J* = 9.8 Hz), 7.08 (1H, d, *J* = 9.0 Hz), 3.9 (3H, m). ^{13}C NMR (D_2O, 100 Mhz) δ 173.4, 149.1, 141.9, 135.6, 130.6, 123.8, 115.0, 54.9, 47.3.

6-amino-2-[(2,4-dinitrophenyl)amino]hexanoic acid (α-DNP Lys)

MALDI-MS-(+): *m*/*z*=313.1 [M+H$^+$]; expected=313.1. ^{1}H NMR (D_2O, 400 Mhz) δ 9.07 (1H, s), 8.24 (1H, d, *J* = 9.5 Hz), 6.91 (1H, d, *J* = 9.0 Hz), 4.21 (1H, t, *J* = 6.25 Hz), 2.95 (1H, t, *J* = 6 Hz), 2.85 (1H, t, *J* = 7.42 Hz), 1.94 (2H, m), 1.61 (2H, m), 1.44 (2H, m). ^{13}C NMR (DMSO d_6, 400 Mhz) δ 170.7, 147.2, 134.4, 130.5, 129.2, 124.8, 116.8, 58.3, 42.4, 34.9, 32.2, 22.5.

2-amino-6-[(2,4-dinitrophenyl)amino]hexanoic acid (ε-DNP Lys)

MALDI-MS-(+): *m*/*z*=313.0 [M+H$^+$]; expected=313.1. ^{1}H NMR (D_2O, 400 Mhz) δ 9.11 (1H, d, *J* = 2.7 Hz), 8.27 (1H, dd, *J* = 2.3, 9.8 Hz), 7.13 (1H, d, *J* = 9.8 Hz), 3.72 (1H, t, *J* = 6 Hz), 3.53 (2H, t, *J*=7 Hz), 1.89 (2H, m), 1.79 (2H, m), 1.51 (2H, m). ^{13}C NMR (DMSO d_6, 100 Mhz) δ 170.2, 148.3, 135.3, 130.7, 130.3, 124.4, 116.0, 54.8, 43.4, 31.5, 28.6, 23.2.

5-amino-2-[(2,4-dinitrophenyl)amino]pentanoic acid (α-DNP Orn)

ESI-MS-(+): *m*/*z*=299.1 [M+H$^+$]; expected=299.1. ^{1}H NMR (D_2O, 400 MHz) δ 8.99 (1H, d, *J* = 2.35 Hz), 8.16 (1H, dd, *J* = 2.34, 9.57 Hz), 6.82 (1H, d, *J* = 9.76 Hz), 4.17 (1H, t, *J* = 6 Hz), 2.87 (2H, t, *J* = 6.4 Hz), 1.88 (2H, m), 1.63 (2H, m). ^{13}C NMR (DMSO d_6, 100 Mhz) δ 147.1, 134.9, 130.6, 130.2, 124.7, 116.8, 57.4, 29.3, 24.0.

2-amino-5-[(2,4-dinitrophenyl)amino]pentanoic acid (δ-DNP Orn)

ESI-MS-(+): *m*/*z*=299.1 [M+H$^+$]; expected=299.1. ^{1}H NMR (D_2O, 400 MHz) δ 8.96 (1H, d, *J* = 2.34 Hz), 8.14 (1H, dd, *J* = 2.73, 9.5 Hz), 7.00 (1H, d, *J* = 9.76), 3.53 (1H, t, *J* = 6.25), 3.42 (2H, t, *J* = 6.64), 1.77 (2H, m), 1.65 (2H, m). ^{13}C NMR (DMSO d_6, 100 Mhz) δ 171.6, 148.8, 135.3, 130.6, 130.3, 124.4, 116.2, 54.6, 43.2, 29.5, 25.3.

Table S1. Summary of MALDI-MS data

#	Sequence	Expected *m/z*	Observed *m/z*	#	Products (* denotes linear products only)	Expected *m/z*	Observed *m/z*
5	QGGRGDW**A**AYDGE	1381.6	1381.6	-	–	n/a	n/a
6	(**Dap**)GGRGDWPAYDGE	1365.6	1365.7	**28**	(**Dap**)GGRGDWP	812.4	812.5
7	(**Orn**)GGRGDWPAYDGE	1393.6	1393.7	**29**	(**Orn**)GGRGDWP	840.4	840.5
8	**K**GGRGDWPAYDGE	1407.6	1407.8	**30**	**K**GGRGDWP	854.4	854.6
9	(**Glyc**)GGRGDWPAYDGE	1337.5	1337.4	**31**	(**Glyc**)GGRGDWP*	802.3	802.4
10	GGRGDWPAYDGE	1279.5	1279.8	**32**	GGRGDWP	726.3	726.5
11	**q**GGRGDWPAYDGE	1407.6	1407.7	**33**	**q**GGRGDWP*	872.4	872.4
12	QGG**r**GDWPAYDGE	1407.6	1407.7	**34**	QGG**r**GDWP	854.4	854.4
13	QGGRG**d**WPAYDGE	1407.6	1407.7	**35**	QGGRG**d**WP	854.4	854.5
14	QGGRGD**w**PAYDGE	1407.6	1407.6	-	–	n/a	n/a
15	QGGRGDW**p**AYDGE	1407.7	1407.7	-	–	n/a	n/a
16	QGGWPAYDGE	1079.4	1079.4	-	–	n/a	n/a
17	Q**GG**GRGDWPAYDGE	1464.6	1464.7	**36**	Q**GG**GRGDWP	911.4	911.5
18	Q**GG**GGRGDWPAYDGE	1521.6	1521.6	**37**	Q**GG**GGRGDWP	968.4	968.5
19	Q**GGG**GGRGDWPAYDGE	1578.7	1578.7	**38**	Q**GGG**GGRGDWP	1025.5	1025.5
20	**G**GGRGDWPAYDGE	1336.6	1336.6	**39**	**G**GGRGDWP	783.4	783.4
21	**GG**GGRGDWPAYDGE	1393.6	1393.6	**40**	**GG**GGRGDWP	840.4	840.4
22	(**Ahx**)RGDWPAYDGE	1278.6	1278.8	**41**	(**Ahx**)RGDWP	725.4	725.5
23	QGG(**Ahp**)WPAYDGE	1206.5	1206.6	**42**	QGG(**Ahp**)WP	653.3	675.4 ($M+Na^+$)
24	KPYILPAYDGE	1265.6	1265.7	**43**	KPYILP	840.5	840.5
25	**K**KPYILPAYDGE	1393.7	1393.9	**44**	**K**KPYILP	712.4	712.4
26	**K**KPY**ll**PAYDGE	1393.7	1393.8	-	–	n/a	n/a
27	GWTLNSAGYLLGPAYDGE	1883.9	1885.2	-	–	n/a	n/a

Table S2. Summary of FT-ICR data. Note: for MS-MS data see figure S3 below.

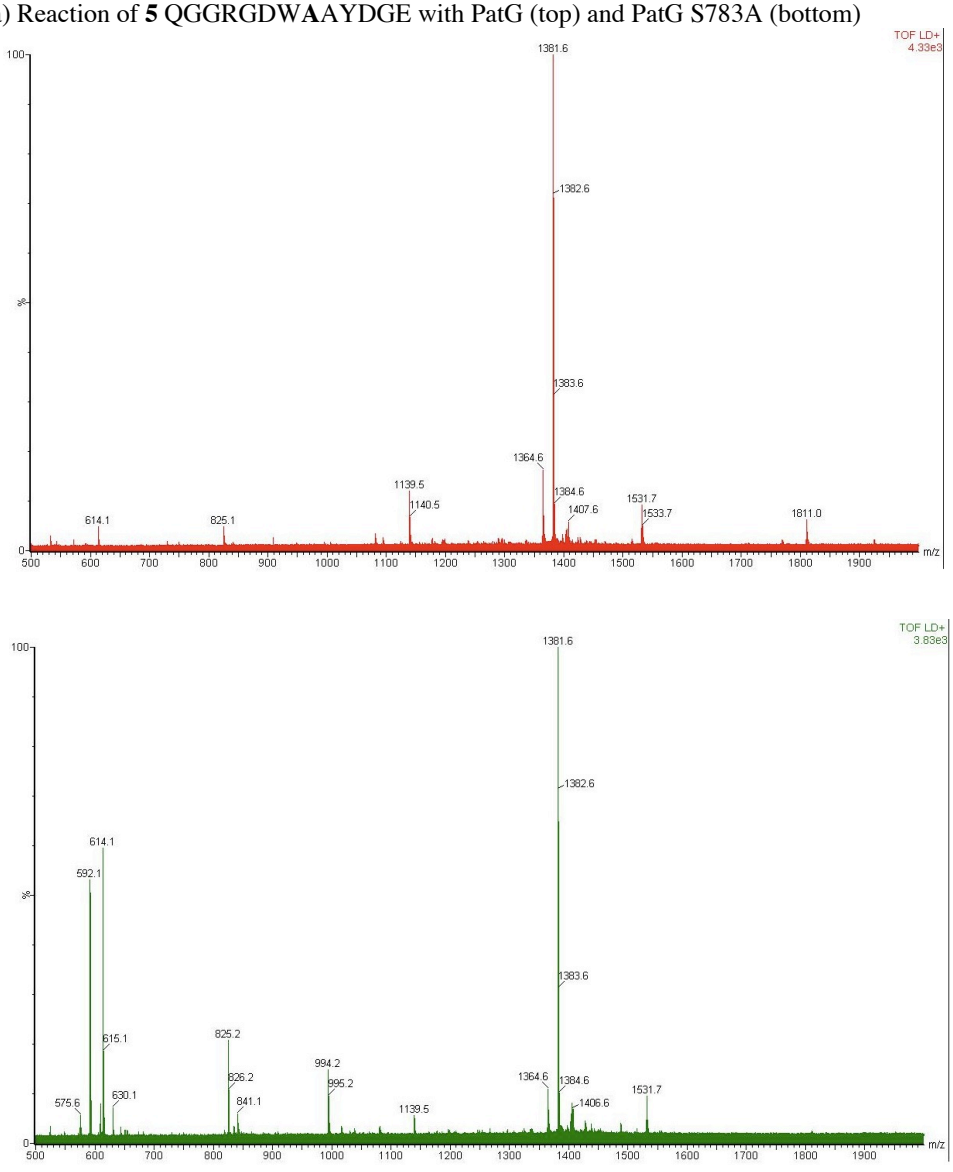
#	Sequence	Expected *m/z*	Observed *m/z*	Δppm
27	α, β-cyclo[(Dap)GGRGDWP]	812.3798, 406.6936	406.6940	0.98
40	cyclo[(Ahx)RGDWP]	725.3729	725.3736	0.91
41	cyclo[QGG(Ahp)WP]	653.3406	653.3413	1.1
42	ε-cyclo[KPYILP]	712.4392	712.4407	2.1
43	cyclo[KKPYILP]	840.5342, 420.7707	420.7714	1.7

Table S3. Ratios of cyclic to linear products

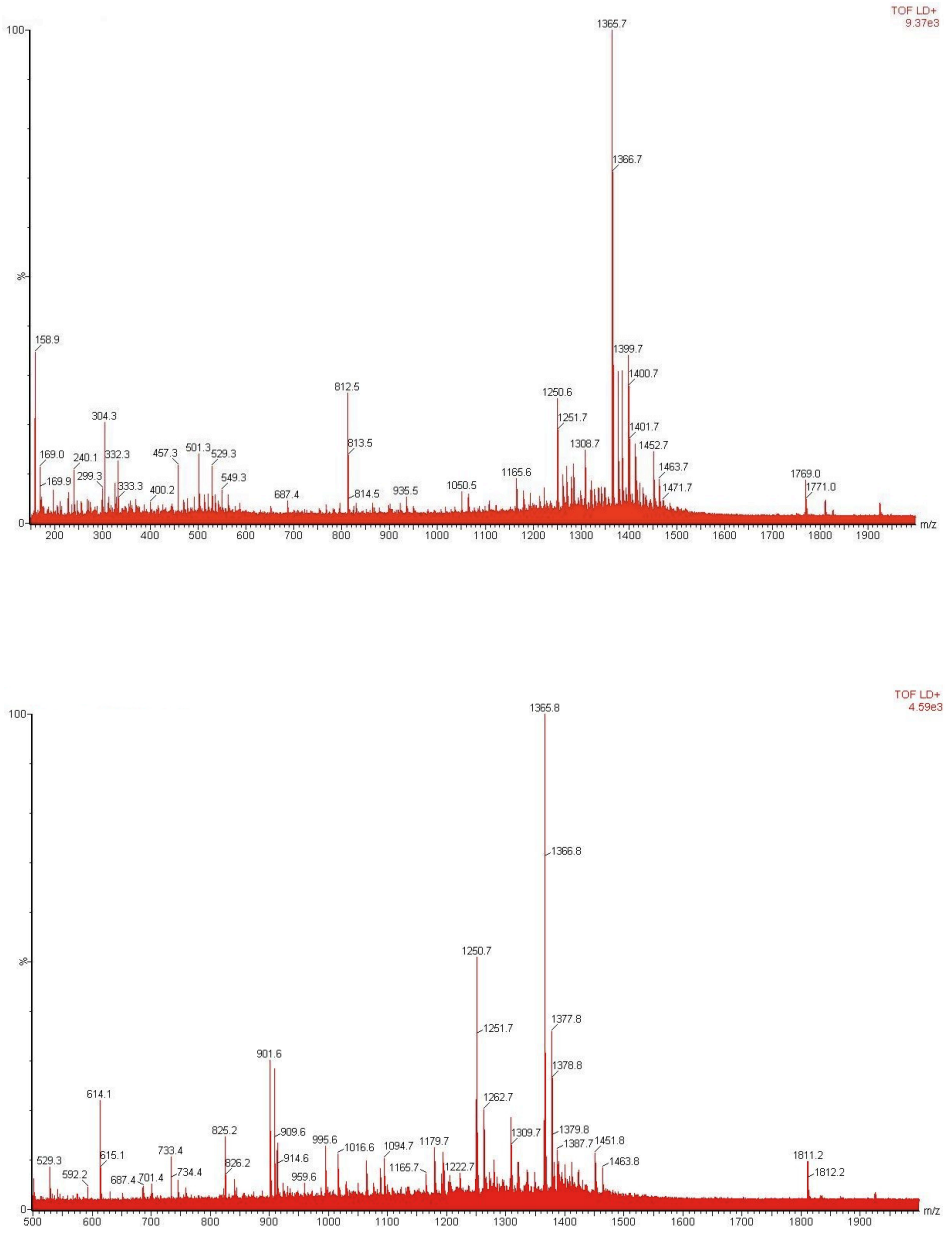
#	Cyclic : Linear	% yield cyclic	Method
6	13:1	-	FT-ICR
7	9:1	-	Fluorescence HPLC
8	17:1	-	Fluorescence HPLC
22	1:2	13%	Fluorescence HPLC
23	1:2	-	FT-ICR
24	1:2	24%	Fluorescence HPLC

Figure S1. Shown below are MALDI-MS data for reactions of PatG with substrates. At the top of each page, reactions with active enzyme are shown, while at bottom, control reactions with active site Ser->Ala mutant are shown. Each page shows the MALDI results for one synthetic peptide (i.e. **5**-**26**). (a) **5** (b) **6** (c) **7** (d) **8** (e) **9** (f) **10** (g) **11** (h) **12** (i) **13** (j) **14** (k) **15** (l) **16** (m) **17**, **20** (n) **18**, **20**, **21** (o) **19** (p) **22** (q) **23** (r) **24-25**

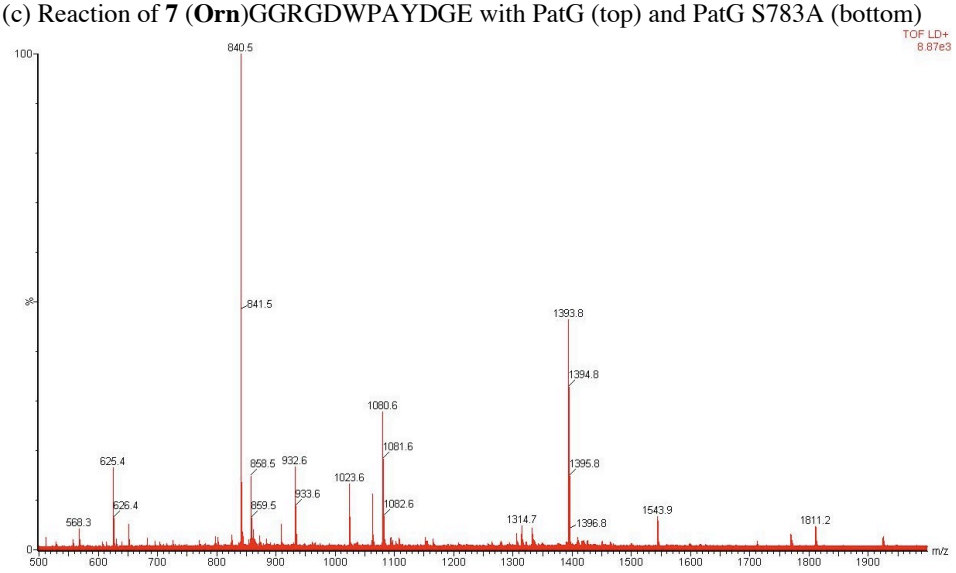
(a) Reaction of **5** QGGRGDW**A**AYDGE with PatG (top) and PatG S783A (bottom)

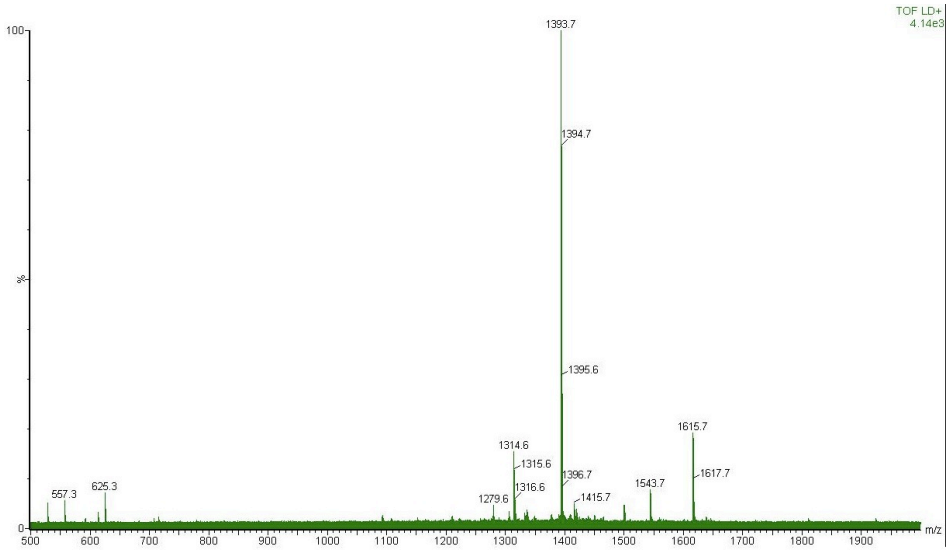


(b) Reaction of **6** (**Dap**)GGRGDWPAYDGE with PatG (top) and PatG S783A (bottom)

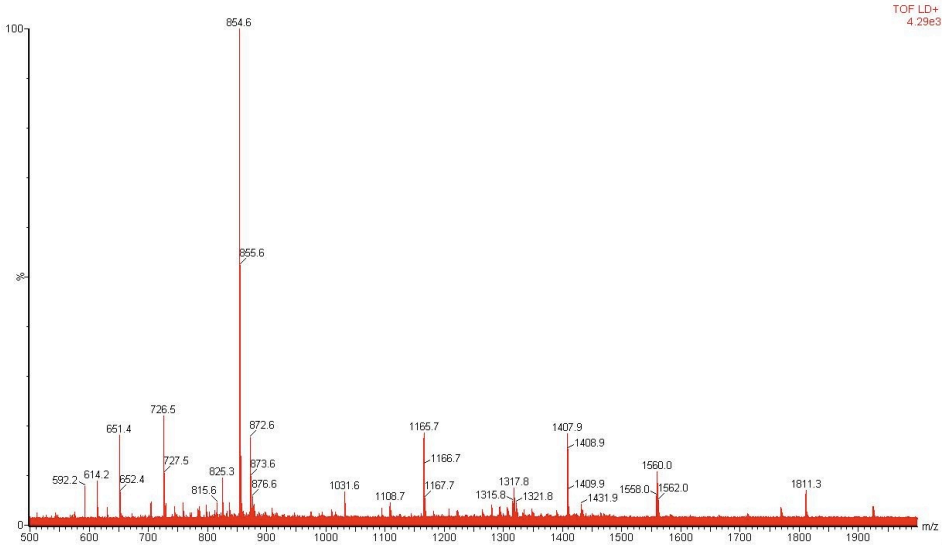


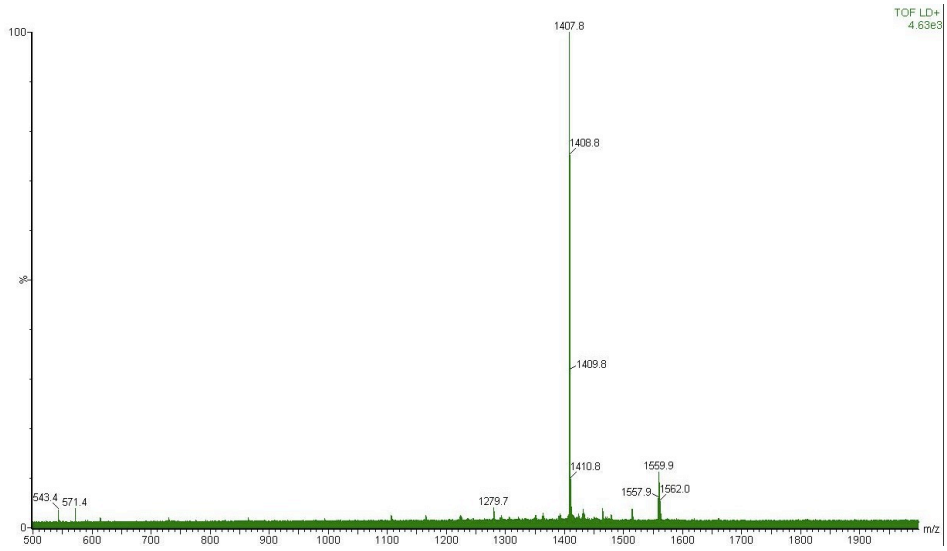
(c) Reaction of **7** (**Orn**)GGRGDWPAYDGE with PatG (top) and PatG S783A (bottom)



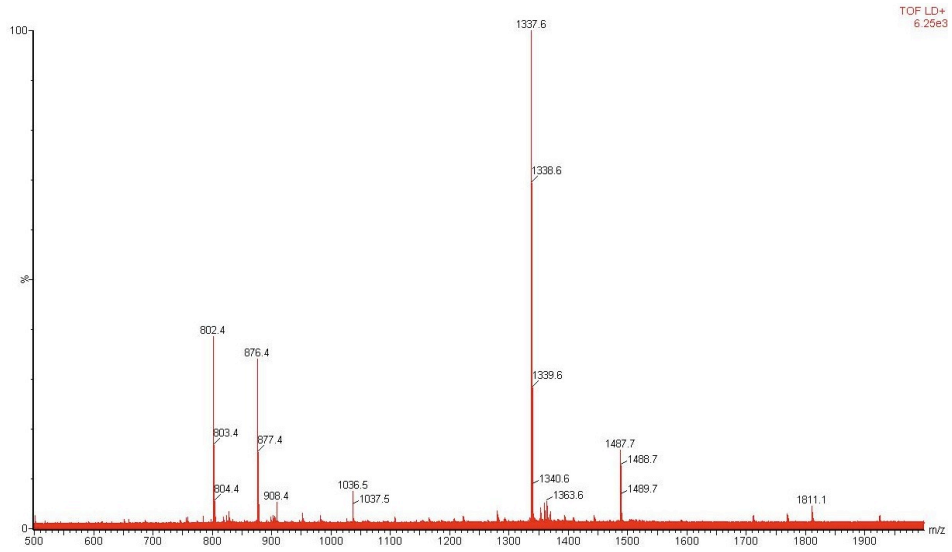


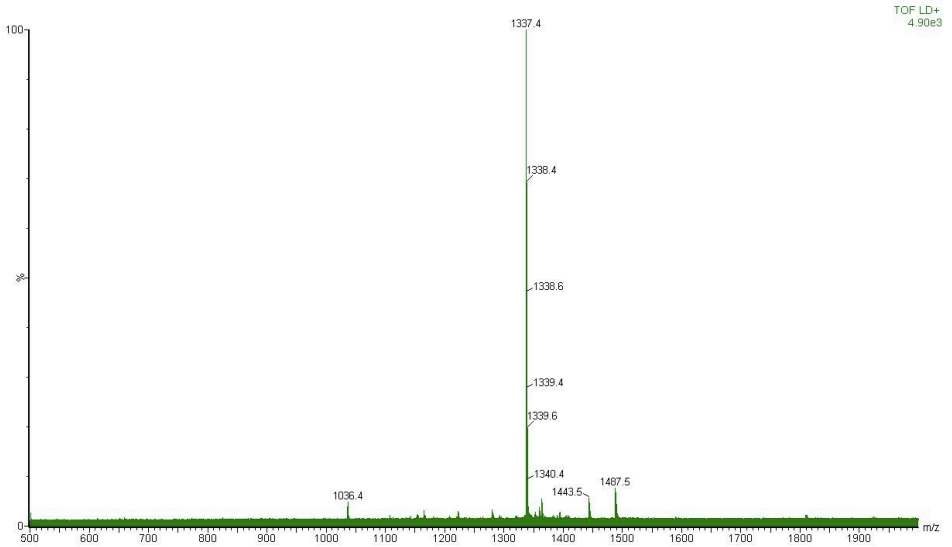
(d) Reaction of **8** **K**GGRGDWPAYDGE with PatG (top) and PatG S783A (bottom)



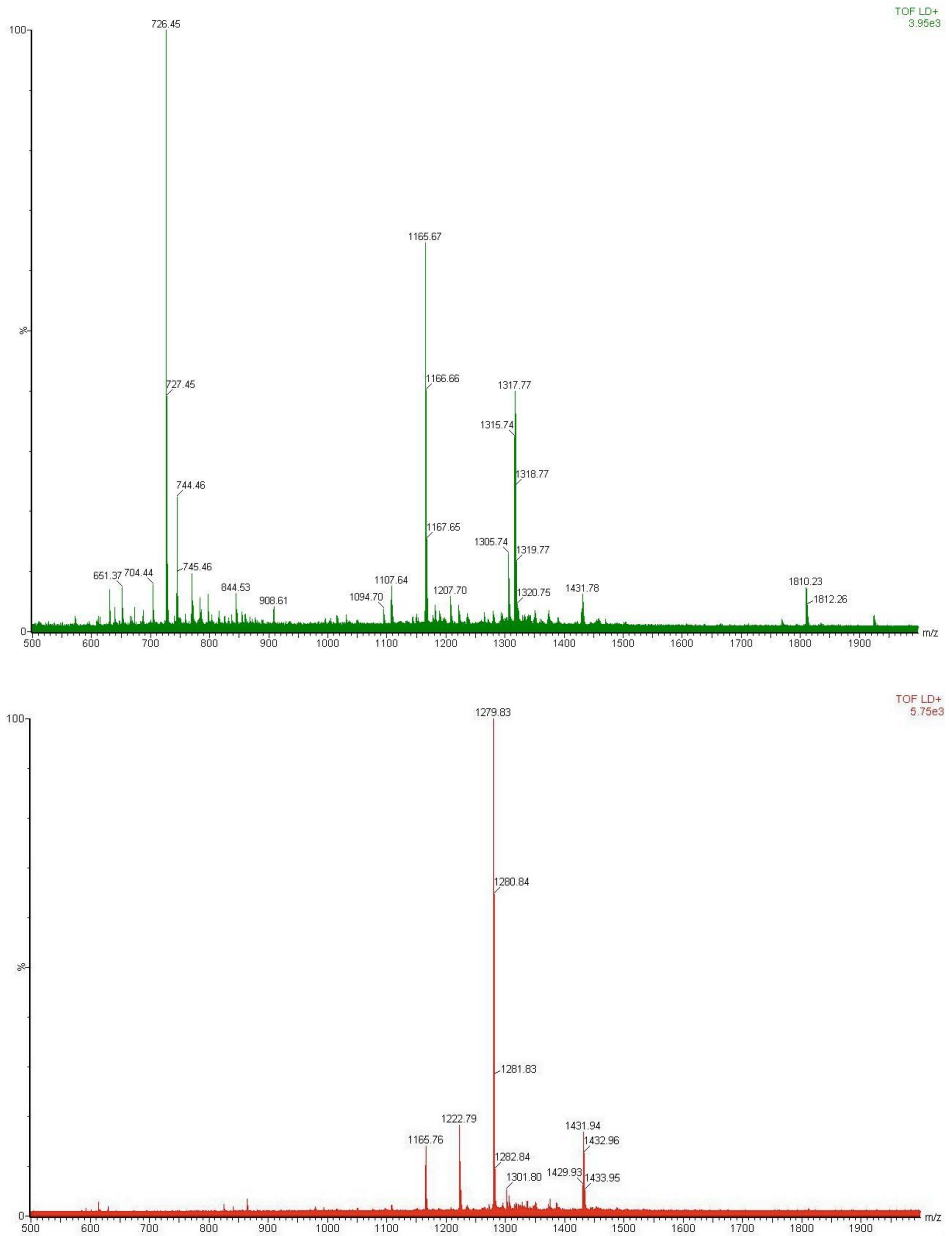


(e) Reaction of **9** (**Glyc**)GGRGDWPAYDGE with PatG (top) and PatG S783A (bottom)

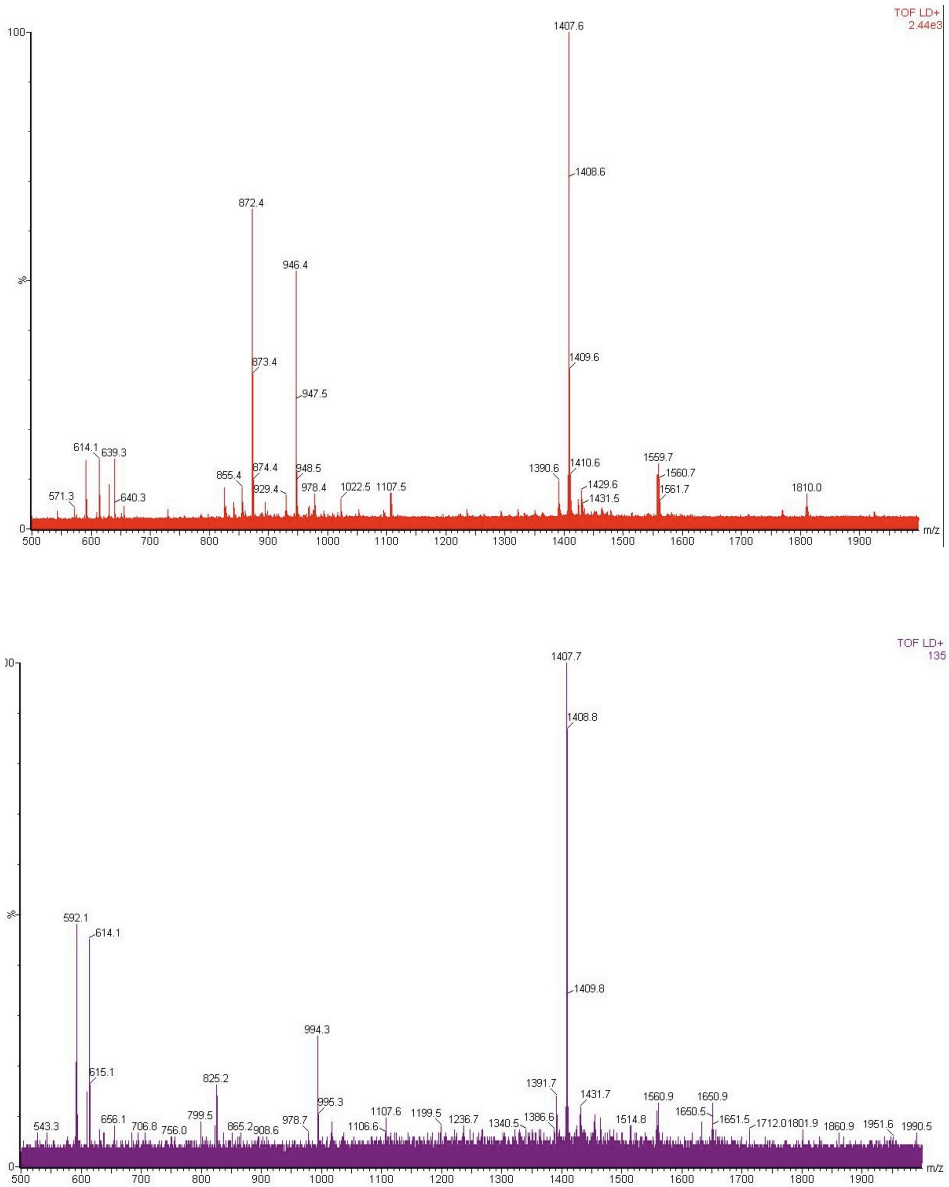




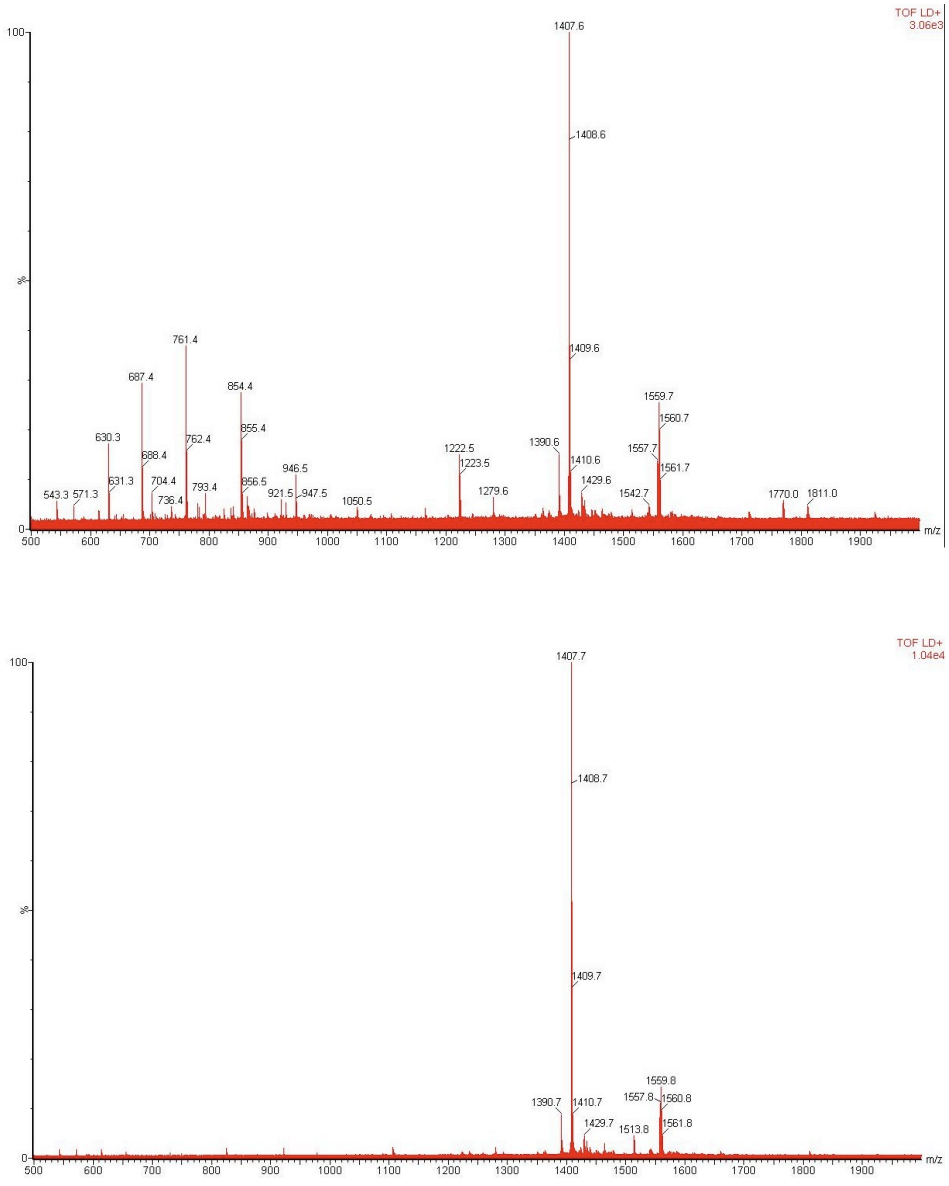
(f) Reaction of **10** GGRGDWPAYDGE with PatG (top) and PatG S783A (bottom)



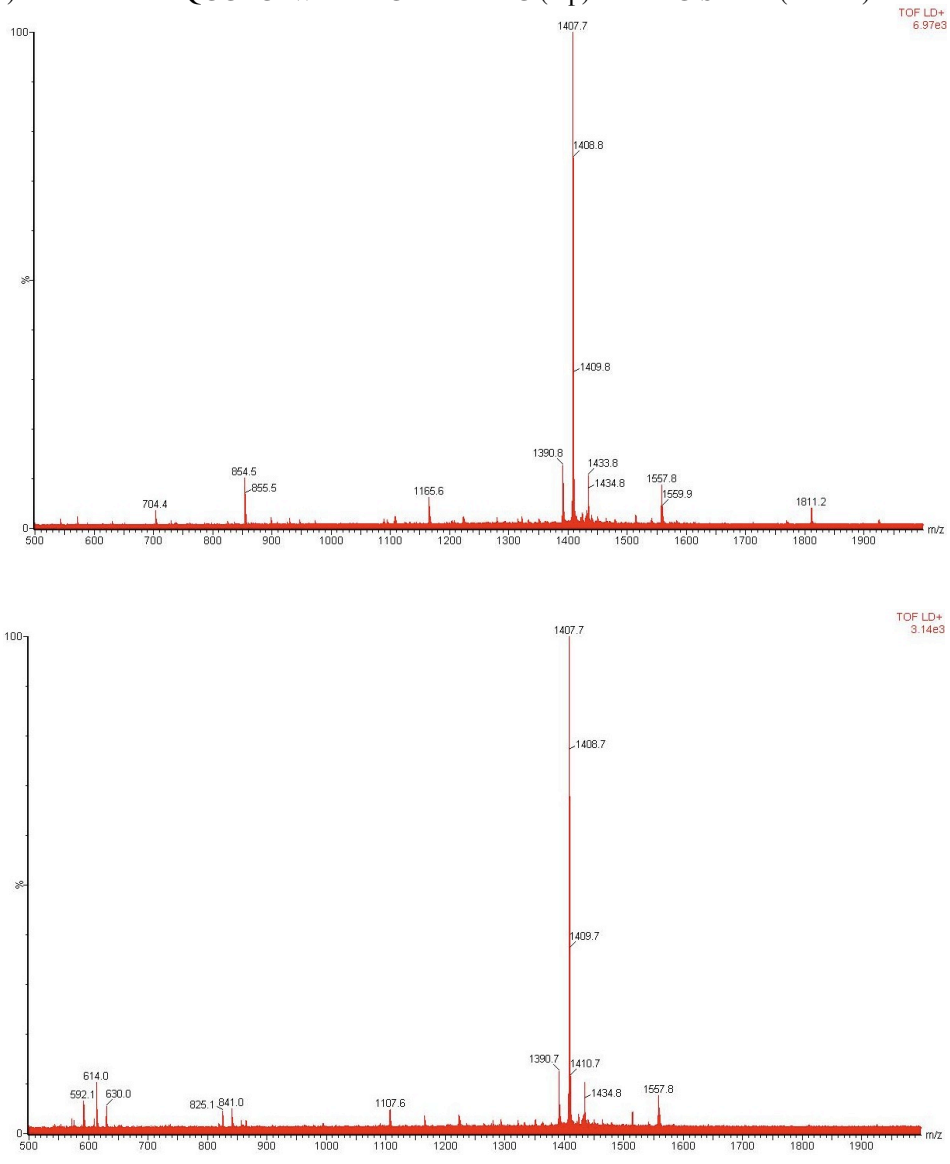
(g) Reaction of **11 q**GGRGDWPAYDGE with PatG (top) and PatG S783A (bottom)



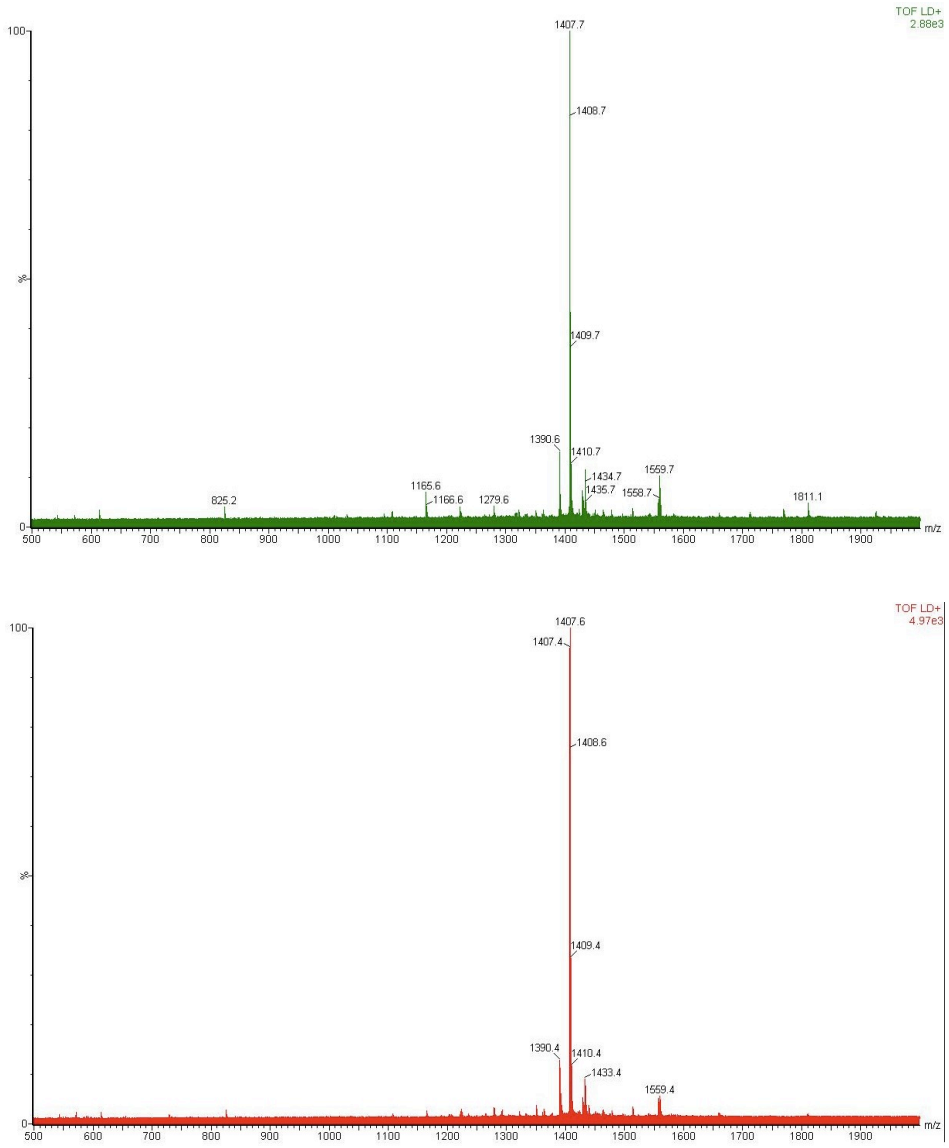
(h) Reaction of **12** QGG**r**GDWPAYDGE with PatG (top) and PatG S783A (bottom)



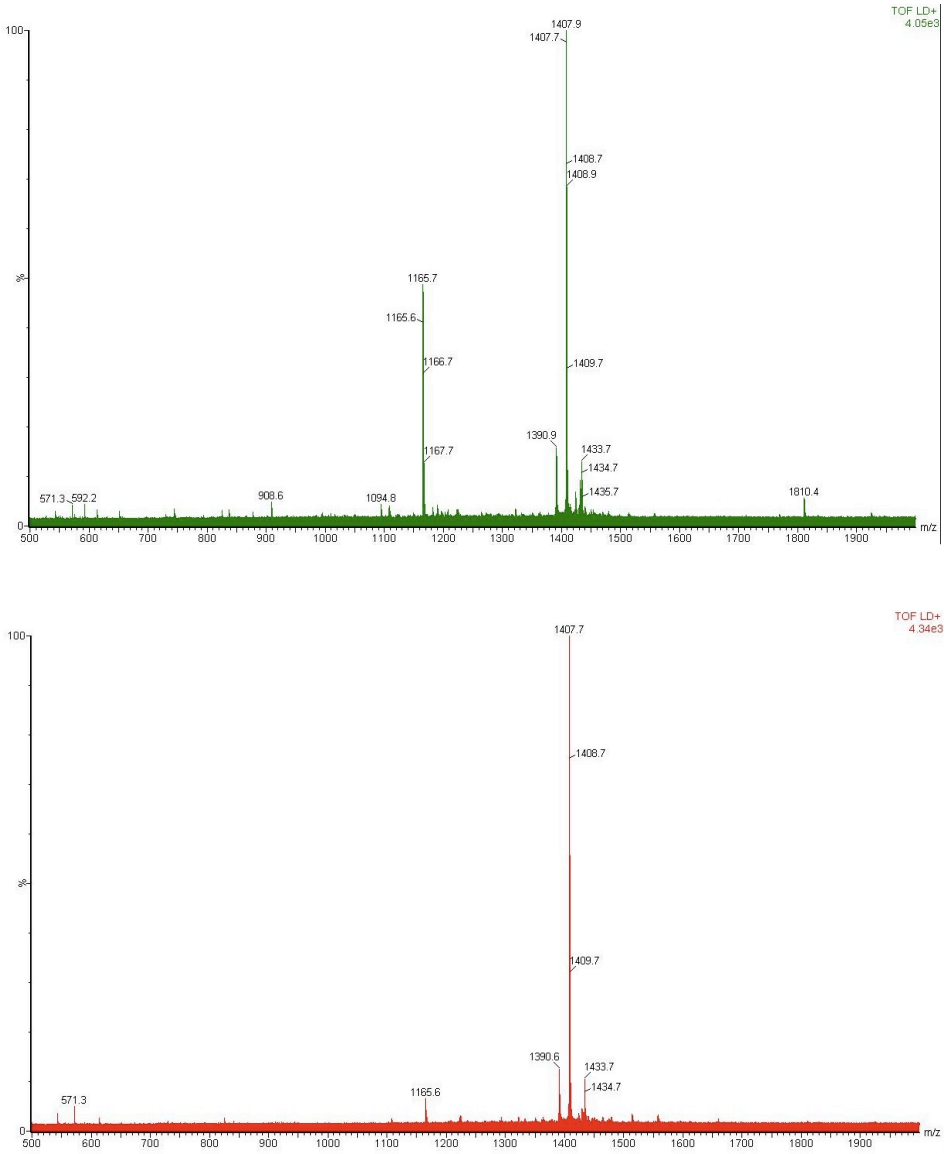
(i) Reaction of **13** QGGRG**d**WPAYDGE with PatG (top) and PatG S783A (bottom)



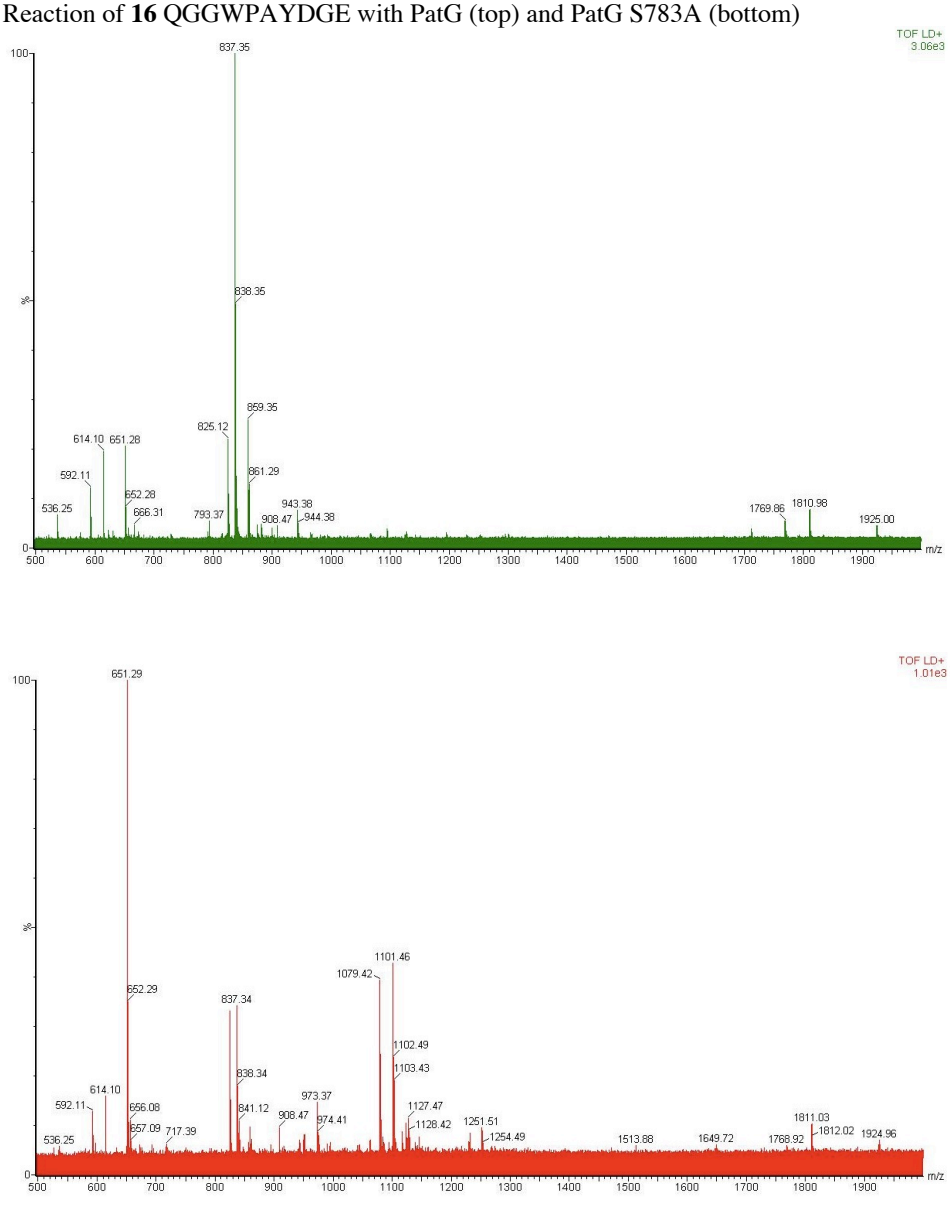
(j) Reaction of **14** QGGRGD**w**PAYDGE with PatG (top) and PatG S783A (bottom)



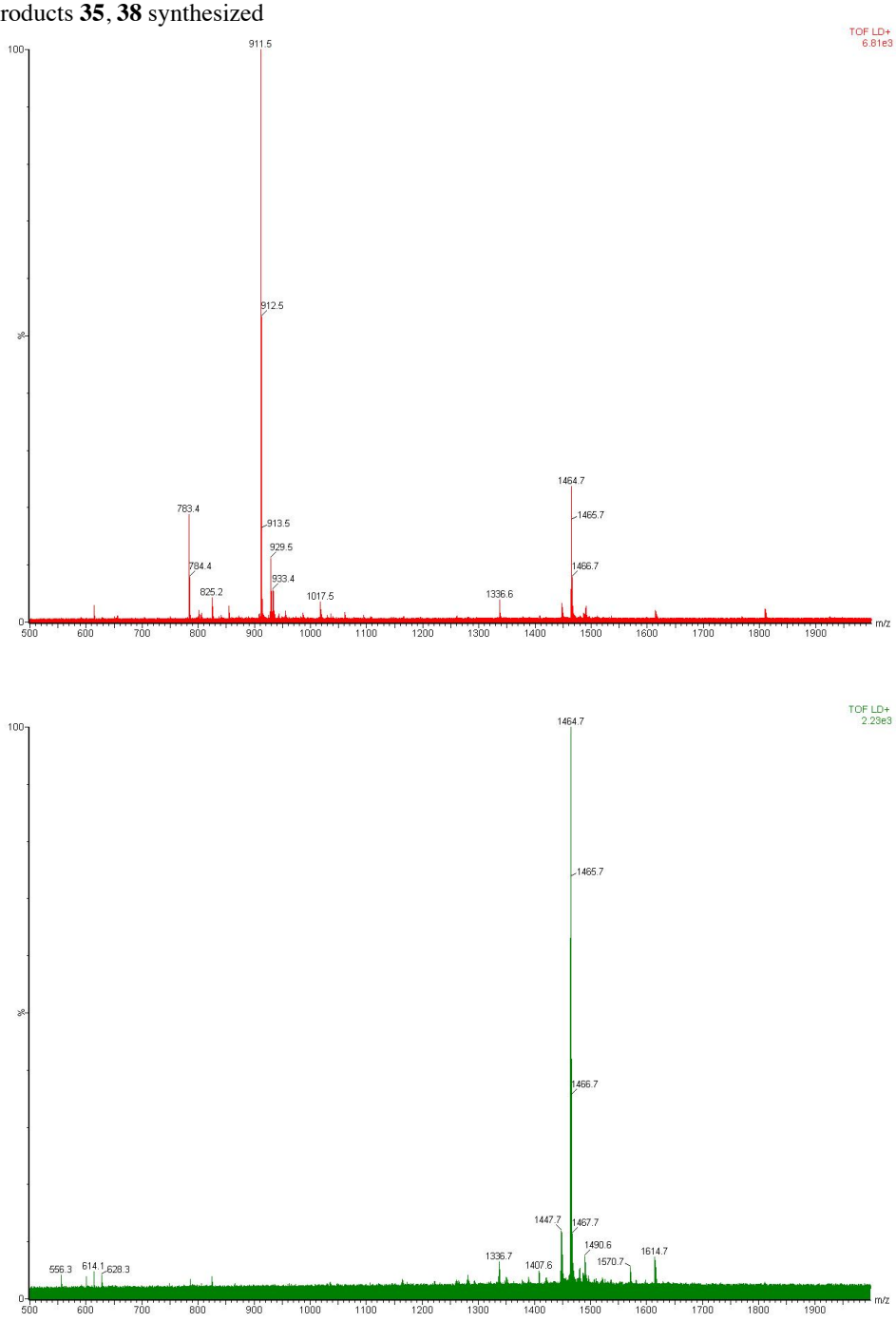
(k) Reaction of **15** QGGRGDW**p**AYDGE with PatG (top) and PatG S783A (bottom)



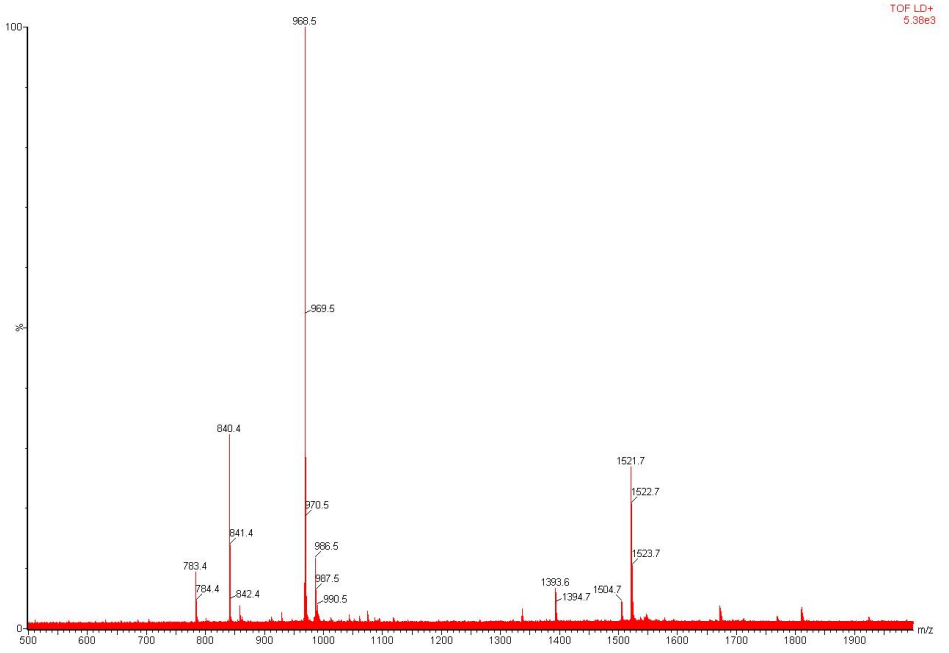
(l) Reaction of **16** QGGWPAYDGE with PatG (top) and PatG S783A (bottom)

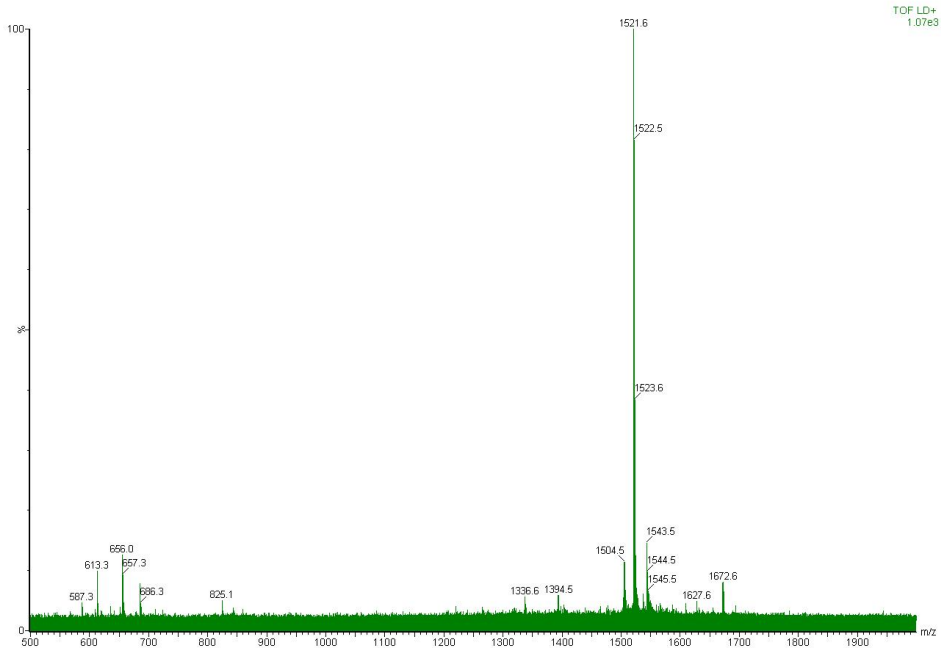


(m) Reaction of **17** Q**G**GGRGDWPAYDGE with PatG (top) and PatG S783A (bottom); products **35**, **38** synthesized

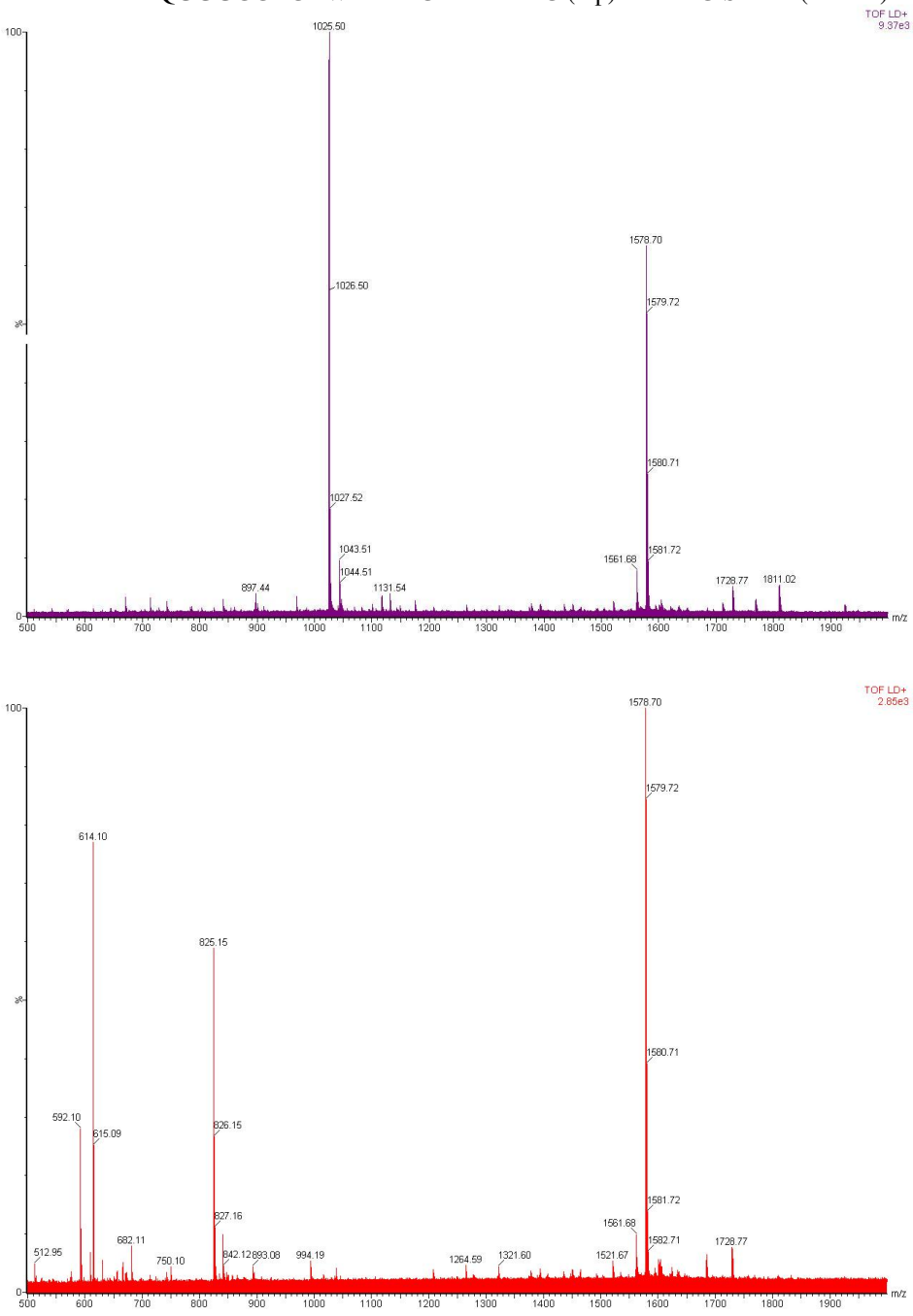


(n) Reaction of **18** Q**GG**GGRGDWPAYDGE with PatG (top) and PatG S783A (bottom); products **36**, **38**, **39** synthesized

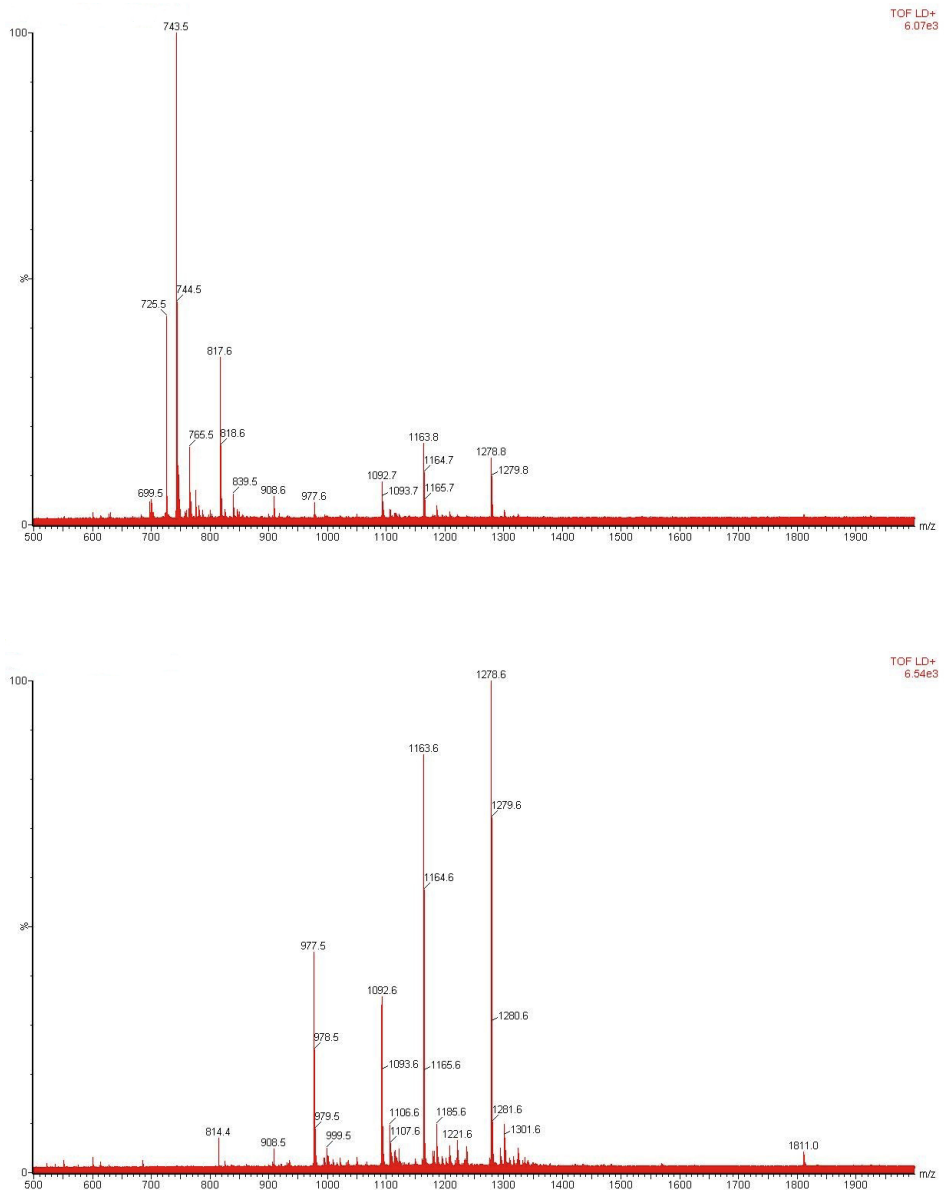




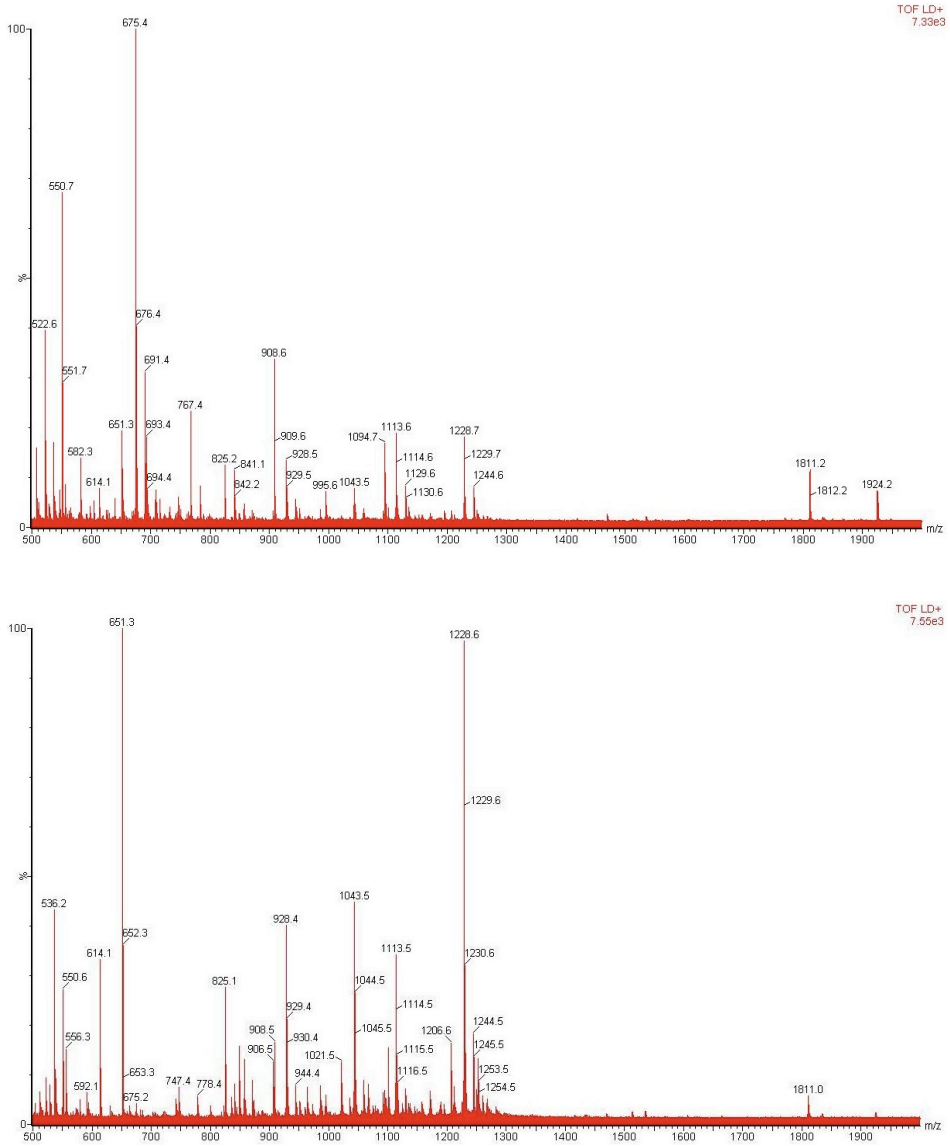
(o) Reaction of **19** Q**GGG**GGRGDWPAYDGE with PatG (top) and PatG S783A (bottom)



(p) Reaction of **22** (**Ahx**)GGRGDWPAYDGE with PatG (top) and PatG S783A (bottom)



(q) Reaction of **23** QGG(**Ahp**)WPAYDGE with PatG (top) and PatG S783A (bottom)



(r) Reaction of **25** **K**KPYILPAYDGE with PatG (top) and PatG S783A (bottom); products **42**, **43** synthesized

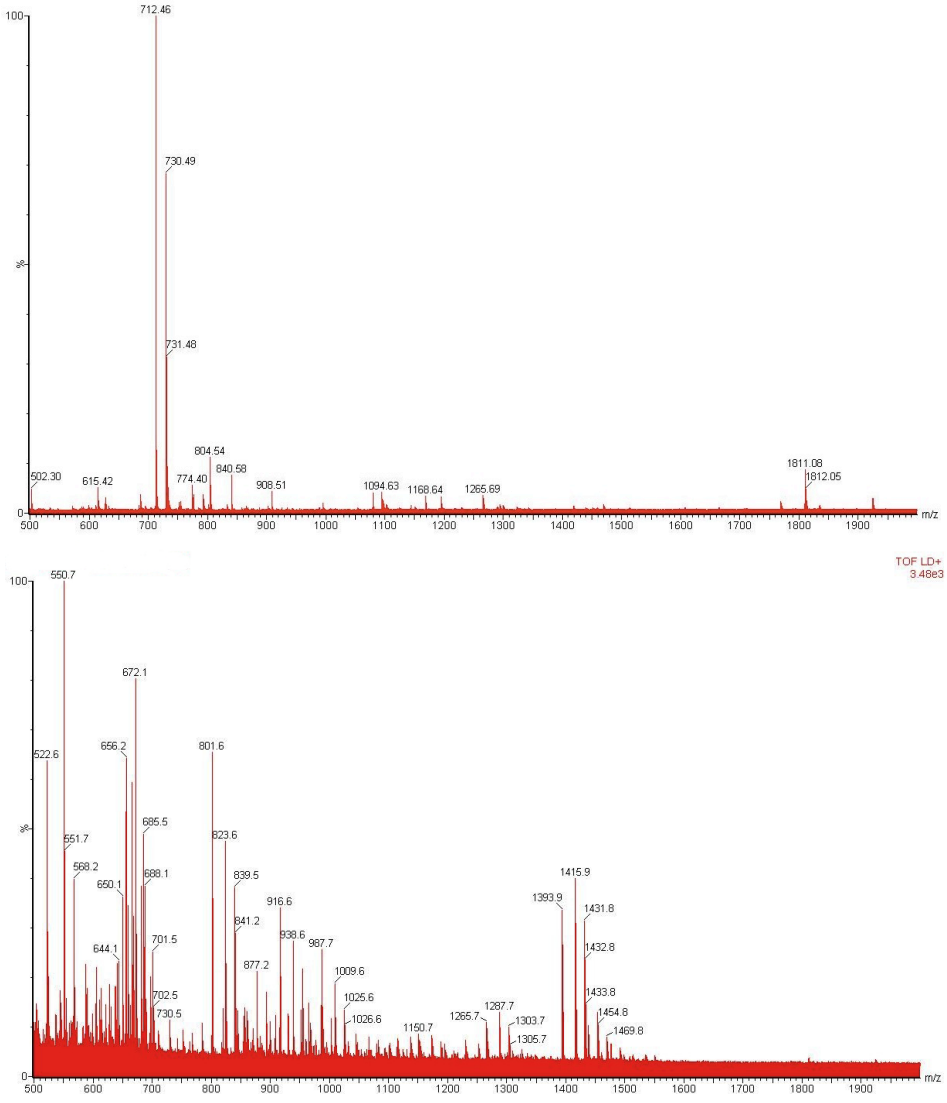
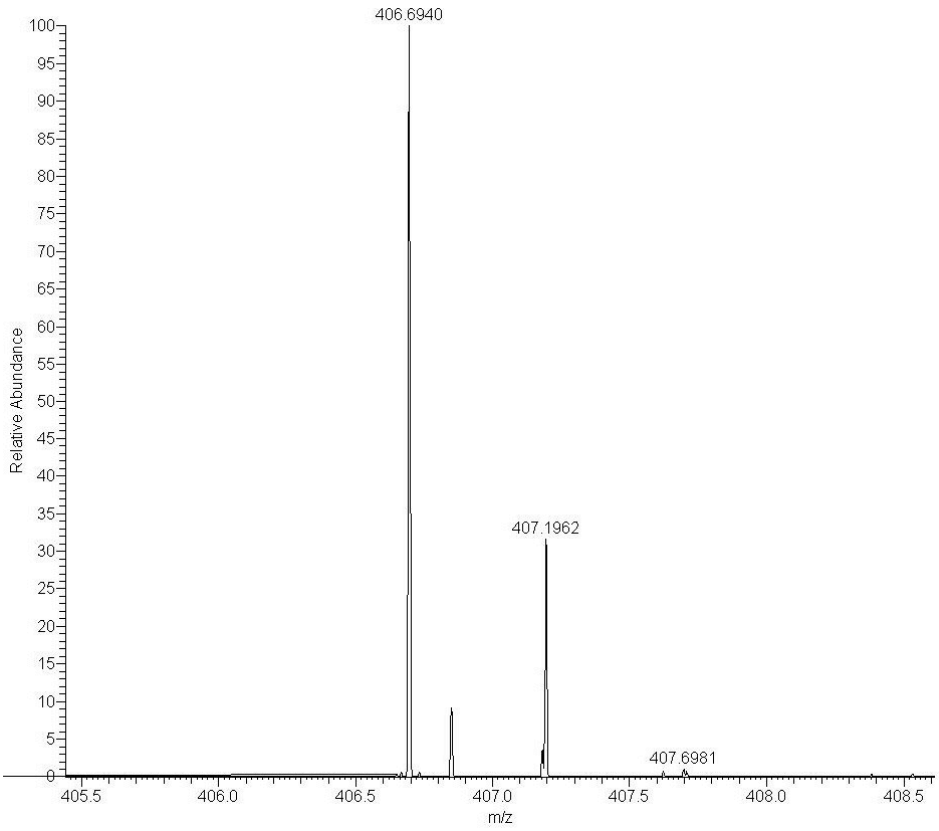
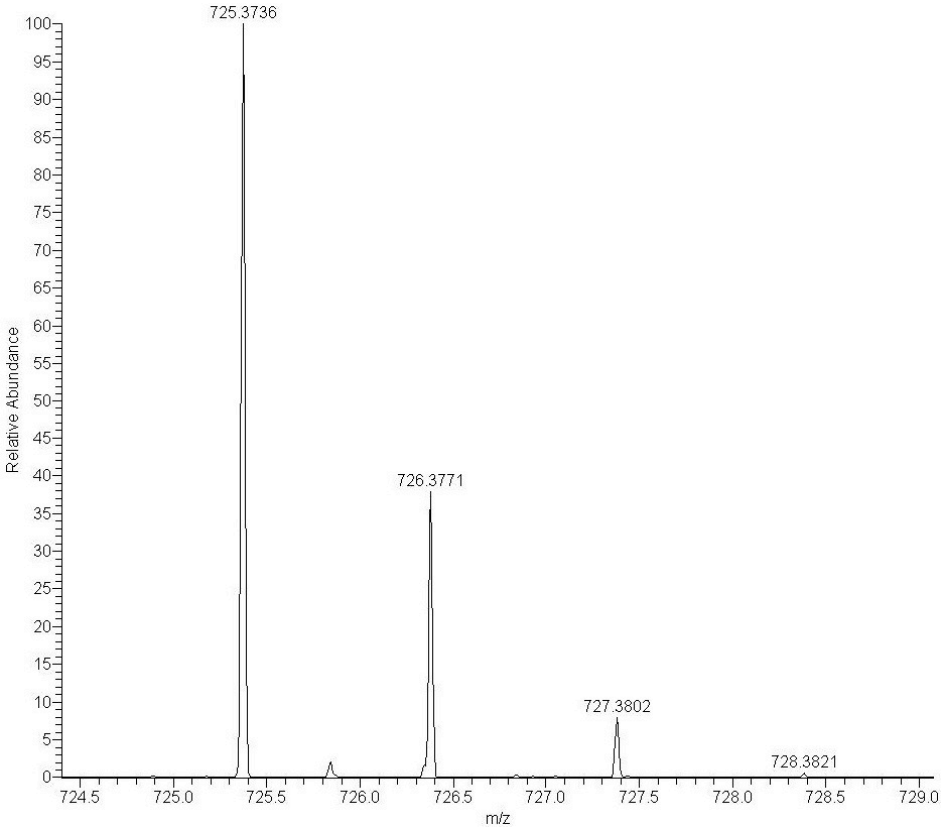


Figure S2. Shown below are LC-FT-ICR data for (a) **27** (b) **40** (c) **41** (d) **42** (e) **43**

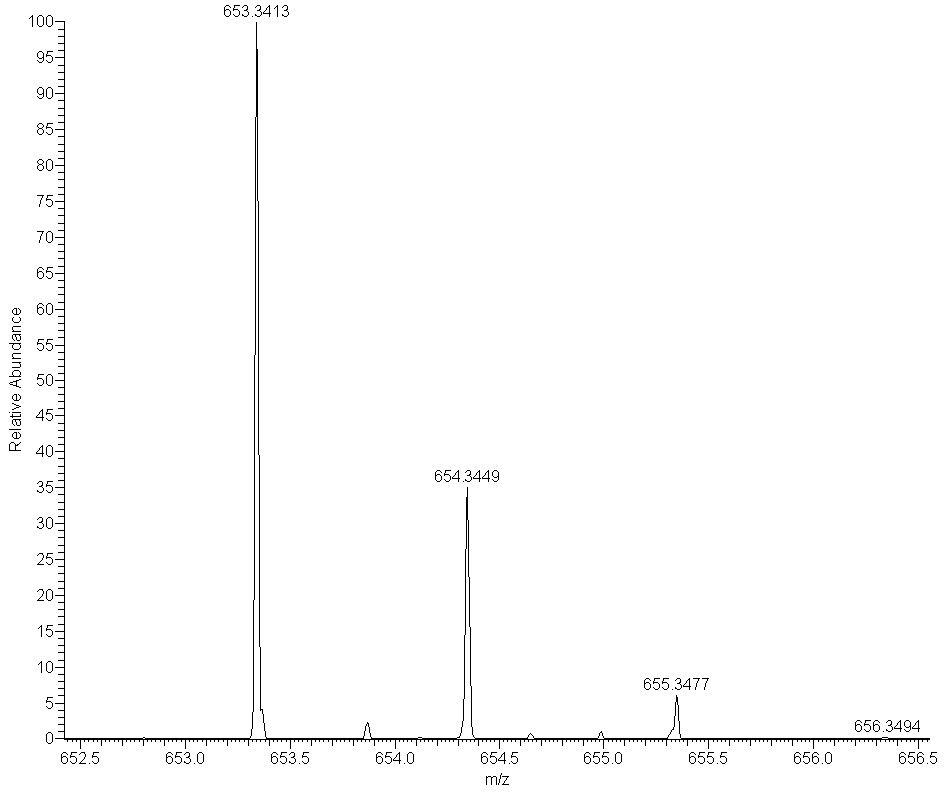
(a)



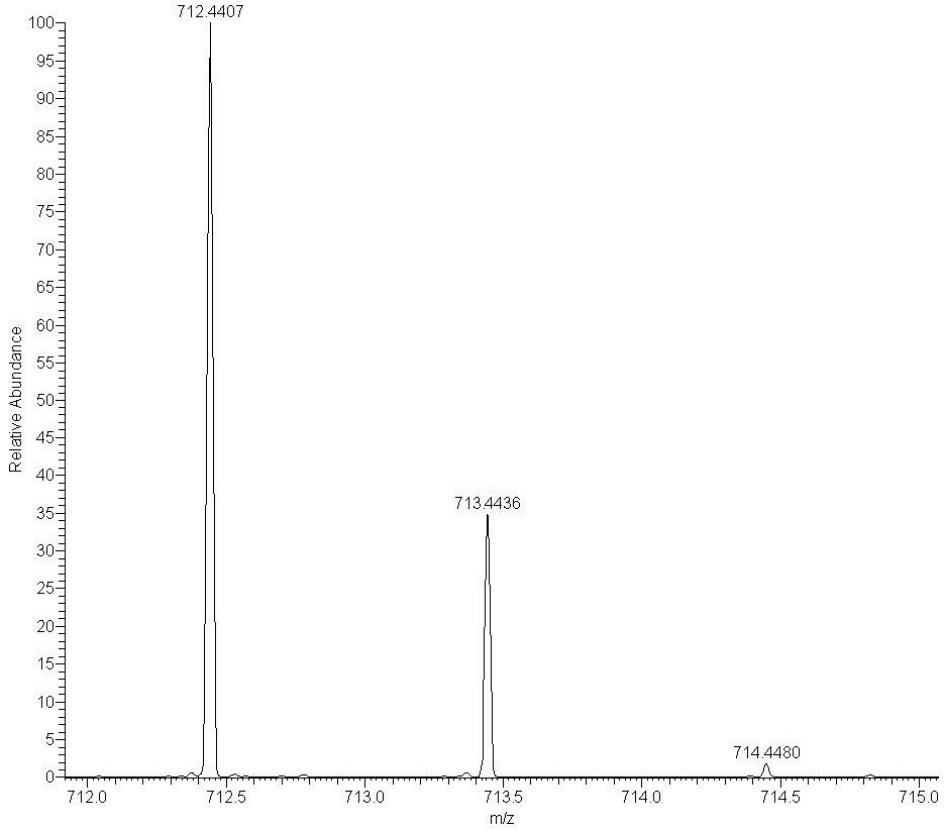
(b)


725.3736
726.3771
727.3802
728.3821
Relative Abundance
m/z
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
724.5
725.0
725.5
726.0
726.5
727.0
727.5
728.0
728.5
729.0

(c)


653.3413
654.3449
655.3477
656.3494
Relative Abundance
m/z
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
652.5
653.0
653.5
654.0
654.5
655.0
655.5
656.0
656.5

(d)


712.4407
713.4436
714.4480
Relative Abundance
m/z
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
712.0
712.5
713.0
713.5
714.0
714.5
715.0

(e)

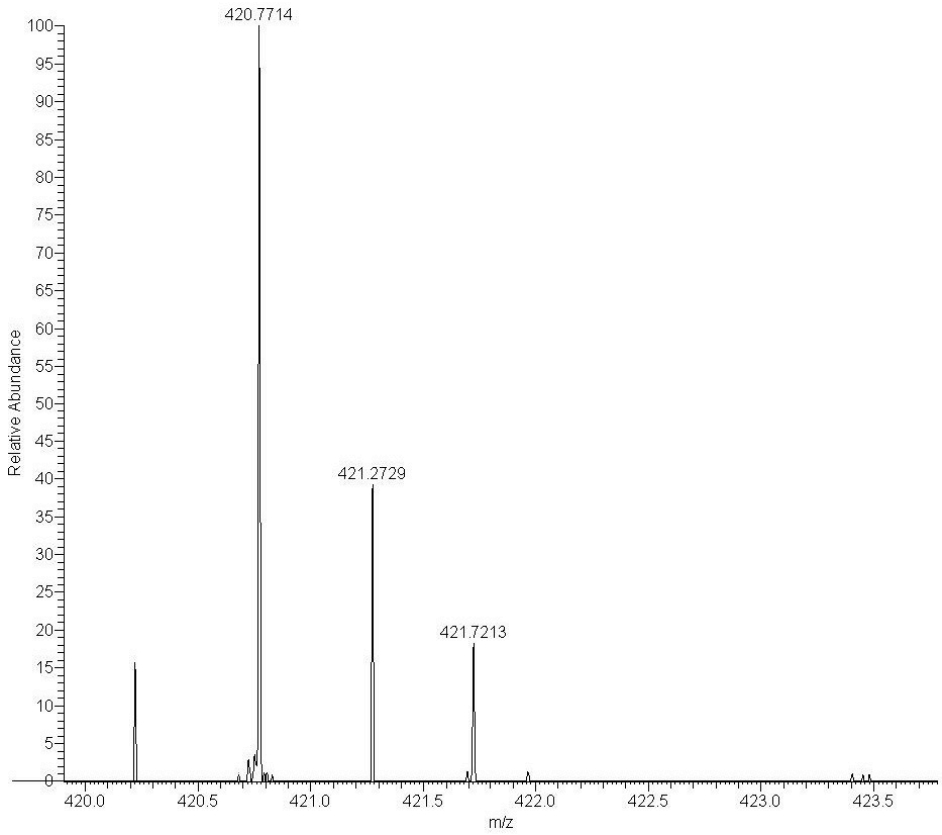
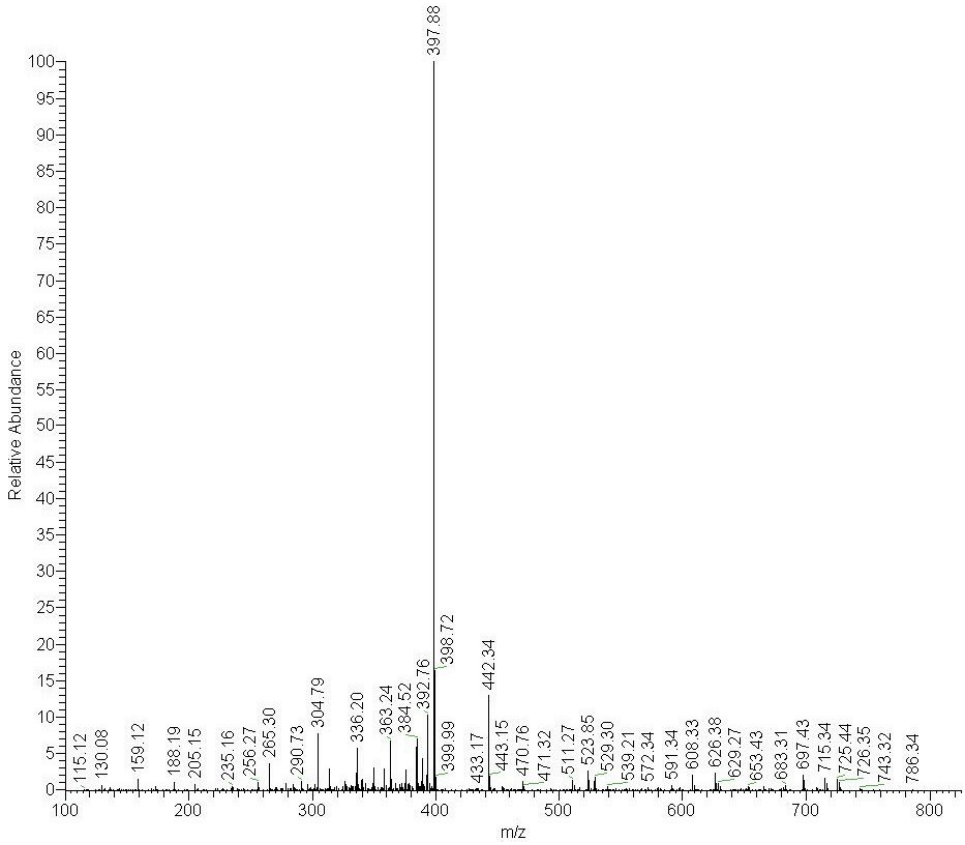
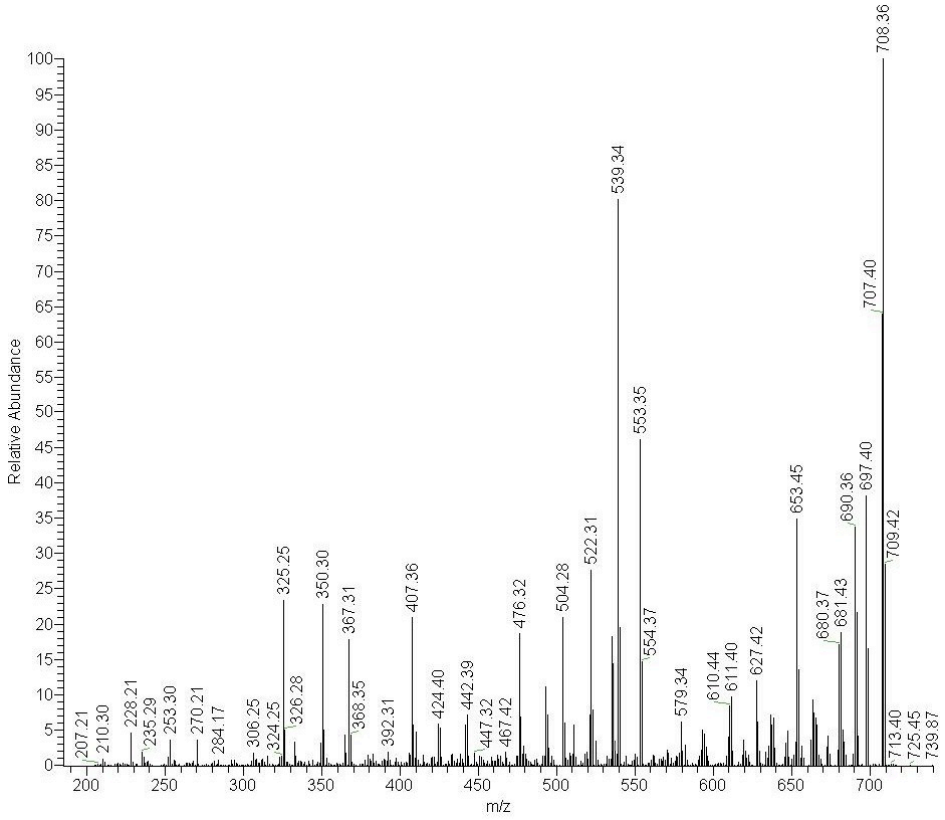

420.7714
421.2729
421.7213
Relative Abundance
m/z
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
420.0
420.5
421.0
421.5
422.0
422.5
423.0
423.5

Figure S3. Shown below are MS-MS data collected in tandem with LC-FT-ICR data for (a) **27** (b) **40** (c) **41** (d) **42** (e) **43** Tables assigning the major peaks of each spectrum are also provided following each spectrum.



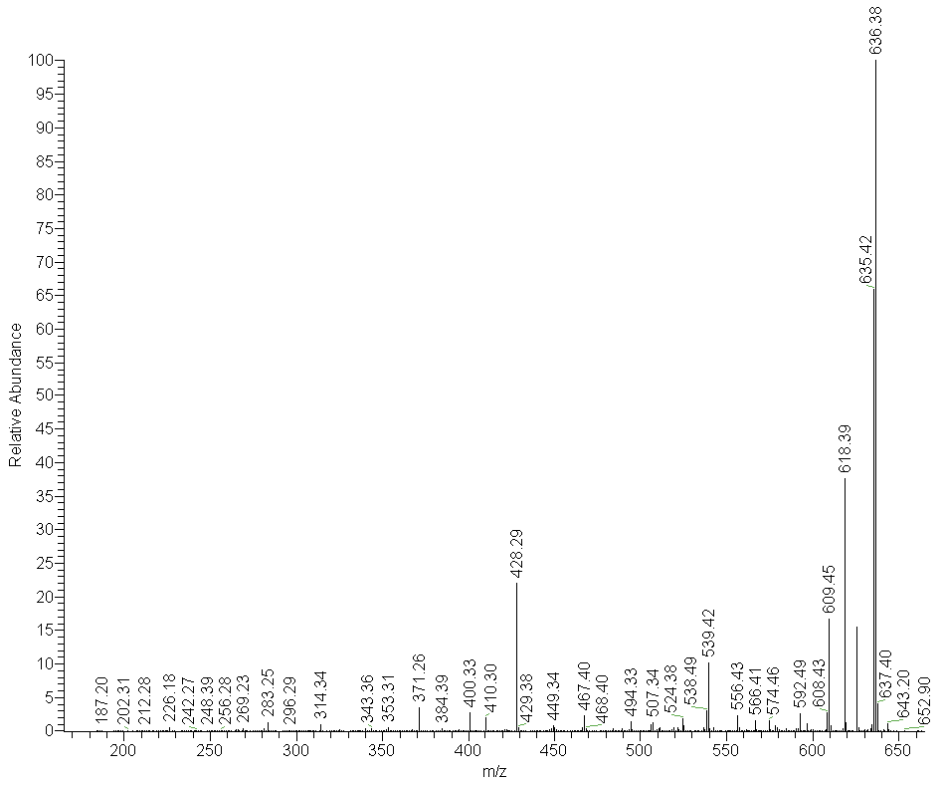
y-, b- cleavages	Expected	Observed	y-, b- cleavages	Expected	Observed
(Dap)G	144.08	144.21	GD	173.06	173.04
(Dap)GG	201.10	-	GDW	329.14	329.21
(Dap)GGR	357.20	-	GDWP	456.19	-
(Dap)GGRG	414.22	-	GDWP(Dap)	542.24	-
(Dap)GGRGD	529.25	529.30	GDWP(Dap)G	599.26	-
(Dap)GGRGDW	715.33	715.33	GDWP(Dap)GG	656.28	-
GG	115.05	115.12	DW	302.12	302.11
GGR	271.15	-	DWP	399.17	-
GGRG	328.17	328.16	DWP(Dap)	485.22	485.36
GGRGD	443.20	-	DWP(Dap)G	542.24	-
GGRGDW	629.28	629.27	DWP(Dap)GG	599.26	-
GGRGDWP	726.33	-	DWP(Dap)GGR	755.36	-
GR	214.13	214.15	WP	284.14	284.20
GRG	271.15	-	WP(Dap)	370.19	-
GRGD	386.18	386.18	WP(Dap)G	427.21	427.23
GRGDW	572.26	572.34	WP(Dap)GG	484.23	-
GRGDWP	669.31	-	WP(Dap)GGR	640.33	640.38
GRGDWP(Dap)	755.36		WP(Dap)GGRG	697.36	697.43
RG	214.13	214.15	P(Dap)	184.11	184.27
RGD	329.16	329.21	P(Dap)G	241.13	241.15
RGDW	515.24	-	P(Dap)GG	298.15	-
RGDWP	612.29	-	P(Dap)GGR	454.25	-
RGDWP(Dap)	698.34	-	P(Dap)GGRG	511.28	511.29
RGDWP(Dap)G	755.36	-	P(Dap)GGRGD	626.30	626.38

(b)


Relative Abundance
m/z
207.21
210.30
228.21
235.29
253.30
270.21
284.17
306.25
324.25
325.25
326.28
350.30
367.31
368.35
392.31
407.36
424.40
442.39
447.32
467.42
476.32
504.28
522.31
539.34
553.35
554.37
579.34
610.44
611.40
627.42
653.45
680.37
681.43
690.36
697.40
707.40
708.36
709.42
713.40
725.45
739.87

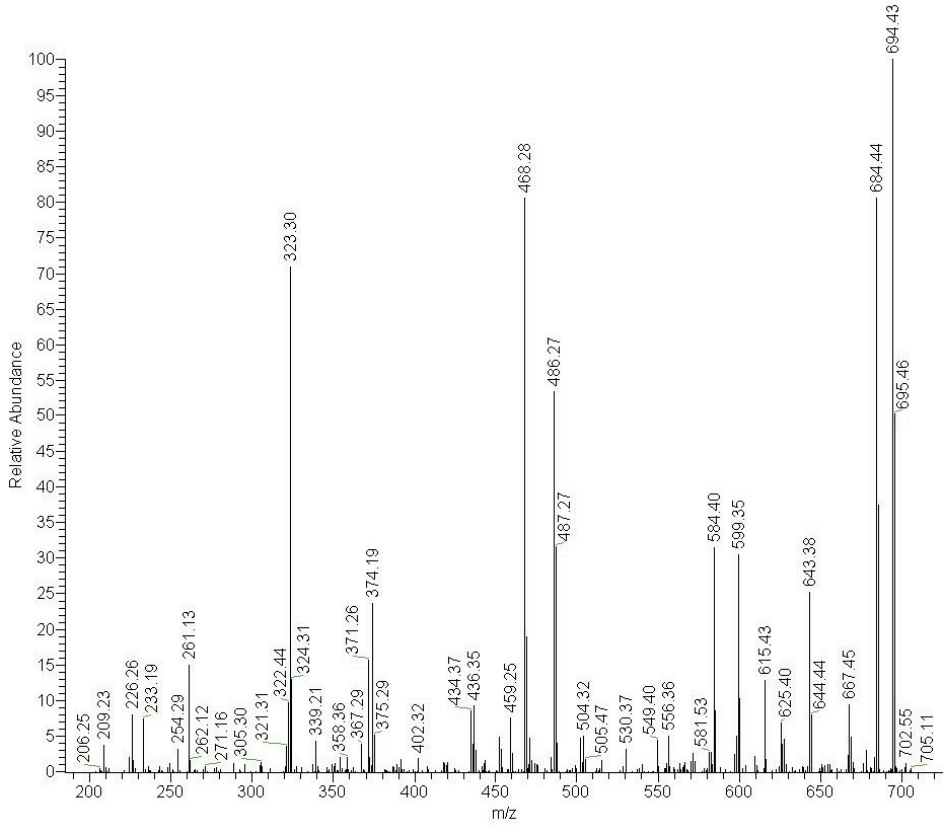
y-, b- cleavages	Expected	Observed	y-, b- cleavages	Expected	Observed
(Ahx)R	270.19	270.21	**P(Ahx)RG***	407.25	407.36
(Ahx)RG	327.22	-	**P(Ahx)RGD**	539.29	539.34
(Ahx)RGD	442.24	-			
(Ahx)RGDW	628.32	-	**(Ahx)RGDWP°**	707.36	707.40
RG	214.13	-	**(Ahx)RGDWP***	708.34	708.36
RGD	329.16	329.21	**(Ahx)RGDWP immonium ion**	697.38	697.4
RGDW	515.24	-			
RGDWP	612.29	-			
GD	173.06	-			
GDW	359.14	359.25			
GDWP	456.19	456.31			
GDWP(Ahx)	569.27	569.33			
DW	302.11	302.15			
DWP	399.17	-			
DWP(Ahx)	512.25	-			
DWP(Ahx)R	668.35	668.38			
WP	284.14	284.17			
WP(Ahx)	397.22	397.30			
WP(Ahx)R	553.33	553.35			
WP(Ahx)RG	610.35	610.44			
P(Ahx)	211.15	-			
P(Ahx)R	367.25	367.31			
P(Ahx)R*	350.21	350.30			
P(Ahx)RG	424.27	424.40			

(c)


Relative Abundance
m/z
636.38
635.42
618.39
609.45
428.29
539.42
637.40
643.20
652.90
187.20
202.31
212.28
226.18
242.27
248.39
256.28
269.23
283.25
296.29
314.34
343.36
353.31
371.26
384.39
400.33
410.30
429.38
449.34
467.40
468.40
494.33
507.34
524.38
538.49
556.43
566.41
574.46
592.49
608.43

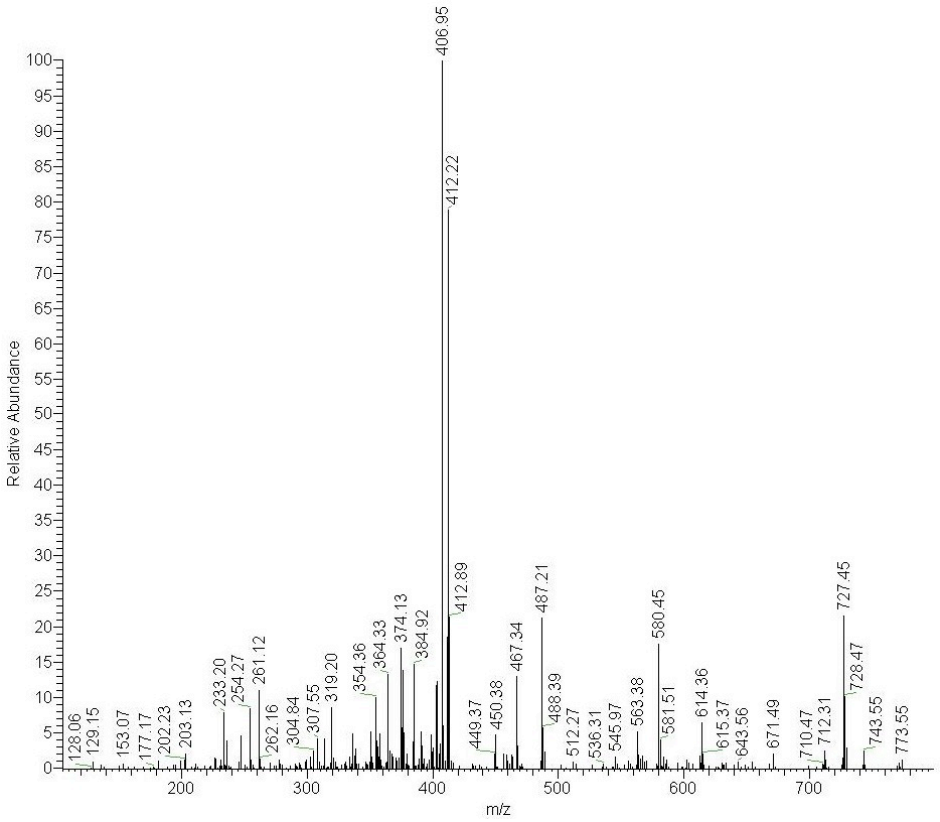
y-, b- cleavages	Expected	Observed	y-, b- cleavages	Expected	Observed
QG	186.09	-	**PQGG**	340.16	340.21
QGG	243.11	243.18	**PQGG(Ahp)**	467.26	467.40
QGG(Ahp)	370.21	-			
QGG(Ahp)*	353.18	353.31	**y-, a- cleavages**		
QGG(Ahp)W	556.29	556.43	**QGG(Ahp)W**	528.29	528.4
GG	115.05	-	**GG(Ahp)**	214.16	214.17
GG(Ahp)	242.15	242.27	**GG(Ahp)W**	400.23	400.33
GG(Ahp)W	428.23	428.23	**(Ahp)W**	286.19	286.21
GG(Ahp)WP	525.28	-	**WP**	256.14	256.28
G(Ahp)	185.13	185.07			
G(Ahp)W	371.21	371.26	**Misc. Ions**		
G(Ahp)WP	468.26	-	**QGG(Ahp)WP***	636.31	636.38
G(Ahp)WPQ	596.32	596.51	**QGG(Ahp)WP°**	635.33	635.42
(Ahp)W	314.19	314.34	**QGG(Ahp)WP (-C=O)**	625.35	625.46
(Ahp)WP	411.24	-	**QGG(Ahp)WP (-NH3, -H2O)**	618.30	618.39
(Ahp)WPQ	539.30	539.42			
(Ahp)WPQG	596.32	596.51			
WP	284.14	-			
WPQ	412.20	-			
WPQG	469.22	-			
WPQGG	526.24	-			
WPQGG	526.24	-			
PQ	226.12	226.18			
PQG	283.14	283.25			

(d)


Relative Abundance
m/z
206.25
209.23
226.26
233.19
254.29
261.13
262.12
271.16
305.30
321.31
322.44
323.30
324.31
339.21
358.36
367.29
371.26
374.19
375.29
402.32
434.37
436.35
459.25
468.28
486.27
487.27
504.32
505.47
530.37
549.40
556.36
581.53
584.40
599.35
615.43
625.40
643.38
644.44
667.45
684.44
694.43
695.46
702.55
705.11

y-, b- cleavages	Expected	Observed	y-, b- cleavages	Expected	Observed
KP	226.16	226.26	**PKPY°**	468.26	468.28
KPY	389.22	389.24	**PKYPI**	599.36	599.35
KPY°	371.21	371.26	**KPYILP°**	694.43	694.43
KPYI	502.30	502.39	**KPYILP (- sidechain C=O, i.e. immonium ion)**	684.44	684.44
KPYIL	615.39	615.43	**KPYIPL (immonium ion - NH3)**	667.42	667.45
PY	261.12	261.13			
PYI	374.21	374.19	**y-, a- cleavages**		
PYIL	487.29	-	**PYILP**	556.35	556.36
PYILP	584.34	584.40	**PYIL**	459.30	459.25
YI	277.16	-			
YIL	390.24	-			
YILP	487.29	-			
YILPK	615.39	615.43			
IL	227.18	-			
ILP	324.23	-			
ILPK	452.32	452.39			
ILPKP	549.38	549.40			
LP	211.14	-			
LPK	339.24	339.21			
LPKP	436.29	436.35			
LPKPY	599.36	599.35			
PK	226.16	-			
PKP	323.21	323.30			
PKPY	486.27	486.27			

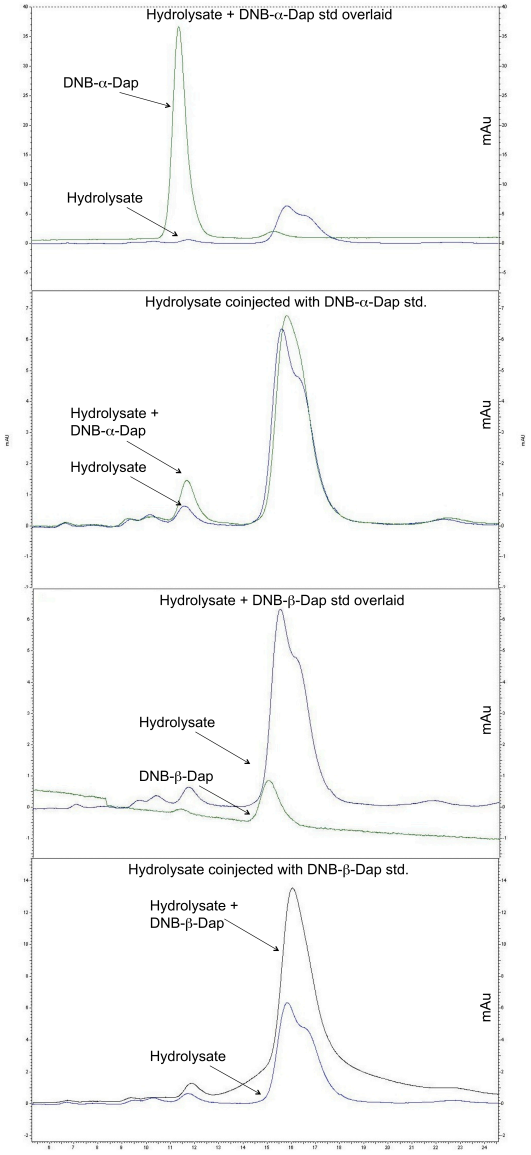
(e)


Relative Abundance
m/z
406.95
412.22
412.89
487.21
727.45
374.13
384.92
467.34
580.45
728.47
233.20
254.27
261.12
364.33
354.36
319.20
614.36
128.06
129.15
153.07
177.17
202.23
203.13
262.16
304.84
307.55
449.37
450.38
488.39
512.27
536.31
545.97
563.38
581.51
615.37
643.56
671.49
710.47
712.31
743.55
773.55

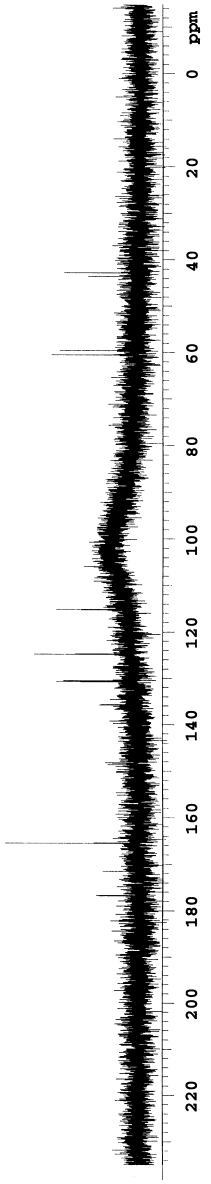
y-, b- cleavages	Expected	Observed	y-, b- cleavages	Expected	Observed
KP	226.16	-	**LPKKPY**	727.45	727.45
KPY	389.22	-	**PK**	226.16	-
KPYI	502.30	-	**PKK**	354.25	354.36
KPYIL	615.39	-	**PKKP**	451.30	
KPYILP	712.44	712.31	**PKKPY**	614.37	614.36
PY	261.12	261.12	**PKKPYI**	727.45	727.45
PYI	374.21	374.13	**KK**	257.20	-
PYIL	487.29	487.21	**KK (+2)**	129.11	129.15
PYILP	584.34	584.32	**KKP**	354.25	354.36
PYILPK	712.44	712.31	**KKPY**	517.31	-
YI	277.16	277.23	**KKPYI**	630.40	-
YIL	390.24	-	**KKPYIL**	743.38	743.55
YILP	487.29	487.21			
YILPK	615.39	-			
YILPKK	743.48	743.55			
IL	227.18	227.07			
ILP	324.23	324.23			
ILPK	452.32	-			
ILPKK	549.38	-			
ILPKKP	677.47	-			
LP	211.14	211.07			
LPK	339.24	-			
LPKK	467.33	467.35			
LPKKP	564.39	-			

Figure S4. (a) HPLC profile of DNFB-reacted hydrolysate of **27** compared with authentic standards of α- and β-DNP Dap (b) ^{13}C, 1H and LC-MS spectra for α-DNP Dap (c) ^{13}C, ^{1}H and LC-MS spectra for β-DNP Dap. (d) LC-MS profile of DNFB-reacted hydrolysate of **28** compared with authentic standards of α- and δ-DNP Orn (e) ^{13}C, ^{1}H, and LC-MS spectra for α-DNP Orn (f) ^{13}C, ^{1}H, and LC-MS spectra for δ-DNP Orn (g) LC-MS profile of DNFB-reacted hydrolysates of **29** compared with authentic standards of α- and ε-DNP Lys. (h) HPLC profile of DNFB reacted hydrolysates of **42** compared with authentic standards of α- and ε-DNP Lys. (i) UHPLC-MS analysis confirming HPLC analysis in 4(h) above (j) ^{13}C, ^{1}H, and MALDI-MS spectra for α-DNP Lys. (k) ^{13}C, ^{1}H and MALDI-MS spectra for ε-DNP Lys.

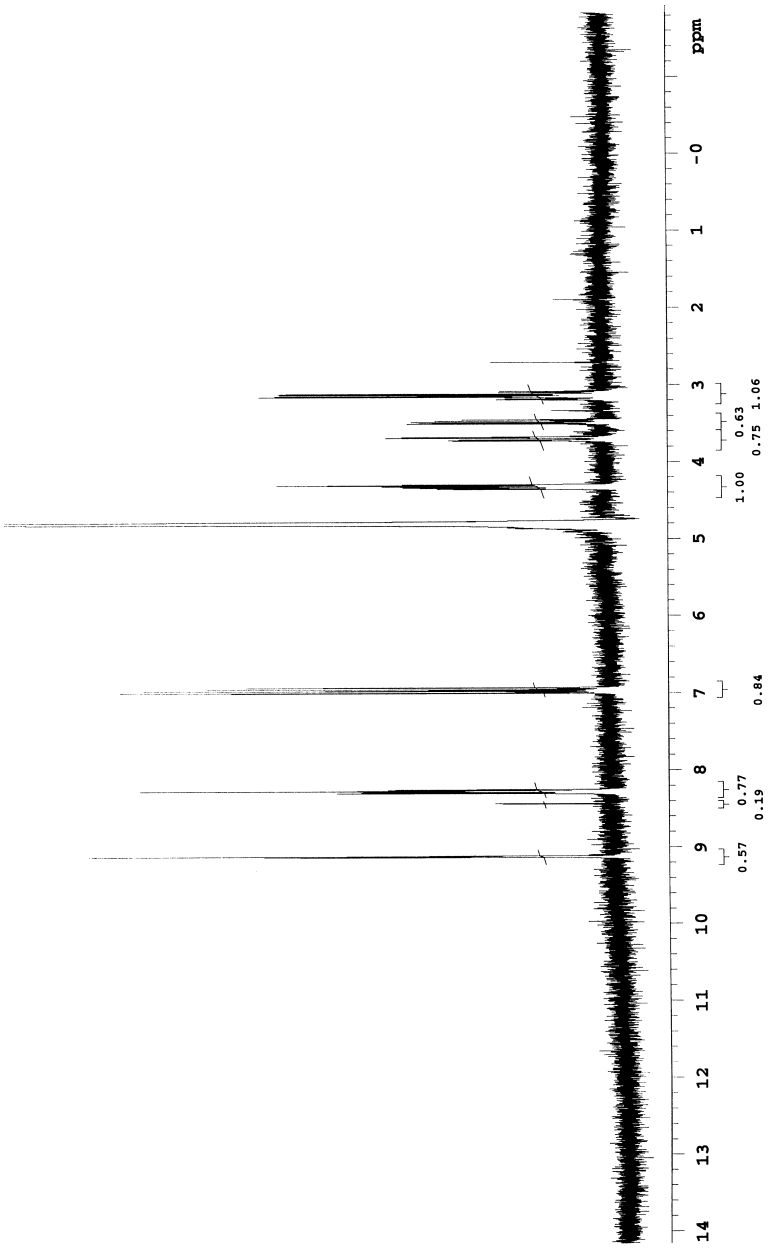
(a)


Hydrolysate + DNB-α-Dap std overlaid
DNB-α-Dap
Hydrolysate
mAu
Hydrolysate coinjected with DNB-α-Dap std.
Hydrolysate +
DNB-α-Dap
Hydrolysate
mAu
Hydrolysate + DNB-β-Dap std overlaid
Hydrolysate
DNB-β-Dap
mAu
Hydrolysate coinjected with DNB-β-Dap std.
Hydrolysate +
DNB-β-Dap
Hydrolysate
mAu

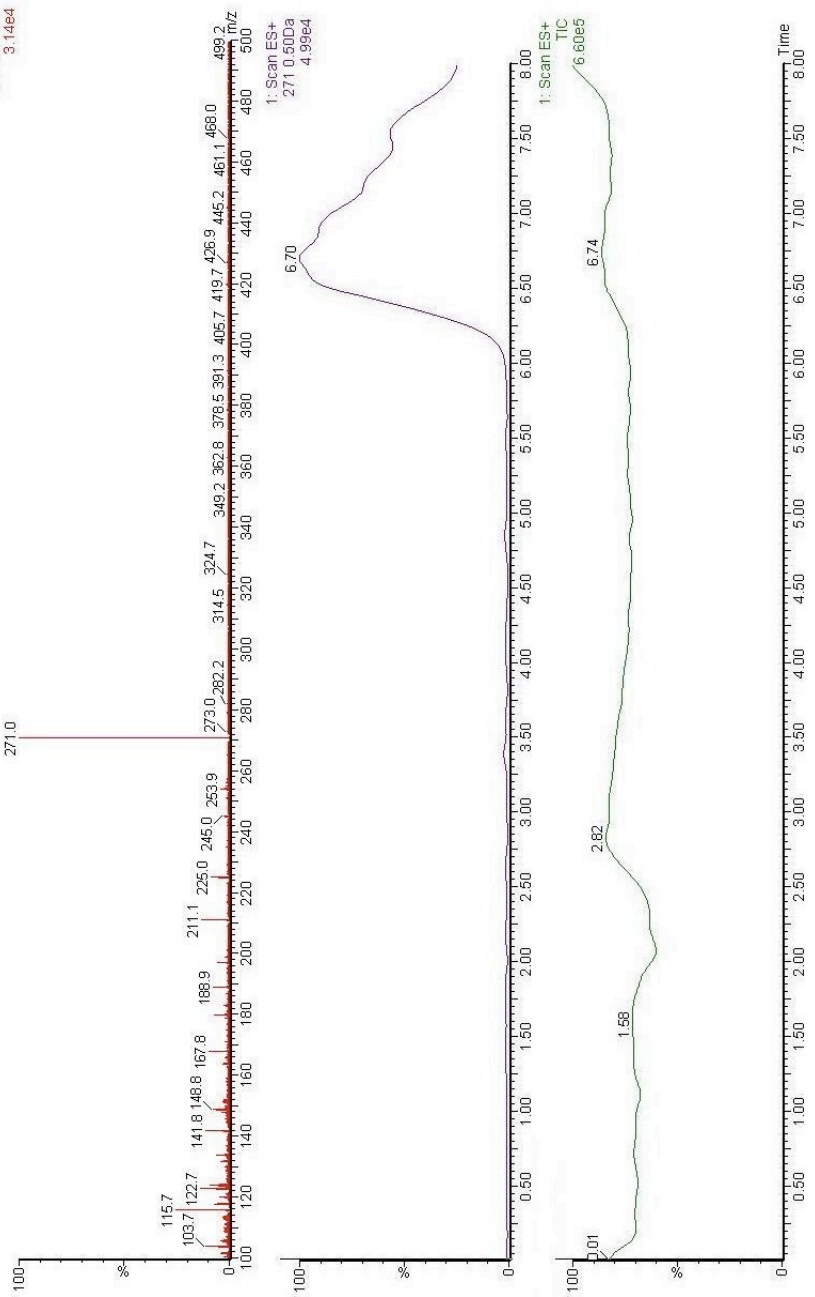
(b) ^{13}C 100 MHz NMR of α-DNP Dap in D_2O with K_2CO_3 (δ=165) added for solubility



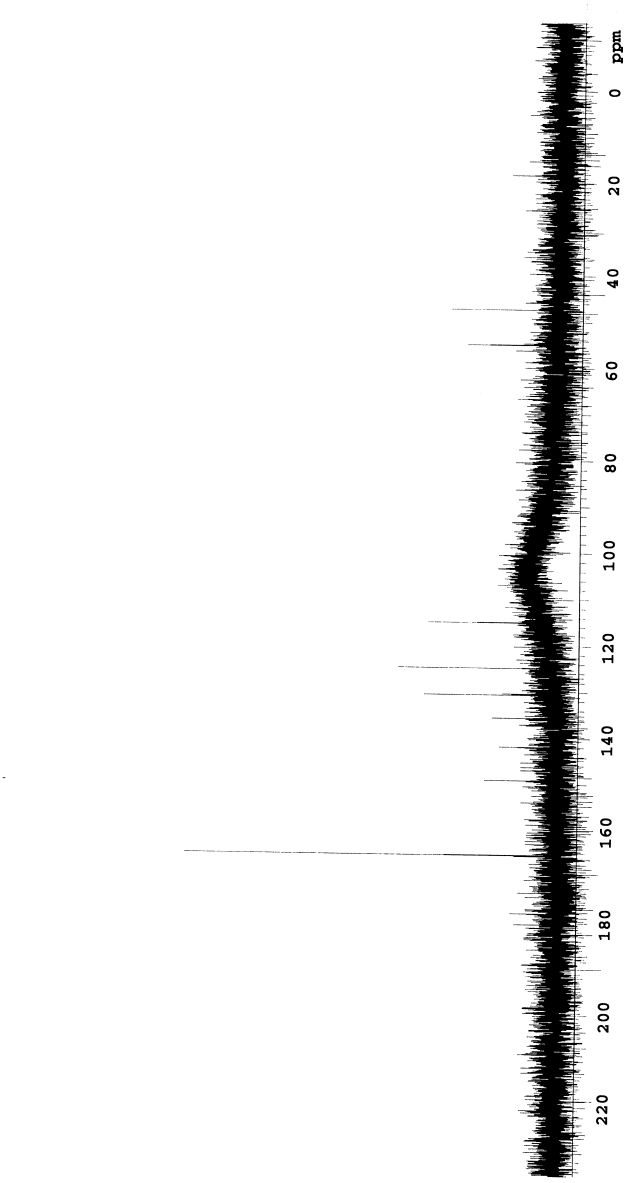
^{1}H 400 MHz NMR of α-DNP Dap.



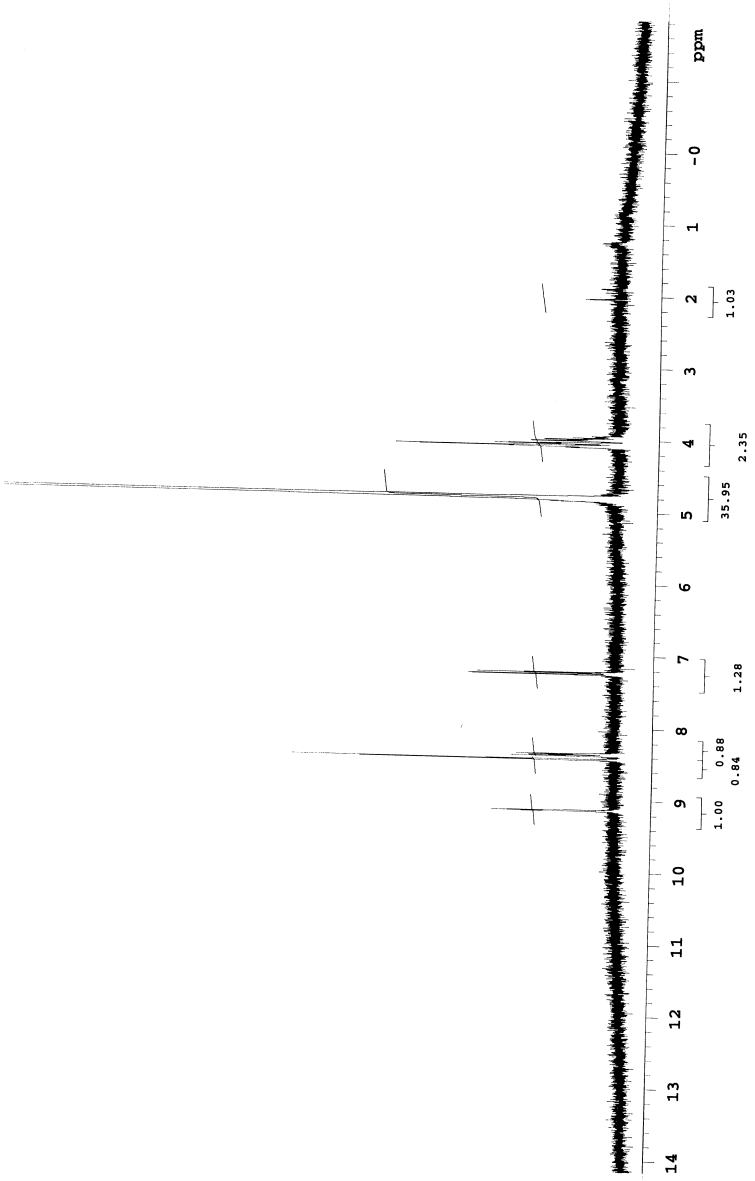
LC-MS of α-DNP Dap



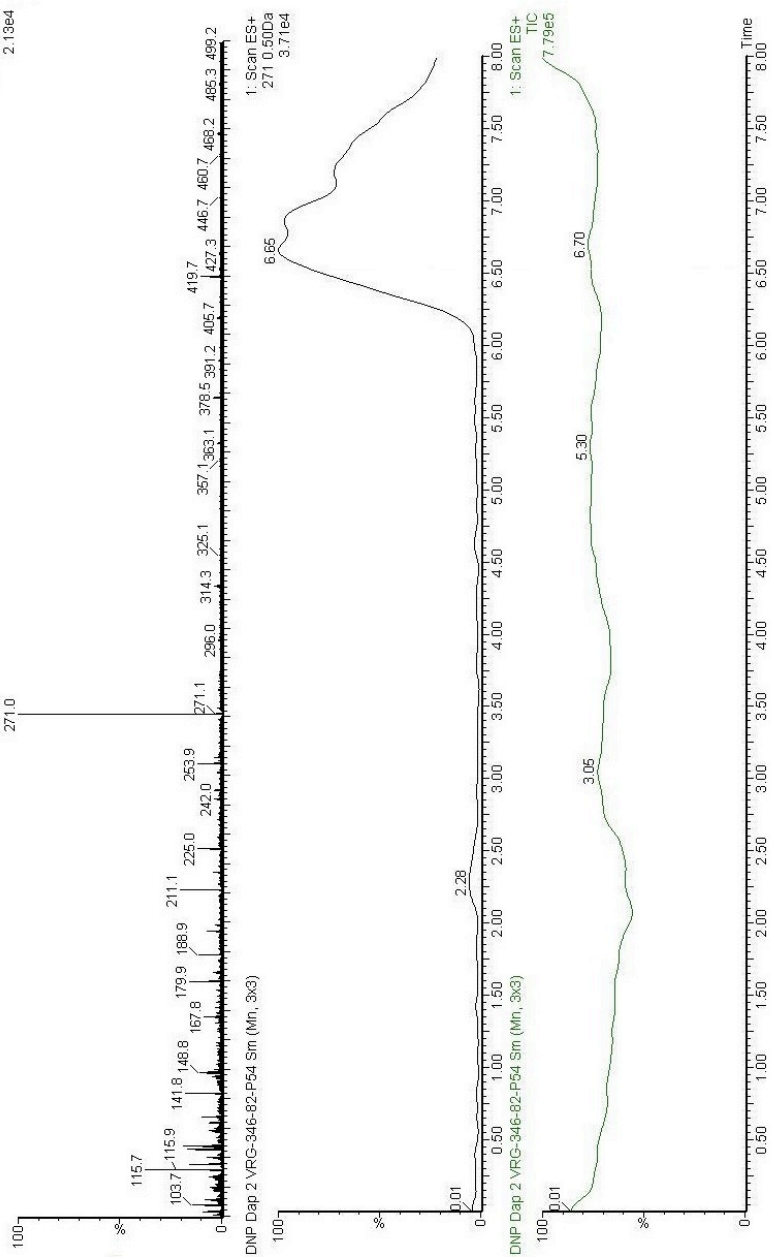
(c) ^{13}C 100 MHz NMR of β-DNP Dap in D_2O with K_2CO_3 (δ=165) added for solubility



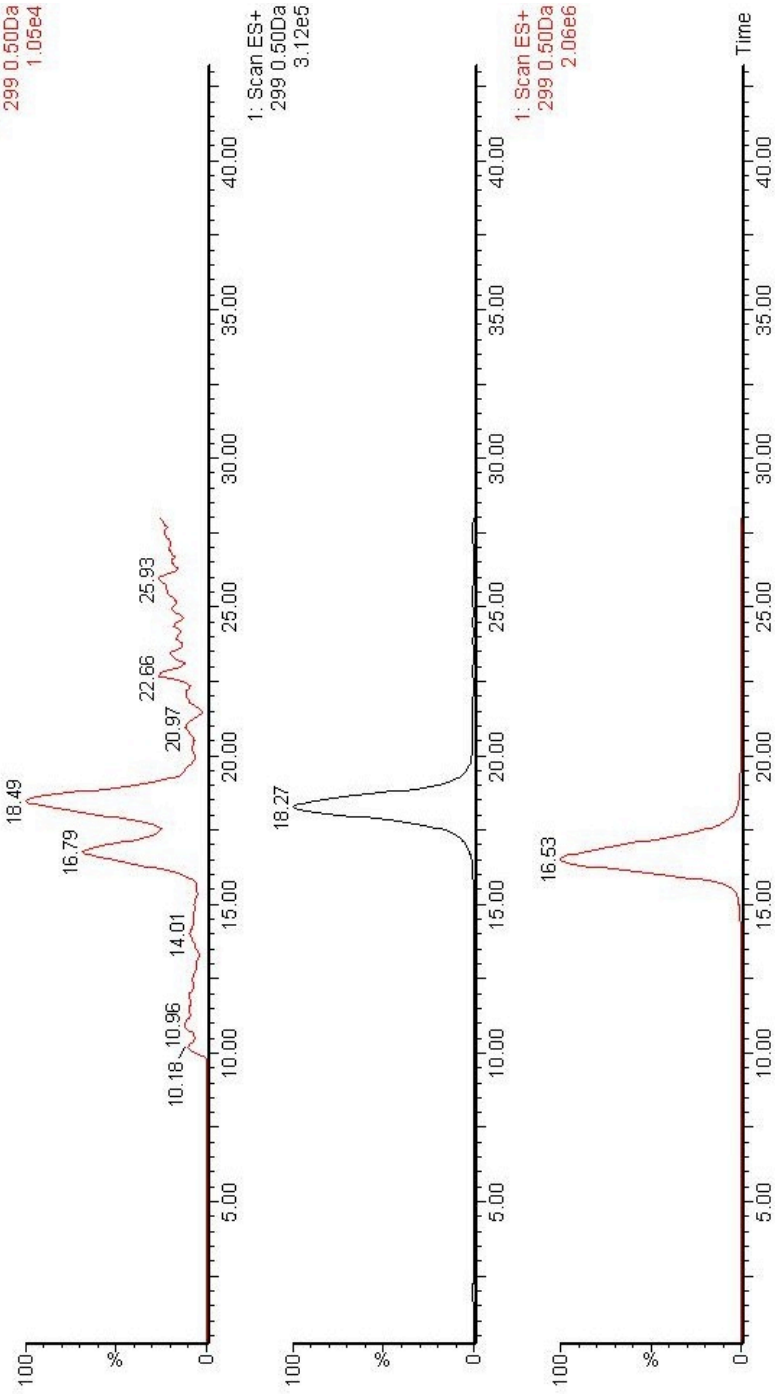
^{1}H 400 MHz NMR β-DNP Dap in D_2O



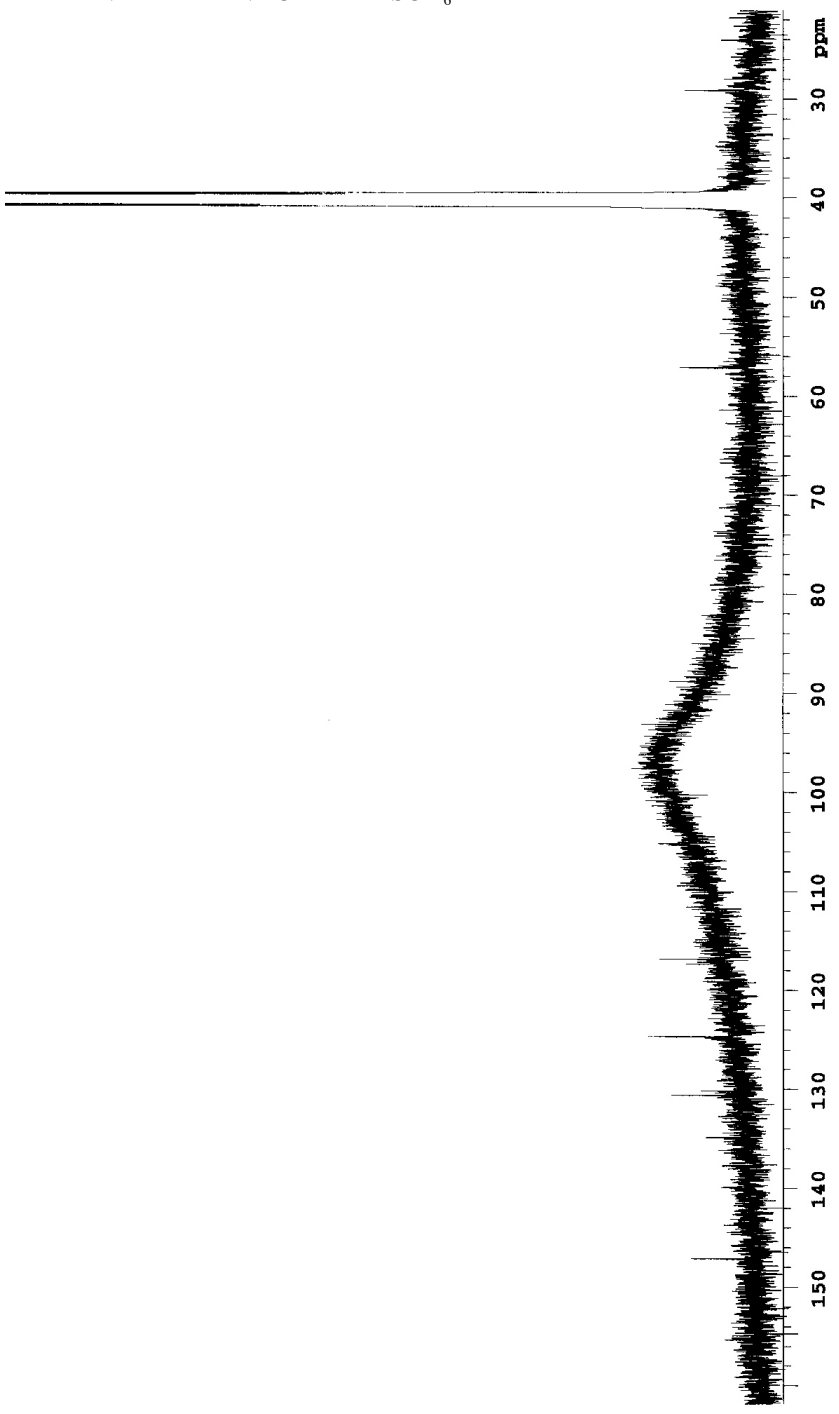
LC-MS of β-DNP Dap



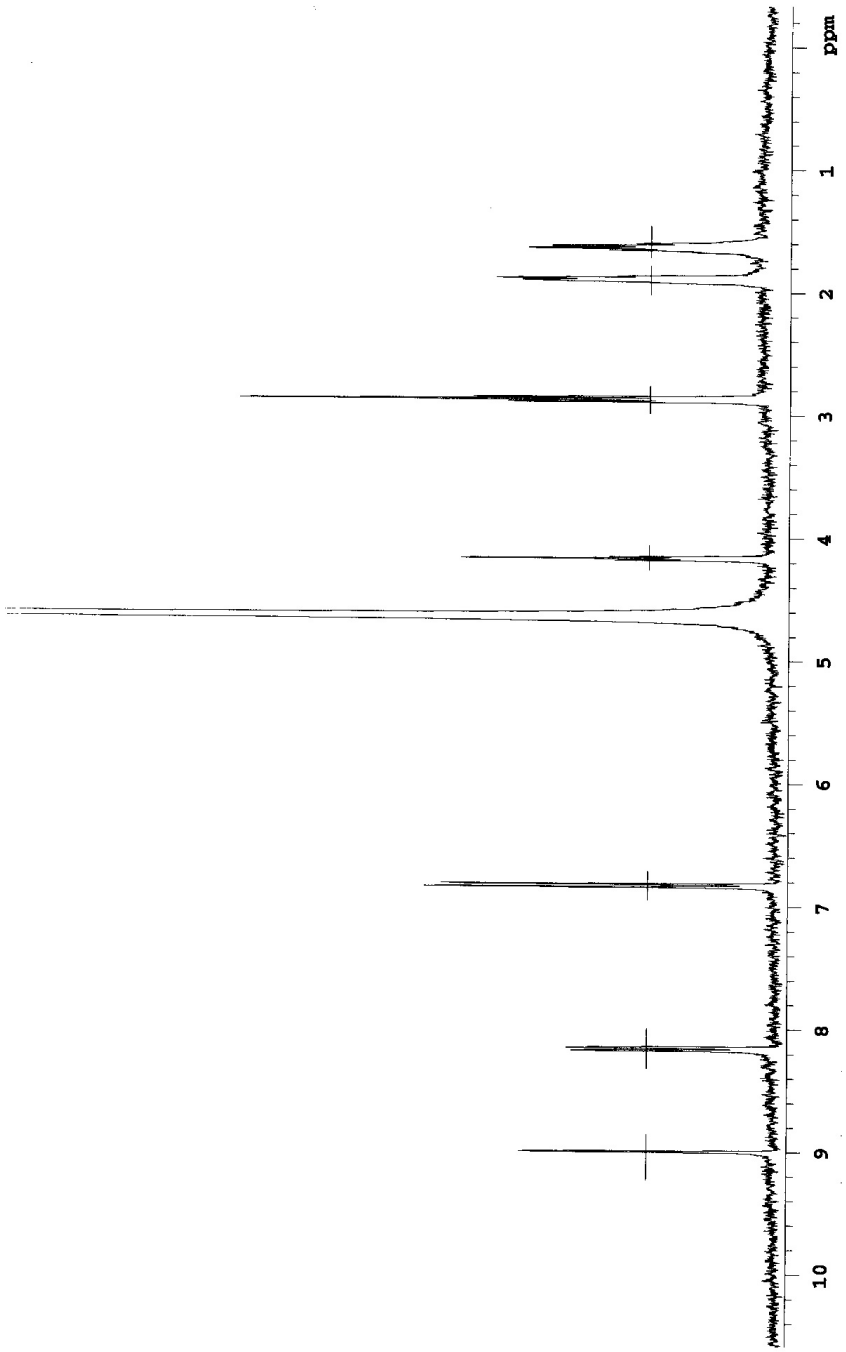
(d) LC-MS chromatograms selected for mass of DNP-Orn; top: hydrolysate, middle: δ-DNP Orn standard; bottom α-DNP Orn standard



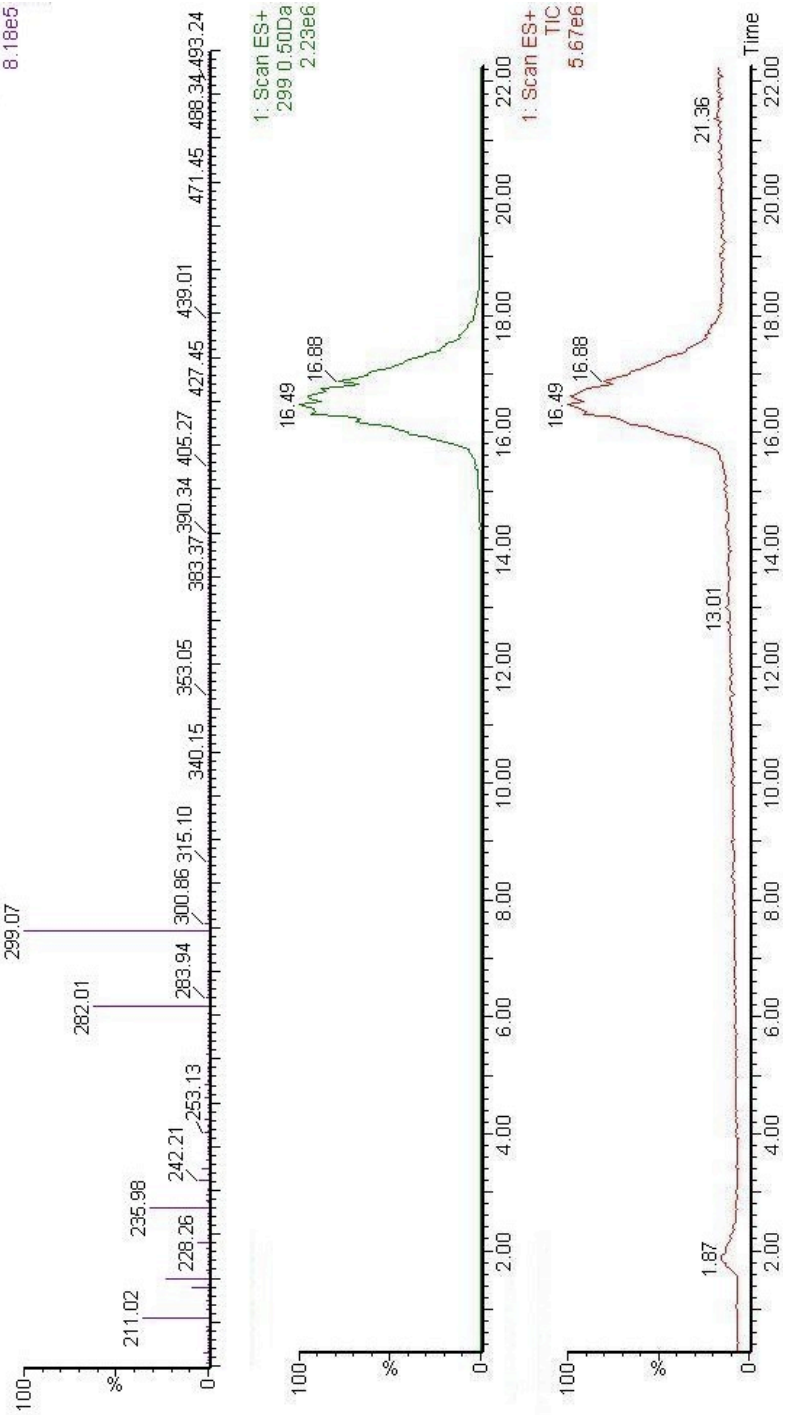
(e) ^{13}C 100 Mhz NMR of α-DNP Orn in DMSO d_6



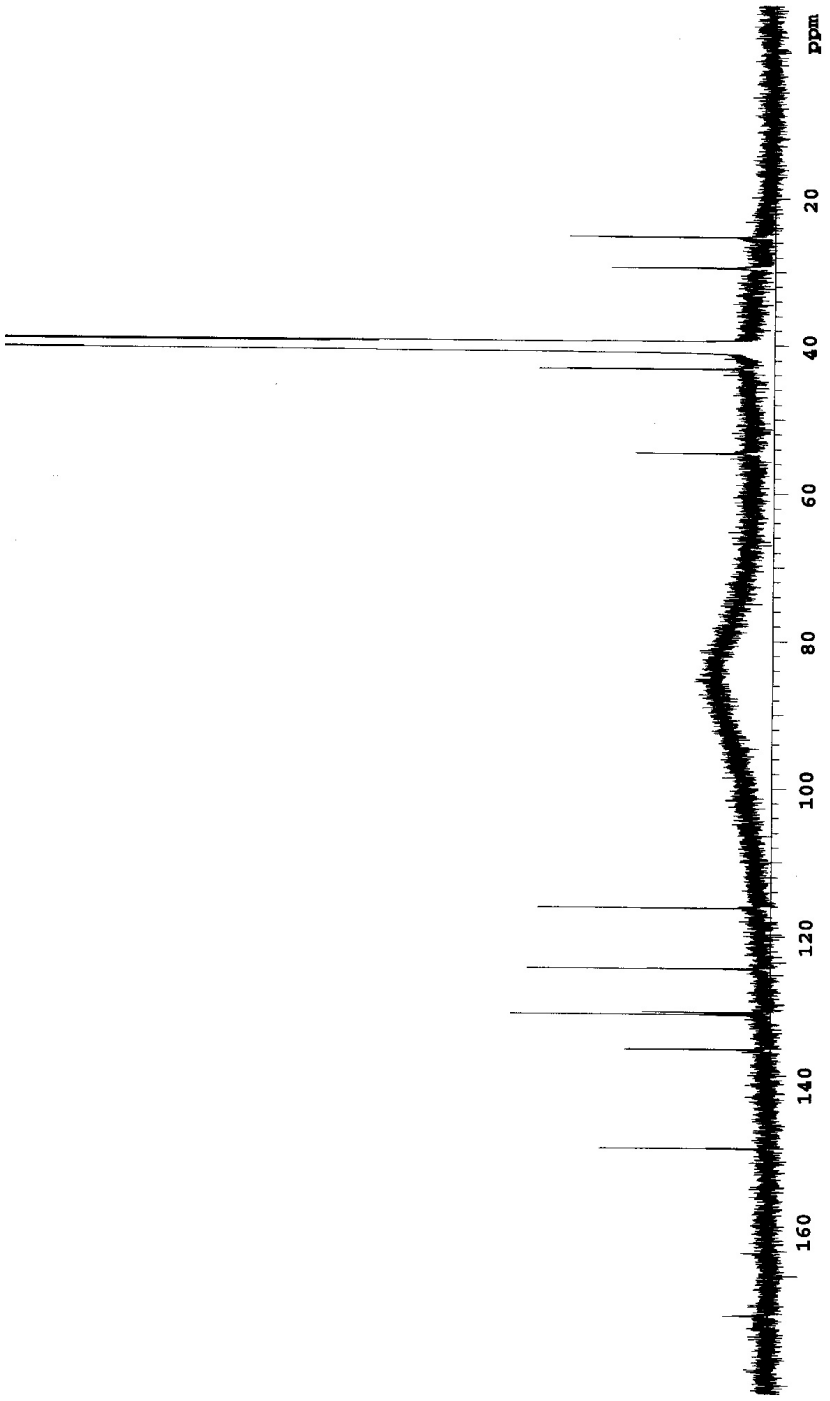
^{1}H 400 MHz NMR of α-DNP Orn



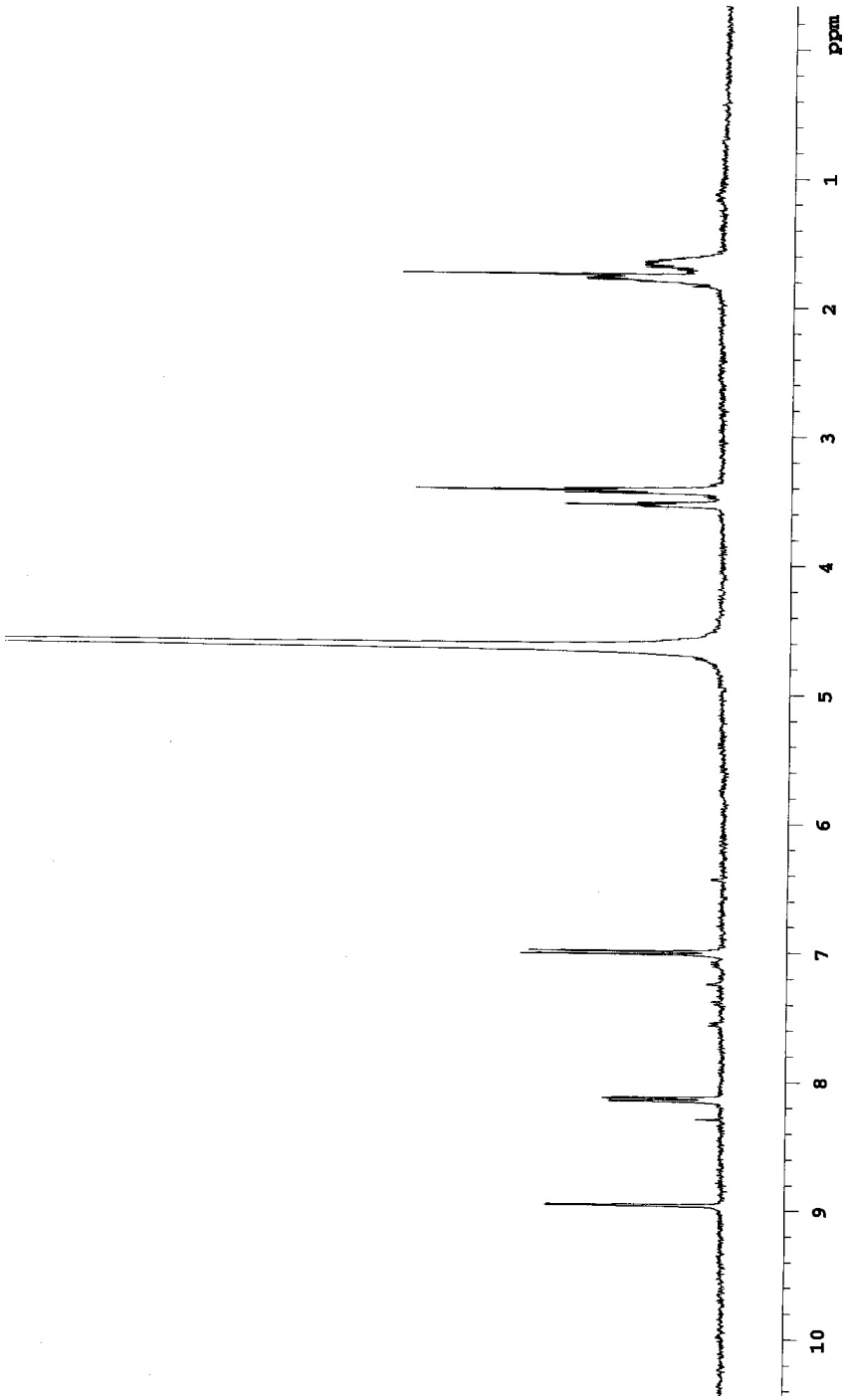
LC-MS of α-DNP Orn



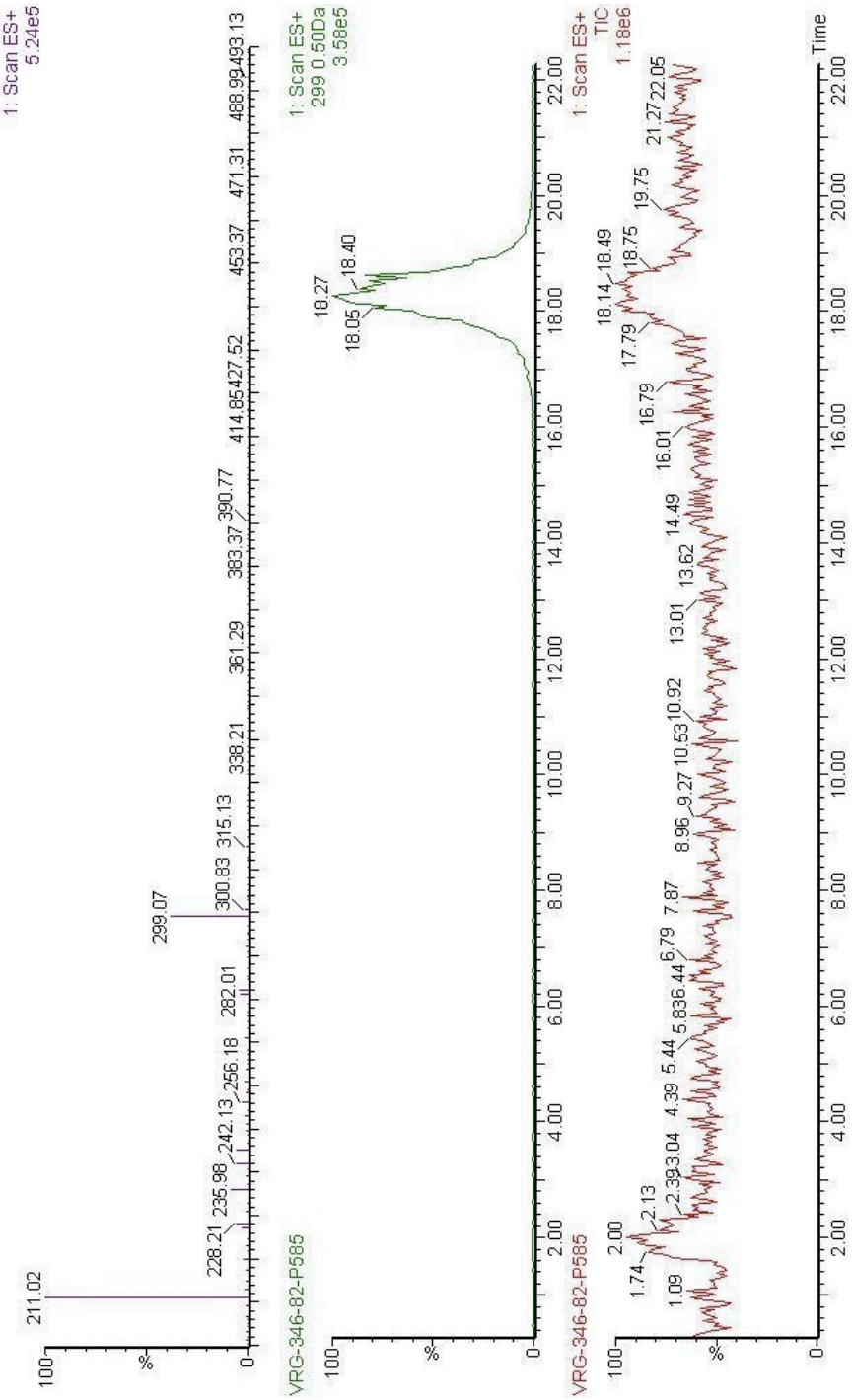
(f) ^{13}C 100 MHz NMR of δ-DNP Orn in DMSO d_6



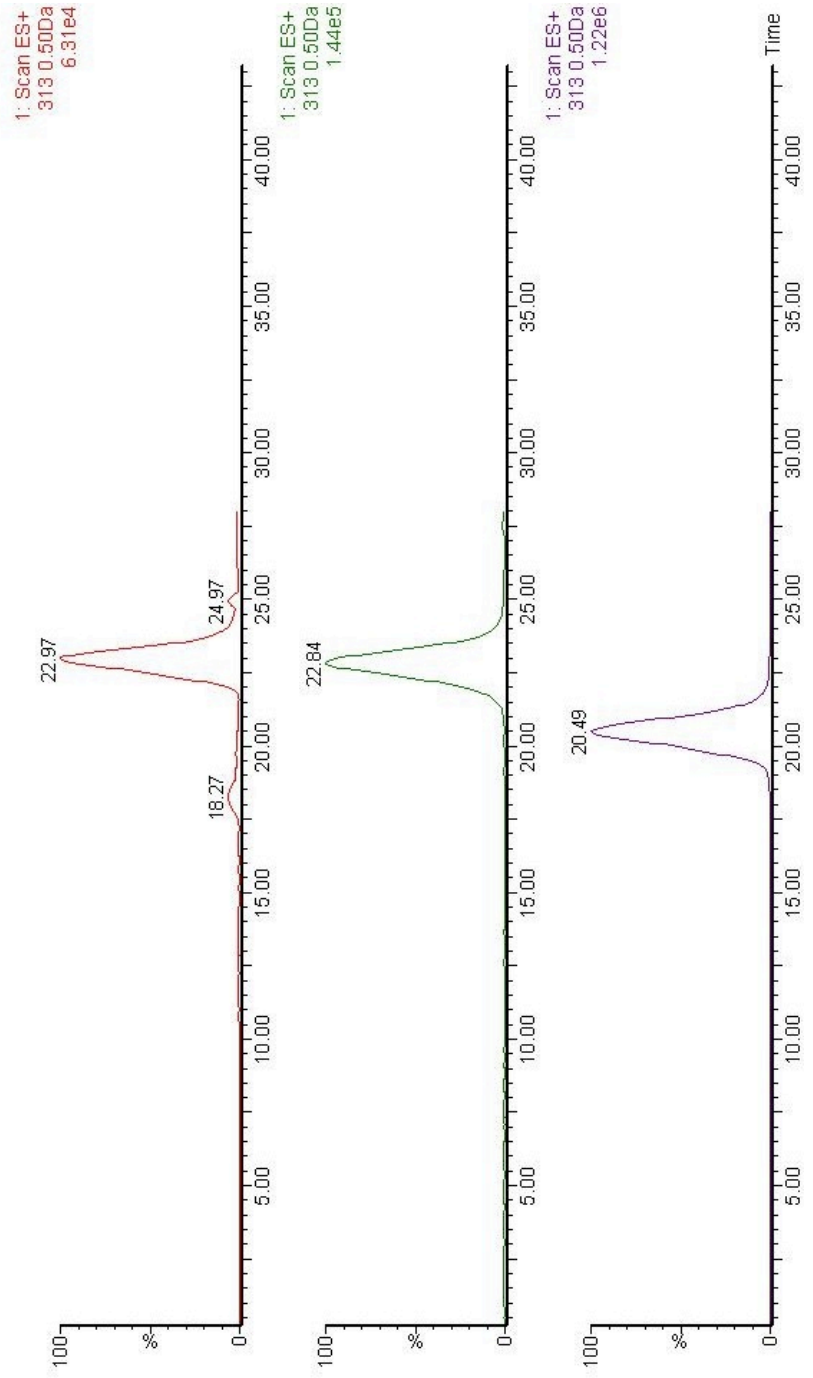
^{1}H 400 MHz NMR of δ-DNP Orn



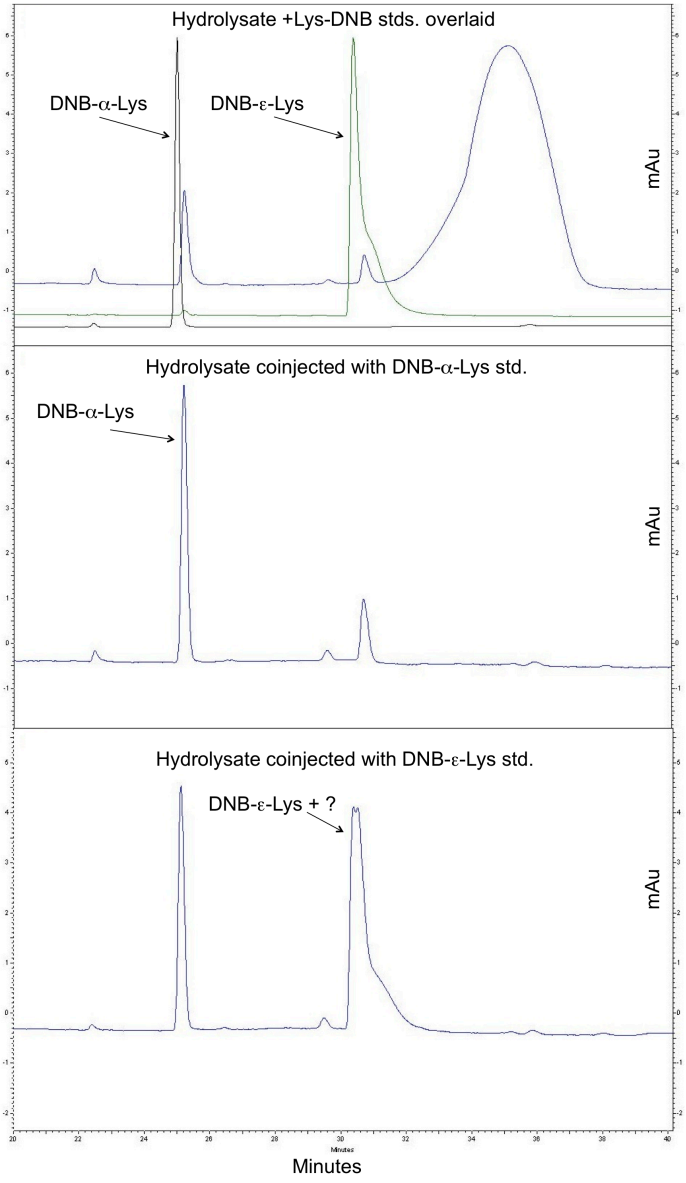
LC-MS of δ-DNP Orn



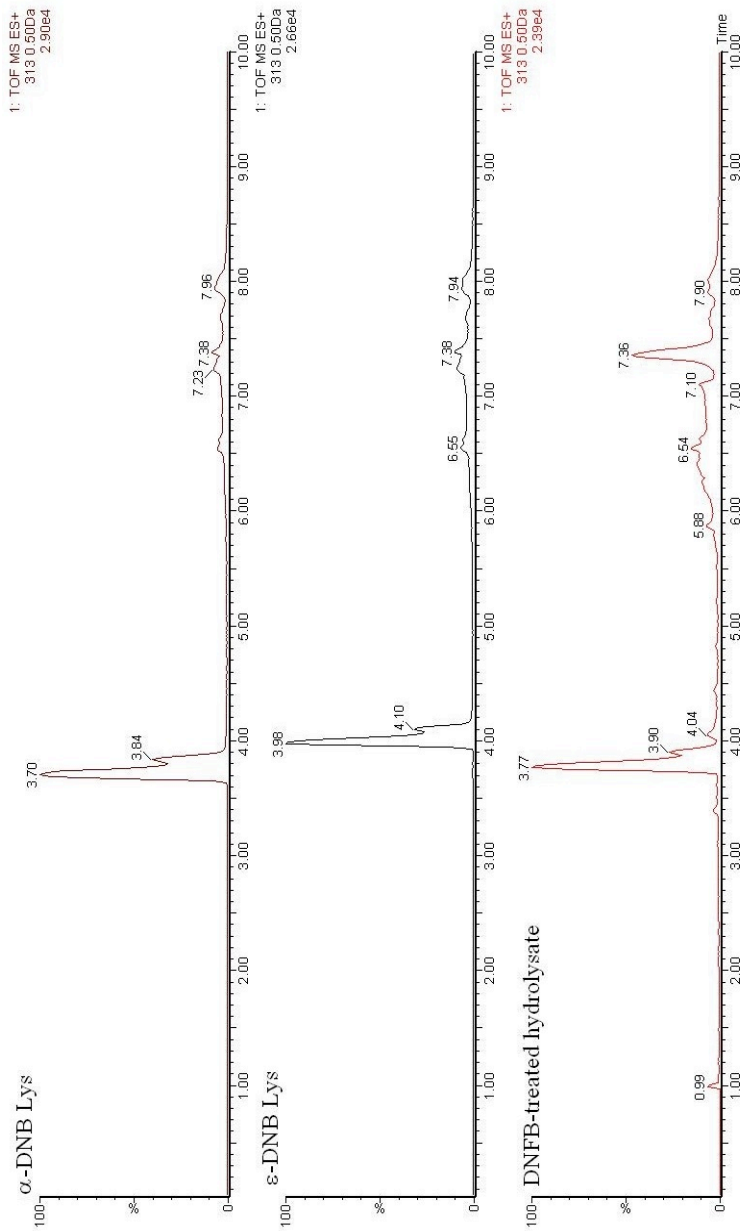
(g) LC-MS chromatograms selected for mass of DNP-Lys; top: hydrolysate, middle: ε-DNP Lys standard; bottom α-DNP Lys standard



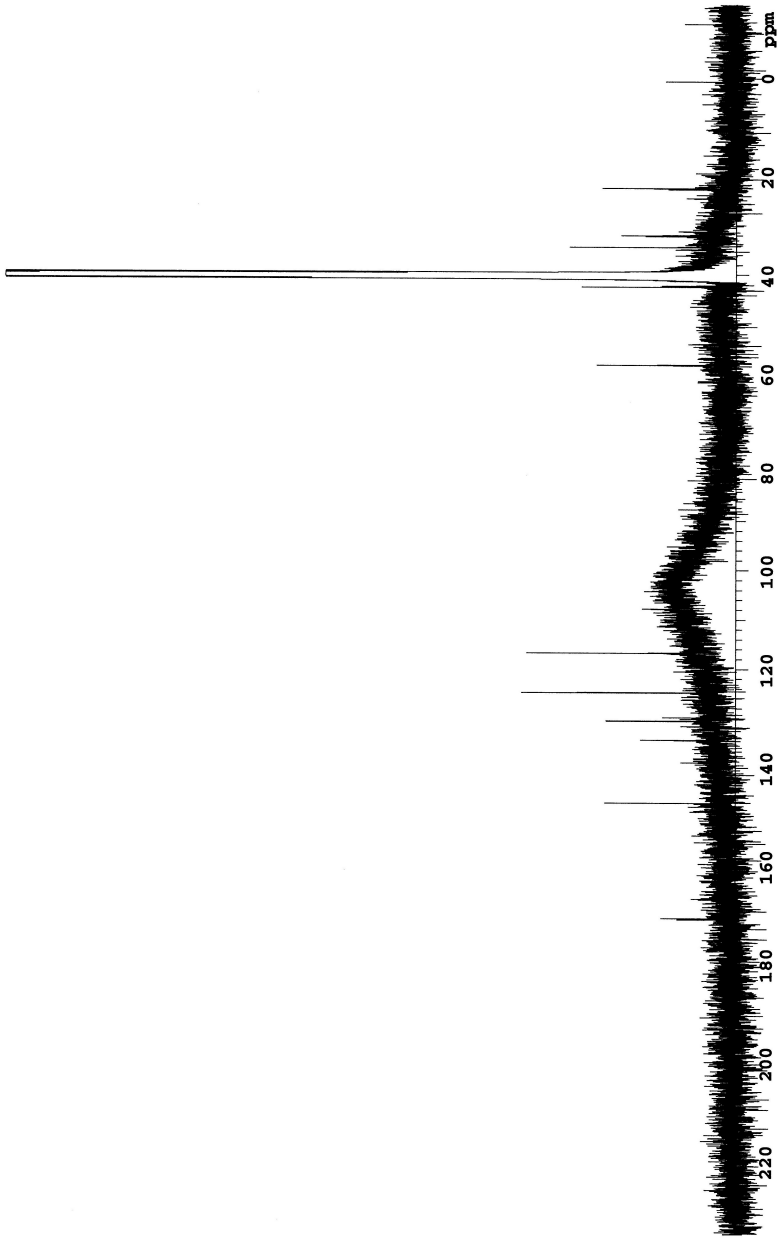
(h)


Hydrolysate +Lys-DNB stds. overlaid
DNB-α-Lys
DNB-ε-Lys
mAu
Hydrolysate coinjected with DNB-α-Lys std.
DNB-α-Lys
mAu
Hydrolysate coinjected with DNB-ε-Lys std.
DNB-ε-Lys + ?
mAu
Minutes

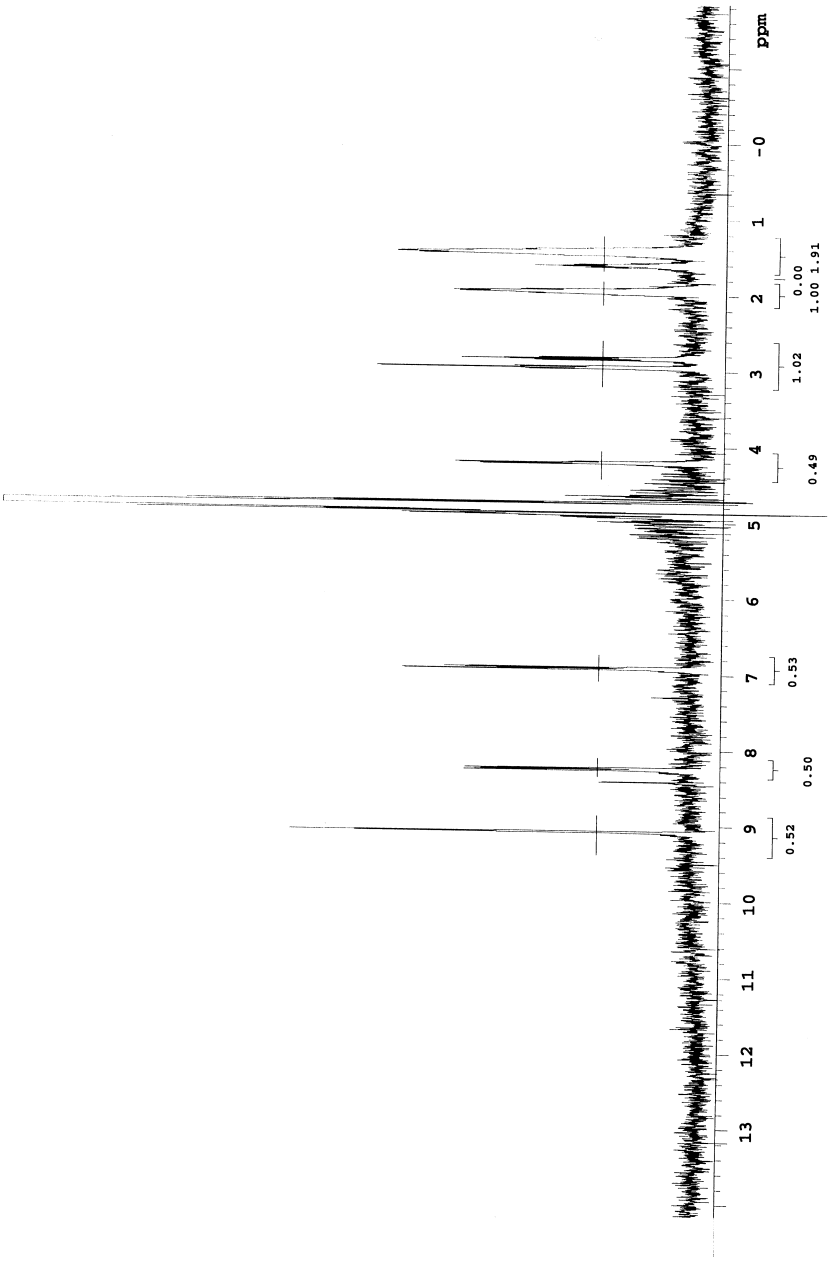
(i)


α-DNB Lys
ε-DNB Lys
DNFB-treated hydrolysate
1: TOF MS ES+
313 0.50Da
2.90e4
2.66e4
2.39e4
3.70
3.84
7.23
7.38
7.96
3.98
4.10
6.55
7.38
7.94
3.77
0.99
3.90
4.04
5.88
6.54
7.10
7.36
7.90
Time

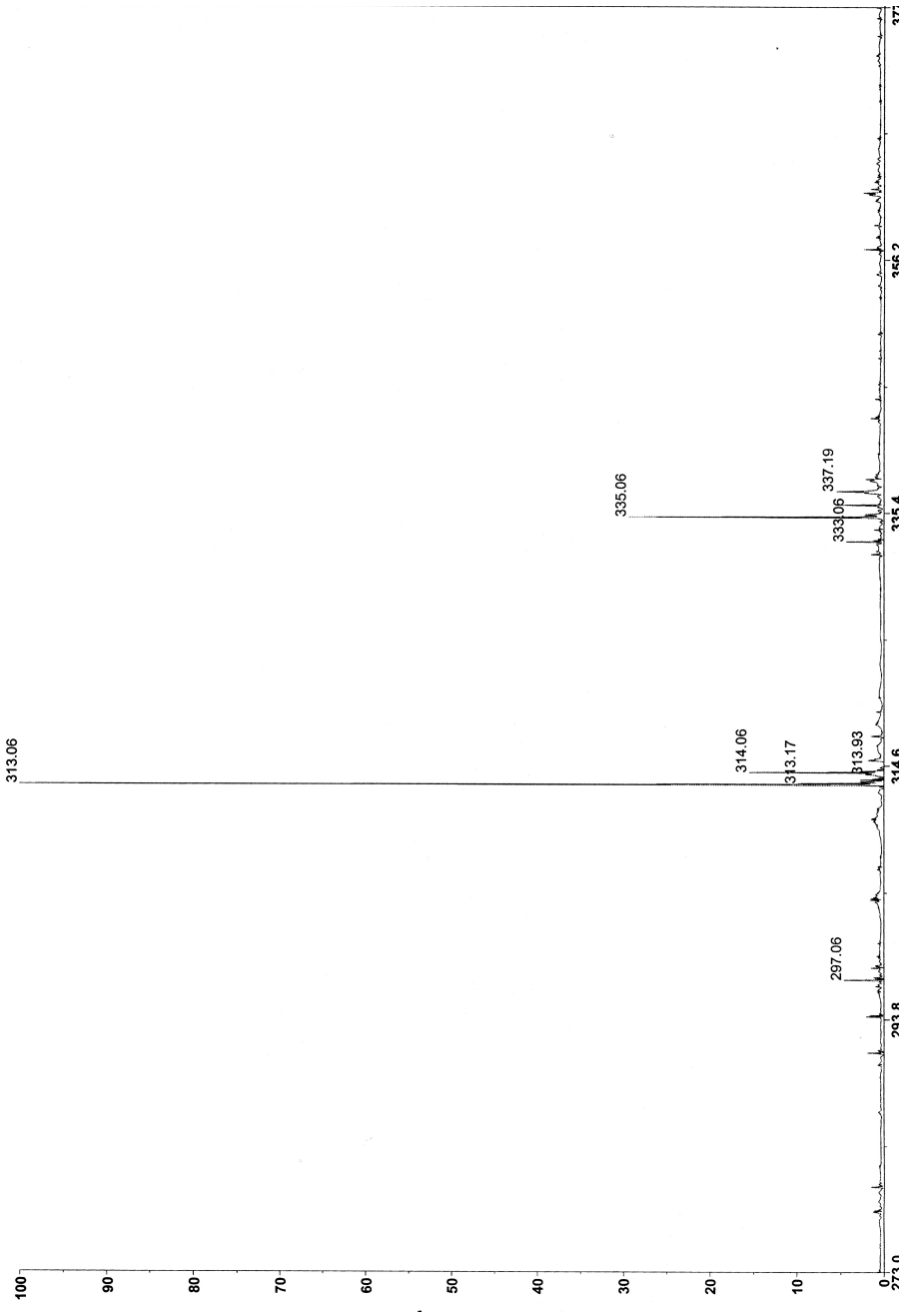
(j) ^{13}C 100 MHz NMR of α-DNP Lys in DMSO d_6



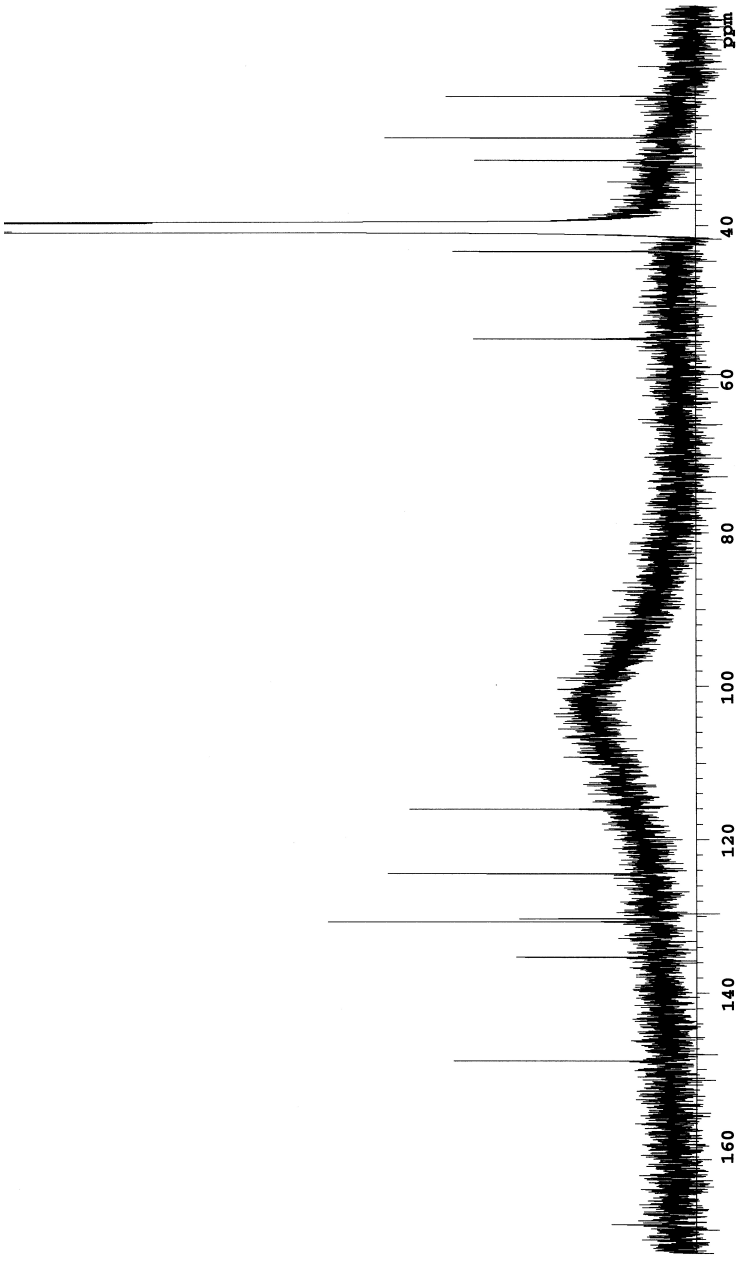
^{1}H 400 MHz NMR of α-DNP Lys in D_2O



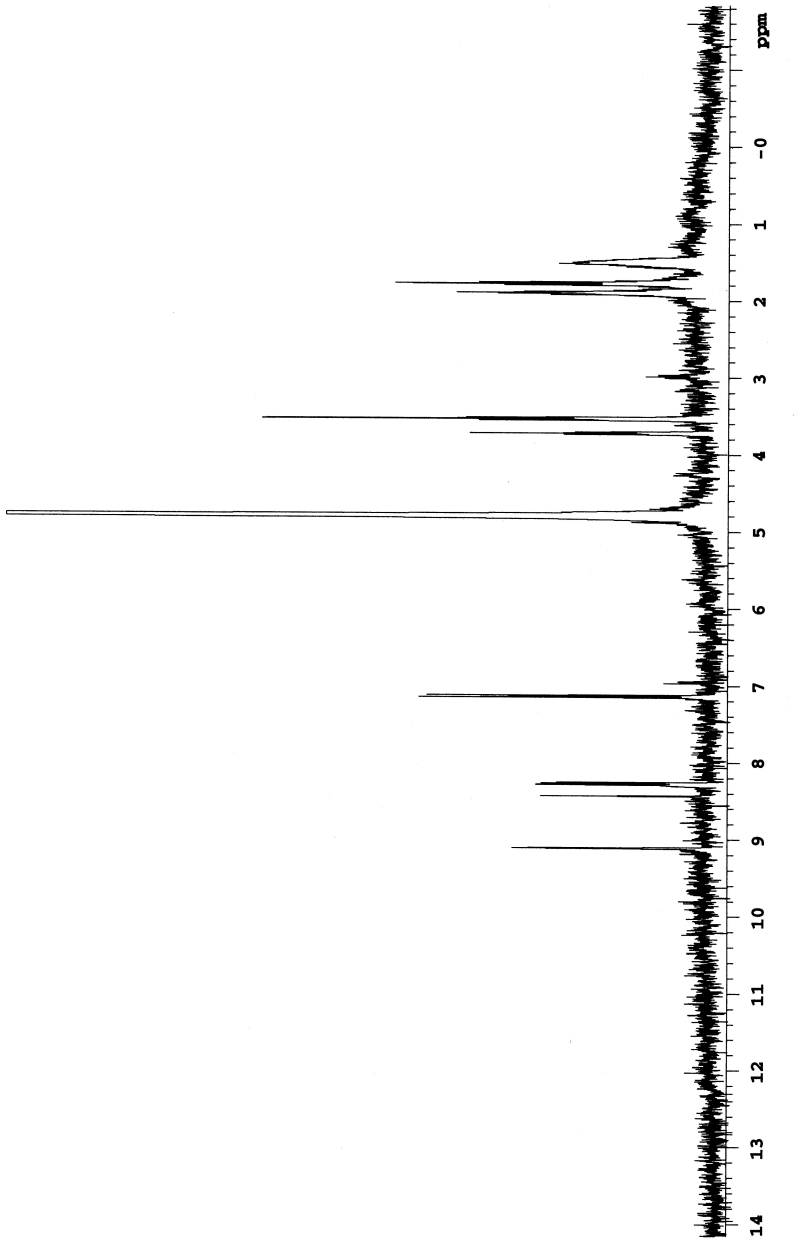
MALDI-MS of α-DNP Lys



(k) ^{13}C 100 MHz NMR of ε-DNP Lys in DMSO (d_6)



^{1}H 400 MHz NMR of ε-DNP Lys in D_2O



MALDI-MS of ε-DNP Lys

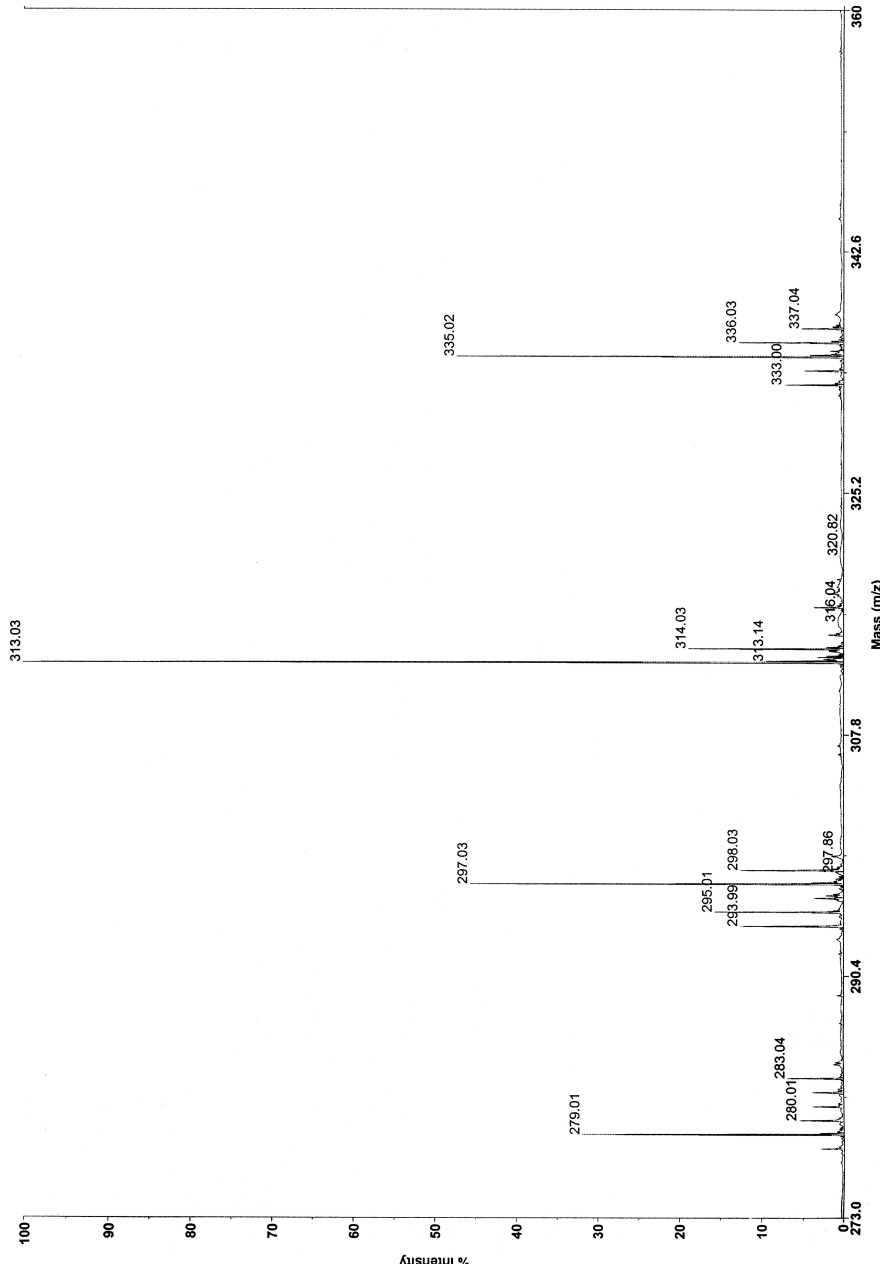
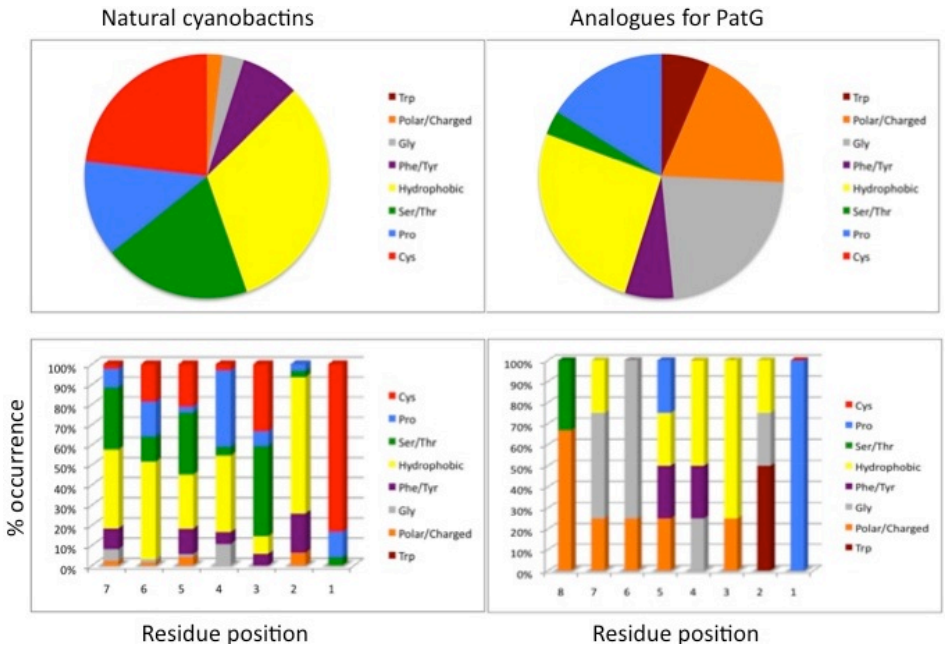


Figure S5. Shown below are pie and bar charts that depict the typical amino acid distribution found in selected analogues tested in this work compared to the distribution found in naturally occurring cyanobactins.



References:
(1) Suh-Lailam, B. B.; Hevel, J. M. *Anal Biochem* **2009**, *387*, 130-2.
(2) Tawfik, K.A.; Jeffs P.; Bray B.; Dubay G.; Falkinham J.O.; Mesbah M.; Youssef D.; Khalifa S.; Schmidt E.W. *Org Lett*. **2010**, *12*, 664-6.
(3) Levy, A. L.; Chung, D. *J. Am. Chem. Soc*. **1955**, *77*, 2899-2900.
(4) Pataki, G. *Techniques of Thin-Layer Chromatography (Revised Edition)*; Ann Arbor Science Publishers, Inc.: Ann Arbor, Michigan, 1968.

CHAPTER 5

INSIGHTS INTO HETEROCYCLIZATION FROM TWO HIGHLY SIMILAR ENZYMES

Manuscript reproduced with permission from:

McIntosh, J. A., Donia, M. S., Schmidt, E. W. (2010) Insights into heterocyclization from two highly similar enzymes, *J. Am. Chem. Soc. 132 (12)*, 4089-4091.

Note: my contribution to this paper was in planning many of the experiments and also in performing all experiments with the exception of the phylogenetic analyses and mass spectrometry data generated by the University of Utah Mass Spectrometry and Proteomics core facility.

J|A|C|S COMMUNICATIONS

Published on Web 03/08/2010

Insights into Heterocyclization from Two Highly Similar Enzymes

John A. McIntosh, Mohamed S. Donia, and Eric W. Schmidt*

Department of Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112

Received December 30, 2009; E-mail: ews1@utah.edu

A problem for chemists in the metagenomic era is how to approach the vast amounts of data coming from metagenome sequences. Metagenomes are mixed samples of multiple organisms from the environment, including symbiotic associations of bacteria with animals.[1−4] We have been developing the use of metagenomes from symbiotic bacteria for pathway engineering, allowing natural sequence variation to define parameters for gene-based organic synthesis.[5,6] Here, we provide an example of how metagenome sequence analysis contributes to understanding biosynthetic enzyme function using the thiazoline and oxazoline synthases, PatD and TruD. These proteins in turn are useful in the synthesis of diverse natural product libraries.

The first natural products pathways identified and characterized by sequencing of metagenomic DNA and comparison are the patellamide (*pat*) and trunkamide (*tru*) pathways to the cyanobactins, ribosomal peptides that are extensively posttranslationally modified (Figure 1A).[5,7] We previously used a metagenome sequence analysis method to determine that enzymes from *pat* and *tru* are extremely broad-substrate selective and can be used for *in vivo* synthesis of diverse natural and unnatural cyanobactin derivatives. Essentially identical (>99%) enzymes within the *tru* or *pat* pathway accept numerous substrates. Although enzymes are identical within pathways, there are differences between *pat* and *tru*. In particular, there are enzyme domains that are virtually identical between pathways, and there are more divergent domains. These changes have functional consequences that can be precisely tracked (Figure 1).

Both *pat* and *tru* were initially cloned from bacteria living symbiotically with marine animals, the ascidians. Over 60 cyanobactins, including the well-known metabolites trunkamide and patellamide C,[8,9] have been isolated from ascidians; sequencing of metagenomic DNA has identified more than 30 additional relatives.[6] Cyanobactins are ribosomal peptide natural products, and comparison of metagenomic sequencing data with known compounds provided a natural mutagenesis in which evolution has mutated every amino acid in the natural products while maintaining identical modifying enzymes.[5] The ribosomal peptides are encoded on precursor peptides with length of ∼70 amino acids. Precursors include leader sequences (∼35 amino acids) and two sets each of recognition and product-coding sequences.

Despite their demonstrated usefulness in synthesizing diverse compounds, the biosynthetic rules behind posttranslational modification patterns remained elusive. For example, products of the *pat* pathway contain Cys, Ser, and Thr residues that are mostly heterocyclic. By contrast, known representatives of the *tru* pathway are heterocyclic at Cys but reverse prenylated at Ser and Thr (Figure 1).

Previous work has shown that essentially identical enzymes within the *pat* or *tru* pathways lead to these diverse products.[5,6,10] The enzymes modify precursor peptides that have conserved leader sequences and enzyme recognition sequences. Between the conserved recognition sequences are cassettes that directly encode the final natural products. These cassettes are modified by heterocyclization, prenylation, and N−C circularization to yield the mature natural products. Cassette sequences are hypervariable, with substitutions accepted at any amino acid position, leading to large natural product libraries.

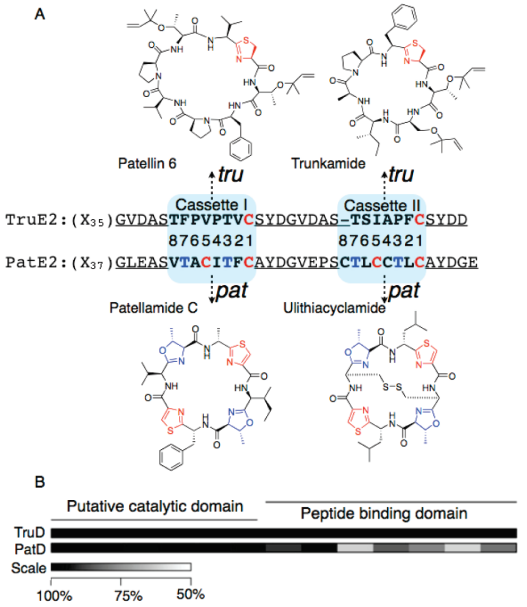


Figure 1. Biosynthesis of heterocyclic cyanobactins. (A) Enzymes modify ribosomal precursor peptides into mature products. Trunkamide group: top; patellamide group: bottom. (B) An alignment of PatD and TruD proteins, showing putative catalytic and peptide-binding domains. Catalytic domains are >99% identical; peptide-binding domains are >77% identical.

Six, seven, or eight amino acids are found in ascidian cyanobactins. Numbering from C- to N-terminus, all positions except position 2 naturally contain the heterocyclizable residues, Cys, Ser, or Thr, in one or more of the 60 known cyanobactins. The *pat* group includes members in which heterocycles occur in positions 1, 3, 5, 6, and 7 (Figure 2). By contrast, cycles are only found in position 1 in the *tru* group. In *tru* products, only Cys is heterocyclized, while Ser and Thr are prenylated; Cys, Ser, or Thr are cyclic in the *pat* products, and prenylation is not found. In all of these peptides, there are numerous Ser residues outsitde of the cassettes that are not modified.

These natural substitution patterns are particularly striking given the overall enzyme conservation. Despite chemical differences between *pat* and *tru*, all enzymes are homologous between these pathways. In fact, the didomain heterocyclase enzymes, PatD and TruD, are >99% identical in their N-terminal domains and only

10.1021/ja9107116

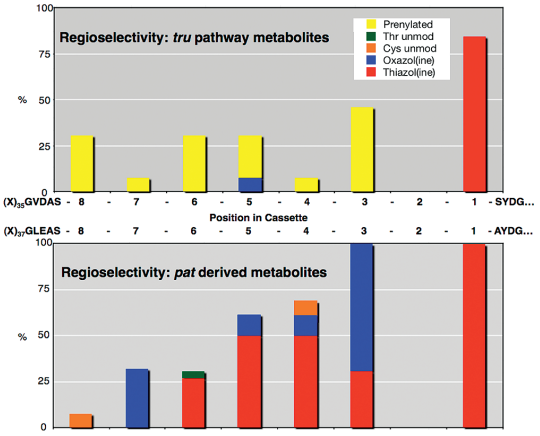


Figure 2. Enzyme selectivity defined by metagenome sequence and chemical analysis. Natural products from the *pat* (top) and *tru* (bottom) pathways contain modified Cys, Ser, and Thr residues in the defined positions. *y*-axis: % of natural products containing these modifications. *x*-axis: amino acid position in cassette. Empty space above the bars denotes hypervariability: diverse amino acids occupy these positions in isolated natural products.

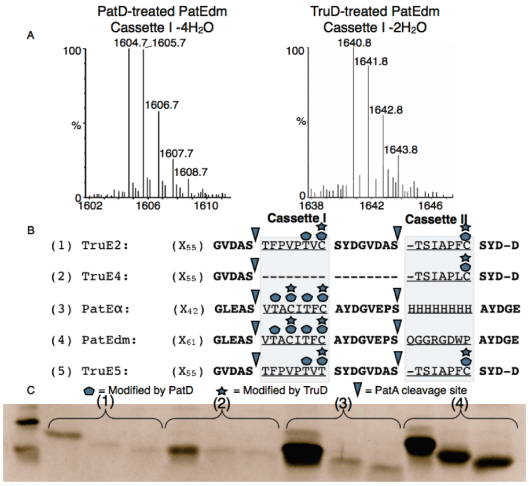


Figure 3. Reaction characterization and substrate specificity of TruD and PatD. (A) MALDI MS of reaction products. (B) Position of modifications identified by ESI-MS/MS. (C) Representative SDS-PAGE gel showing band shifts of precursor peptides 1−4. The left of each triplet is a standard of the peptide, middle is the TruD-modified peptide, and right is the PatD-modified peptide.

∼77% identical in their C-terminal domains. The questions addressed in this study are how such similar enzymes and pathways lead to different posttranslational products and how these enzymes can be used to synthesize diverse chemical derivatives.

Microcin B17 and streptolysin S synthetases were the first and second heterocyclase enzymes to be characterized *in vitro*.[11−13] Although they are only distantly related to PatD and TruD, comparison of these enzymes showed that the PatD/TruD catalytic domain is at the N-terminus, while the C-terminus functions primarily to bind the substrate peptide. This led to the question, do these enzymes operate in a chemoselective fashion (*O* vs *S* in heterocycles), or is regioselectivity important, as might be implicated by peptide binding differences? Here, we use biochemical experiments to define PatD and TruD as regioselective heterocyclases, which catalyze thiazoline and oxazoline biosynthesis.

When the *pat* gene cluster encoding PatD was previously expressed in *Escherichia coli*, we observed products containing the natural heterocycle pattern, including thiazoline and oxazoline.[6,7] Similarly, heterologous expression of the *tru* pathway led to production of thiazoline-containing natural products with prenylated Ser/Thr residues.[5] Thus, PatD and TruD were implicated as probable heterocyclases, but they had not been purified or characterized. Both genes were cloned and expressed here as N-terminal His-tagged proteins (Figure S1, Supporting Information). The *patD* clone includes two point mutations in the N-terminal domain, which did not change enzyme function. The *truD* gene was cloned into the C-terminus of an existing *patD* construct because of toxicity problems. Practically, this cloning strategy ensured that PatD and TruD were in fact 100% identical (instead of merely >99%) in their presumed catalytic N-terminal domains. A series of substrates were constructed by cloning and expressed recombinantly, which was extremely efficient in comparison to peptide synthesis because of the large size of the substrates (∼70 amino acids).

Purified PatD or TruD were used in experiments with the substrate analogues, PatEdm, PatEα, TruE2, TruE4, and TruE5 (Figure 3). All enzyme experiments in this study were done with varying substrate concentrations, with each condition performed independently at least in triplicate. We first noticed that PatD and TruD products exhibited band shift differences by SDS-PAGE. In all cases, TruD products migrated more rapidly than unmodified peptides, while PatD products migrated more rapidly still (Figure 3). The mobility shift was consistent with formation of thiazoline and/or oxazoline rings given that these heterocycles introduce a significant conformational restraint on the peptide backbone.[14−17]

Given that heterocycle formation causes a loss of 18 Da, we employed an MS approach to confirm that each of the five precursor peptides was modified by heterocyclization. First, intact mass electrospray ionization (ESI) was used to determine the total number of dehydrations catalyzed by PatD or TruD on these ∼9 kDa substrates, in comparison to unmodified control (Figure S2, Supporting Information). Subsequently, the enzyme products were treated by a specific protease, PatA, that cleaves upstream from each cassette's start site.[10] These smaller fragments were analyzed by MALDI and ESI to localize dehydrations to single cassettes (Figures S3 and S8, Supporting Information). Finally, these PatA-digested cassettes were also subjected to LC-Fourier transform ion cyclotron resonance (FT-ICR) and MS/MS to determine which amino acids within cassettes were heterocyclized (Figures S4 and S5, Supporting Information). For Cys, MS data confirmed that thiazoline was formed and ruled out other possible dehydration routes. In addition, for Thr, all available evidence supported oxazoline formation. This evidence included observation of the same type of fragmentation suppression seen in other heterocycle-containing peptides,[18] the absence of observed MS/MS fragments consistent with other modifications, and the SDS-PAGE mobility shift, which as noted above, is consistent with heterocycle formation. Additionally, the facts that these genes lead to oxazoline formation *in vivo* and that no other type of Thr dehydration has been observed in this compound family also support this interpretation. Despite this evidence, we nevertheless sought to completely eliminate the possibility of a reverse-Michael reaction (Figure S6, Supporting Information). Because activated double bonds are not very reactive with acids, while oxazolines are labile in acidic conditions, substrates were subjected to very mild acidic conditions. The resulting rehydrated products confirmed that all Thr modifications

were indeed due to oxazolines (Figure S6, Supporting Information). Therefore, PatD and TruD catalyzed the synthesis of thiazoline and oxazoline; other products were not observed in extensive experimental analysis.

TruE2 is a natural substrate that contains two Cys residues that are found as thiazoline in the final natural products, as well as a number of Ser and Thr residues that are prenylated naturally (Figure 1A). TruE2 contains two cassettes, with Cys in position 1 of each cassette. When the *tru* cluster was expressed in *E. coli*, these natural products were synthesized,[5] indicating that *tru* gene products modify TruE2 to produce thiazoline. When treated with TruD, both Cys residues were heterocyclized as expected, while none of the Ser/Thr residues in the molecule were modified. By contrast, when treated with PatD, which is not normally associated with TruE2, both Cys residues and an additional Thr residue in position 3 of cassette I were cyclized (Figure 3). An unnatural analog of TruE2, TruE4, contained only cassette II and only a single Cys residue; this residue was also cyclized by both PatD and TruD.

We next analyzed reactivity using unnatural substrate analogs from the *pat* pathway. In cassette I, both PatEdm and PatEα encode the natural patellamide C sequence which would normally be modified to contain two thiazole and two oxazoline residues (Figure 1). Indeed, patellamide C is synthesized in *E. coli* when PatEdm is coexpressed with *pat* enzymes.[6,7] When these peptide substrates were treated with PatD, two Cys and two Thr residues were cyclized in both PatEdm and PatEα, as found *in vivo* in *E. coli* expression. TruD, which is not normally associated with the *pat* pathway, behaves differently. None of the *tru* products we have so far examined contain Cys in position 5, as found in PatEdm. However, TruD readily cyclized both Cys residues in position 1 and position 5 (Figure 3).

From these experiments, we could not determine whether TruD was truly chemoselective for Cys or whether regioselectivity played a role. We therefore synthesized an unnatural variant, TruE5, that contained Thr in place of Cys at position 1 of cassette I. Both PatD and TruD were able to cyclize this new Thr residue, indicating that the reaction specificity of these enzymes is mainly due to regioselectivity (Figure S7, Supporting Information), though this reaction was much slower than Cys heterocyclization. Thus, although the enzymes select residues for modification based primarily on their position within cassettes, the chemical features of the modified residue may influence selectivity as well.

The regioselectivity of these enzymes clearly explains the observed product patterns in the ~60 ascidian-derived cyanobactins. In fact, previously the cyanobactin comoramide A was isolated with Thr heterocyclized in position 5 and prenylated in position 3.[19] Although genes for comoramide synthesis have not been cloned, the data described here allow this to be defined as a *tru*-like pathway. Another pathway type contains a prenylated Ser in position 5. In TruE2, PatD did not modify a Ser in position 6.

These experiments also indicate that TruD and PatD do not determine the regioselectivity of prenylation, since TruD leaves unmodified the residues that are presumably later prenylated. Although there are no prenyltransferase homologues in the pathways, our current hypothesis is that the TruF1 and TruF2 proteins are involved in this step.

The usefulness of these enzymes is primarily for *in vivo* synthesis of new compounds, not necessarily for the *in vitro* modification of discrete substrates. Previously, we demonstrated that, in principle, large libraries of natural and unnatural cyanobactin derivatives could be synthesized and screened in *E. coli*.[6] The enzymatic specificity results reported here will be greatly helpful in determining which sequences are appropriate for the development of a chemically diverse cyanobactin library. Metagenome sequence analysis enabled the discovery of methods to synthesize these libraries, and in this study the sequencing methods allowed us to obtain new insights into enzyme function. These methods should be applicable to other enzyme systems. Because the underlying genetic methods are now extremely fast and inexpensive, they are of practical utility in the enzymatic synthesis of new molecules.

Acknowledgment. This work was funded by NIH GM071425 and a Willard Eccles Fellowship as well as an ACS Medicinal Chemistry Predoctoral Fellowship sponsored by Sanofi Aventis to J.A.M. We thank Brian Hathaway and Michael Mathews for cloning PatEα and PatD; Chad Nelson, Krishna Parawar, and Jim Muller for mass spectrometry assistance; Adele Flail for graphical assistance; and Archana Yerra for technical assistance.

Supporting Information Available: Additional mass spectrometry data including FT-ICR, MS-MS, ESI-MS, and MALDI data, and SDS-PAGE analyses.This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Stein, J. L.; Marshall, T. L.; Wu, K. Y.; Shizuya, H.; DeLong, E. F. *J. Bacteriol.* **1996**, *178*, 591–599.
(2) Haygood, M. G.; Davidson, S. K. *Appl. Environ. Microbiol.* **1997**, *63*, 4612–4616.
(3) Schloss, P. D.; Handelsman, J. *Curr. Opin. Biotechnol.* **2003**, *14*, 303–310.
(4) Handelsman, J.; Rondon, M. R.; Brady, S. F.; Clardy, J.; Goodman, R. M. *Chem. Biol.* **1998**, *5*, R245–R249.
(5) Donia, M. S.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2008**, *4*, 341–343.
(6) Donia, M. S.; Hathaway, B. J.; Sudek, S.; Haygood, M. G.; Rosovitz, M. J.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2006**, *2*, 729–735.
(7) Schmidt, E. W.; Nelson, J. T.; Rasko, D. A.; Sudek, S.; Eisen, J. A.; Haygood, M. G.; Ravel, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7315–7320.
(8) Ireland, C. M.; Durso, A. R.; Newman, R. A.; Hacker, M. P. *J. Org. Chem.* **1982**, *47*, 1807–1811.
(9) Carroll, A. R.; Coll, J. C.; Bourne, D. J.; MacLeod, J. K.; Zabriskie, T.; Ireland, C. M.; Bowden, B. F. *Aust. J. Chem.* **1996**, *49*, 659–667.
(10) Lee, J.; McIntosh, J. A.; Hathaway, B. J.; Schmidt, E. W. *J. Am. Chem. Soc.* **2009**, *131*, 2122–2124.
(11) Lee, S. W.; Mitchell, D. A.; Markley, A. L.; Hensler, M. E.; Gonzalez, D.; Wohlrab, A.; Dorrestein, P. C.; Nizet, V.; Dixon, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5879–5884.
(12) Milne, J. C.; Roy, R. S.; Eliot, A. C.; Kelleher, N. L.; Wokhlu, A.; Nickels, B.; Walsh, C. T. *Biochemistry* **1999**, *38*, 4768–4781.
(13) Li, Y. M.; Milne, J. C.; Madison, L. L.; Kolter, R.; Walsh, C. T. *Science* **1996**, *274*, 1188–1193.
(14) Milne, B. F.; Long, P. F.; Starcevic, A.; Hranueli, D.; Jaspars, M. *Org. Biomol. Chem.* **2006**, *4*, 631–638.
(15) Abbenante, G.; Fairlie, D. P.; Gahan, L. R.; Hanson, G. R.; Piersens, G. K.; van den Brenk, A. L. *J. Am. Chem. Soc.* **1996**, *118*, 10382–10388.
(16) Bernhardt, P. V.; Comba, P.; Fairlie, D. P.; Gahan, L. R.; Hanson, G. R.; Lotzbeyer, L. *Chemistry* **2002**, *8*, 1527–1536.
(17) Walsh, C. T.; Nolan, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5655–5656.
(18) Belshaw, P. J.; Roy, R. S.; Kelleher, N. L.; Walsh, C. T. *Chem. Biol.* **1998**, *5*, 373–384.
(19) Rudi, A.; Aknin, M.; Gaydou, E. M.; Kashman, Y. *Tetrahedron* **1998**, *54*, 13203–13210.

JA9107116

Supporting Online Material for

Insights into heterocyclization from two highly similar enzymes

John A. McIntosh, Mohamed S. Donia, Eric W. Schmidt*

Department of Medicinal Chemistry, University of Utah, Salt Lake City, UT 84112, USA

ews1@utah.edu

This file contains:

Figures S1-S8

Table S1

Experimental methods

Figure S1. SDS-PAGE of purified TruD and PatD proteins. From the left, lanes are: (1) TruD (2) Ladder (3) PatD.

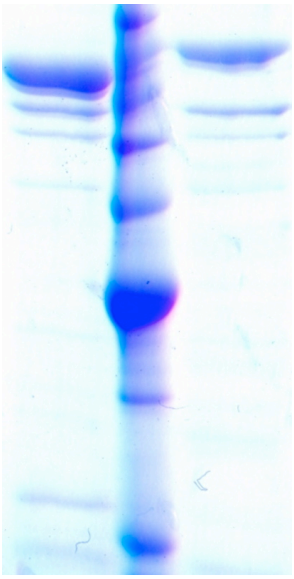


Figure S2. Intact analysis of TruD- and PatD-treated TruE2. (a) shows ESI-(-) intact analysis of PatD-treated TruE2. Result is consistent with the formation of two thiazoline rings and a single oxazoline ring; (b) shows ESI-(-) intact analysis of TruD-treated TruE2. Result is consistent with the formation of two thiazoline rings; (c) SDS-PAGE showing typical band shifts resulting from TruE2 modification reactions. From right to left, lanes are as follows: (1) TruE2 standard (2) TruD-treated TruE2 (3) TruD-treated TruE2 + TruE2 standard (4) PatD-treated TruE2 (5) PatD-treated TruE2 + TruE2 standard (6) TruE2 standard (7) ladder.

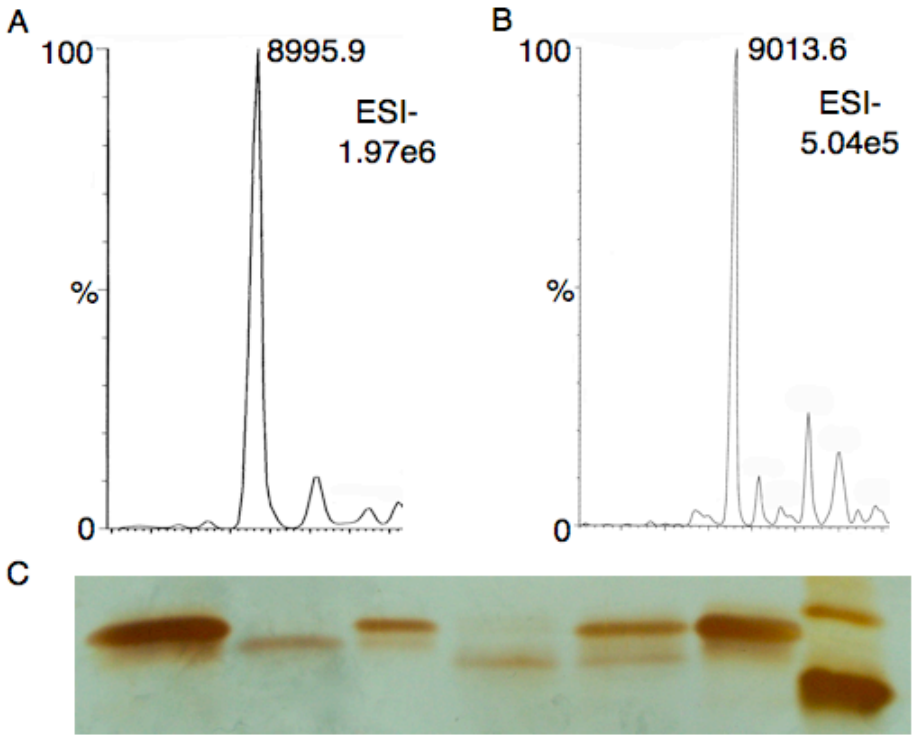
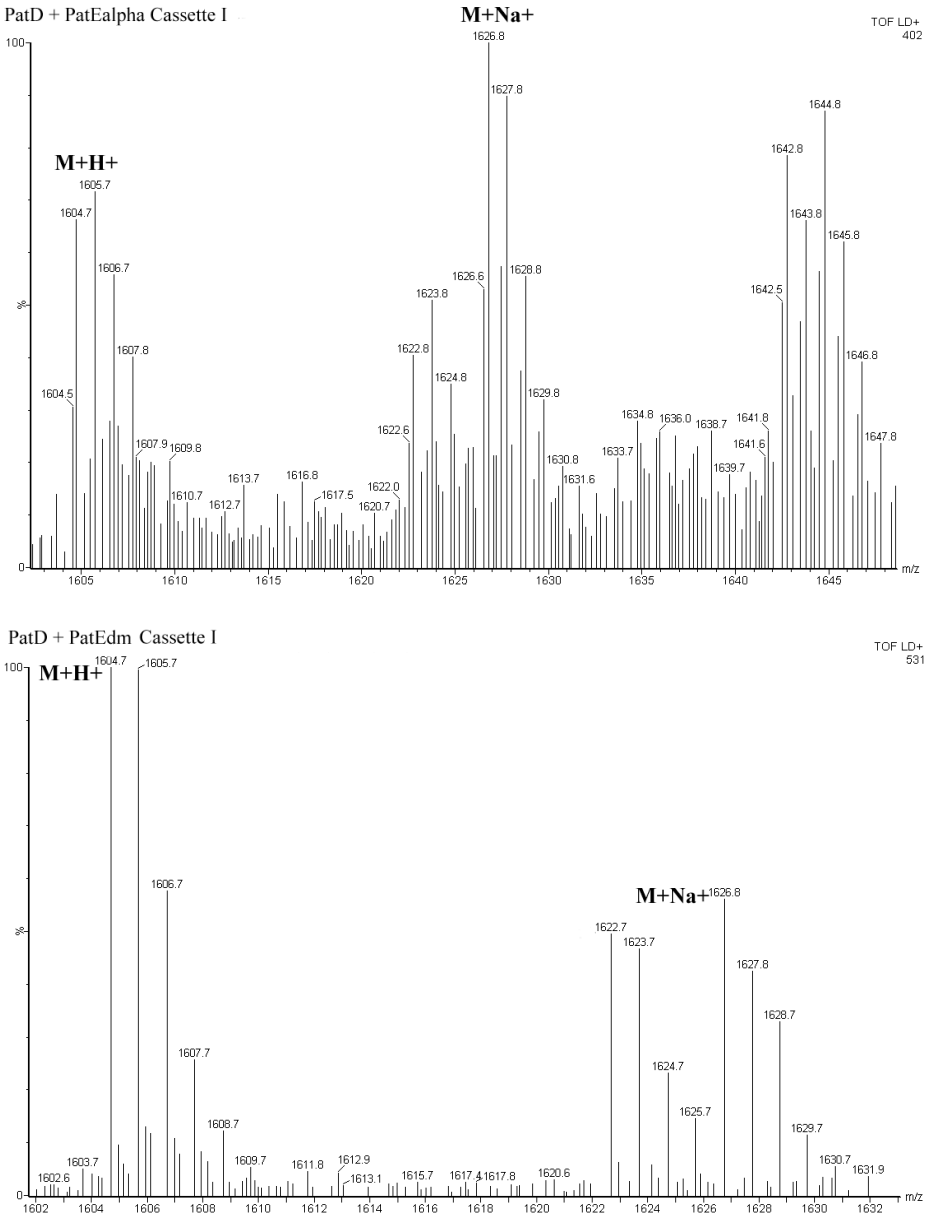
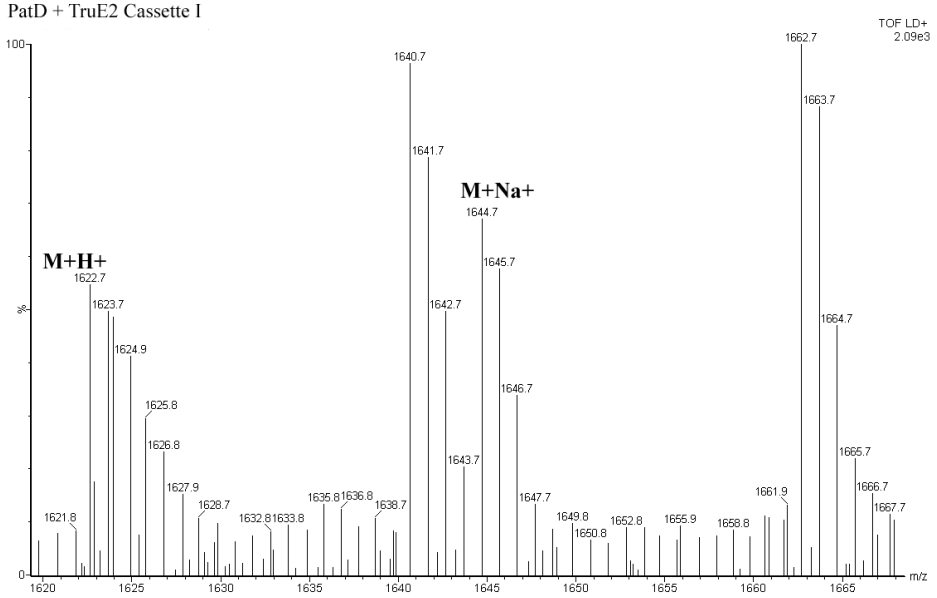


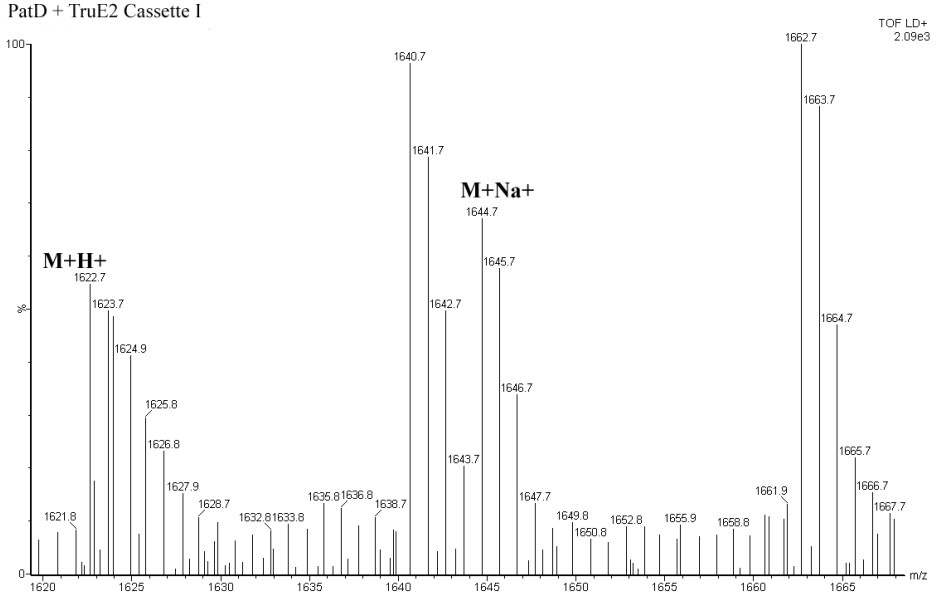
Figure S3. MALDI-MS spectra demonstrating modification of TruE2, PatEα, and PatEdm product cassettes by PatD and TruD; peptides were all digested with PatA prior to analysis.

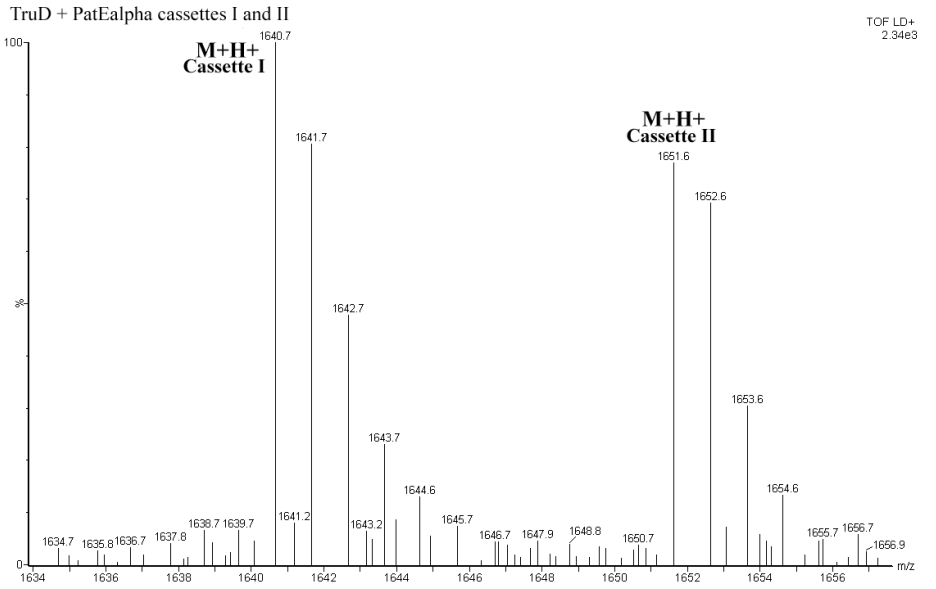


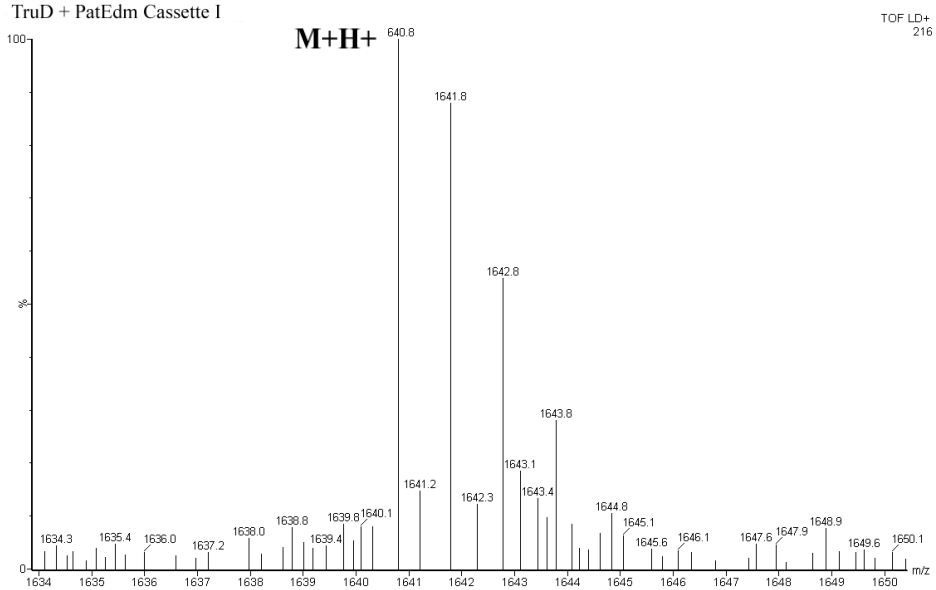
PatD + TruE2 Cassette I

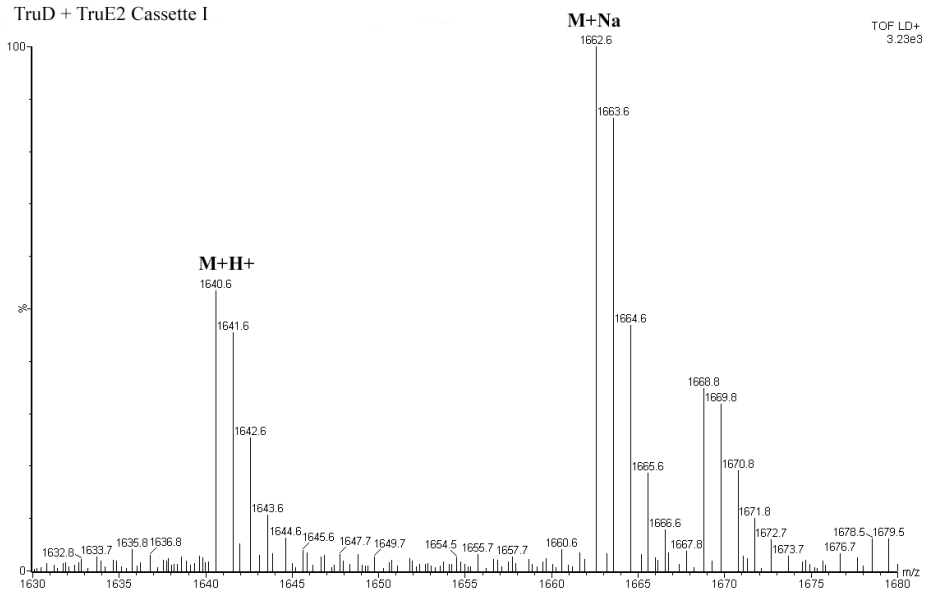


PatD + TruE2 Cassette I




TruD + PatEalpha cassettes I and II
TOF LD+
2.34e3
M+H+
Cassette I
M+H+
Cassette II
1640.7
1641.7
1642.7
1643.7
1644.6
1645.7
1646.7
1647.9
1648.8
1650.7
1651.6
1652.6
1653.6
1654.6
1655.7
1656.7
1656.9
1634.7
1635.8
1636.7
1637.8
1638.7
1639.7
1641.2
1643.2
m/z


TruD + PatEdm Cassette I
TOF LD+
216
M+H+
640.8
1641.8
1642.8
1643.8
1643.1
1643.4
1641.2
1642.3
1644.8
1645.1
1645.6
1646.1
1647.6
1647.9
1648.9
1649.6
1650.1
1634.3
1635.4
1636.0
1637.2
1638.0
1638.8
1639.4
1639.8
1640.1
m/z


TruD + TruE2 Cassette I
M+Na
1662.6
TOF LD+
3.23e3
100
%
0
M+H+
1640.6
1641.6
1642.6
1643.6
1644.6
1645.6
1647.7
1649.7
1654.5
1655.7
1657.7
1660.6
1663.6
1664.6
1665.6
1666.6
1667.8
1668.8
1669.8
1670.8
1671.8
1672.7
1673.7
1676.7
1678.5
1679.5
1632.8
1633.7
1635.8
1636.8
1630
1635
1640
1645
1650
1655
1660
1665
1670
1675
1680
m/z

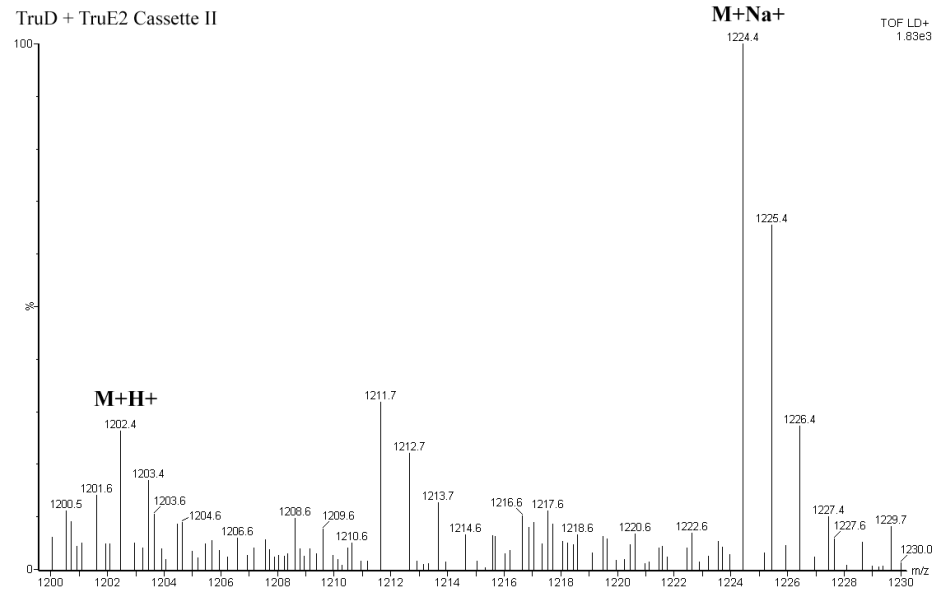
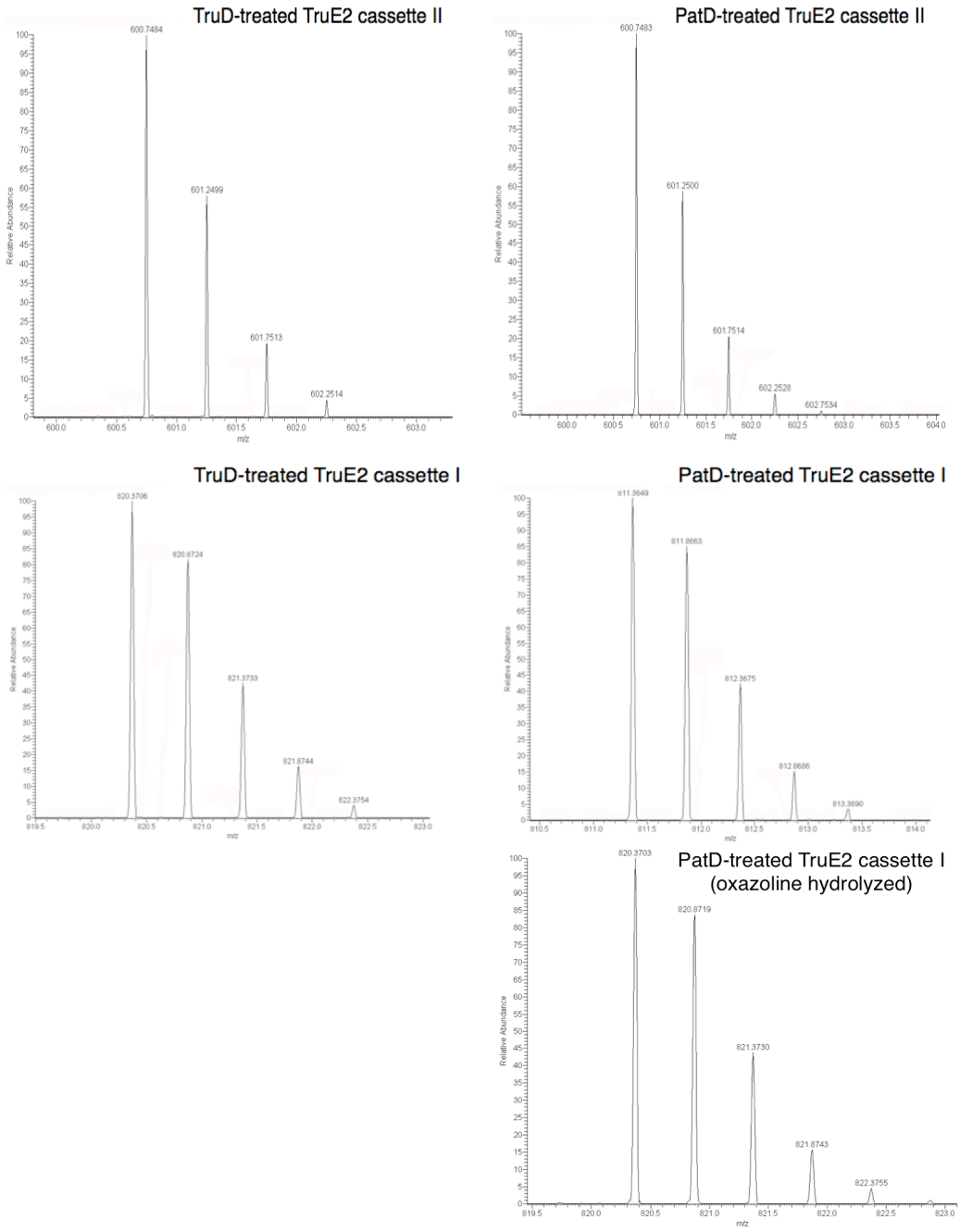

TruD + TruE2 Cassette II
M+Na+
1224.4
TOF LD+
1.83e3
100
%
0
M+H+
1202.4
1203.4
1201.6
1200.5
1203.6
1204.6
1206.6
1208.6
1209.6
1210.6
1211.7
1212.7
1213.7
1214.6
1216.6
1217.6
1218.6
1220.6
1222.6
1225.4
1226.4
1227.4
1227.6
1229.7
1230.0
1200
1202
1204
1206
1208
1210
1212
1214
1216
1218
1220
1222
1224
1226
1228
1230
m/z

Figure S4. (a) FT-ICR characterization of TruD- and PatD-treated TruE2, digested with PatA. (b) FT-ICR characterization of TruD- and PatD treated PatEdm, digested with PatA.

(a)



(b)

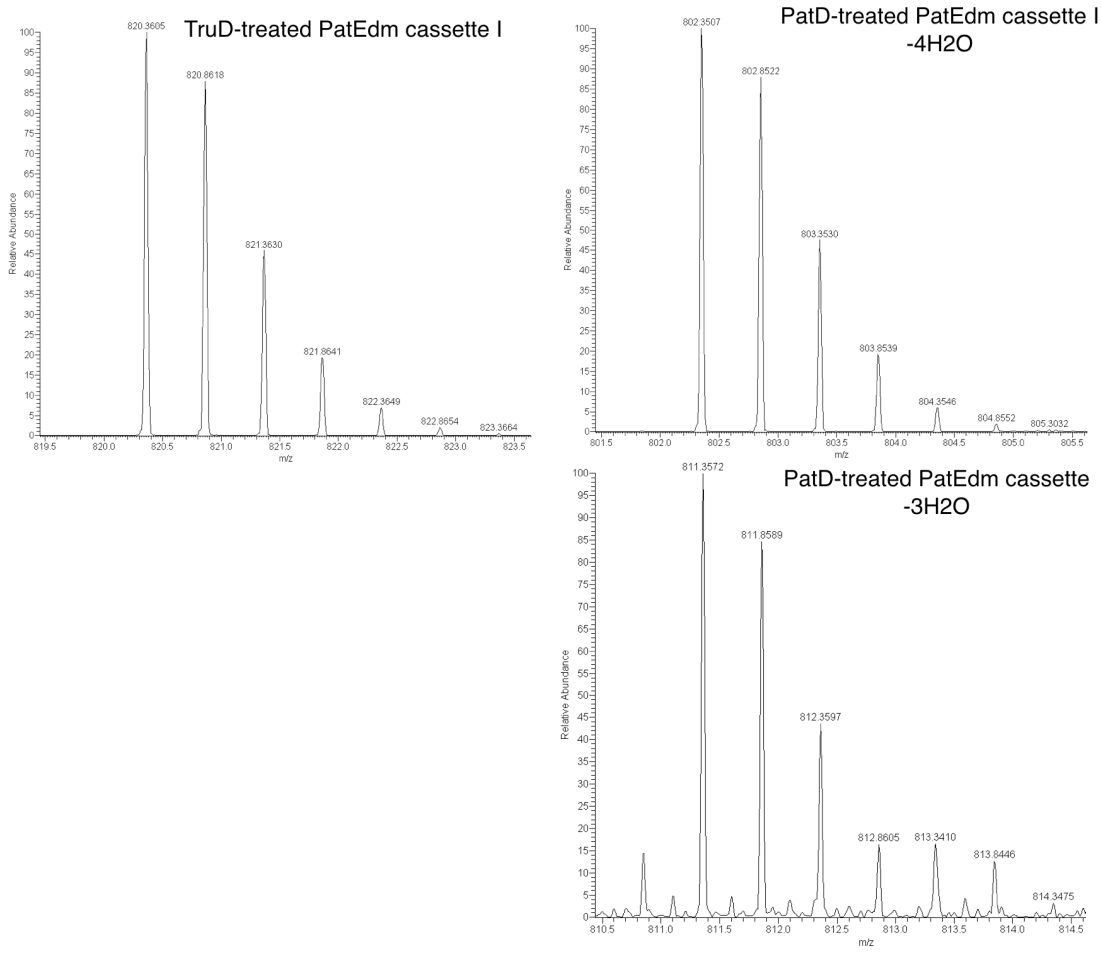
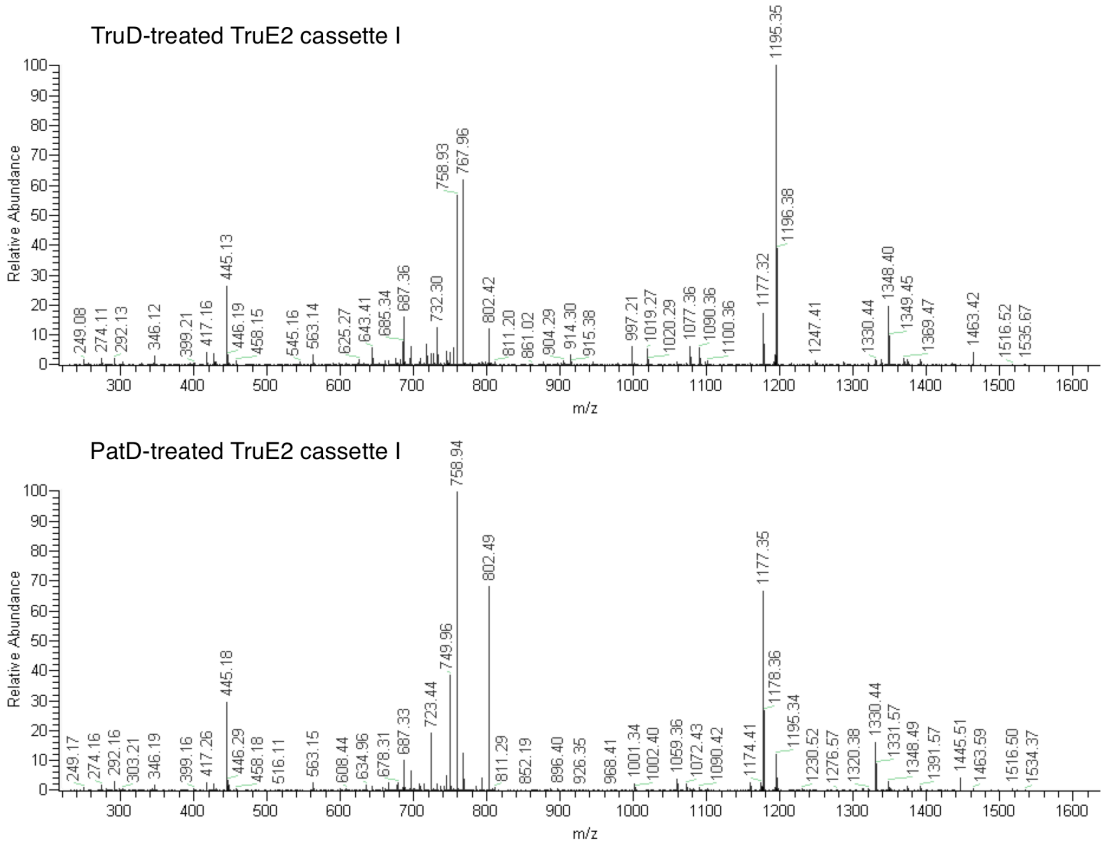
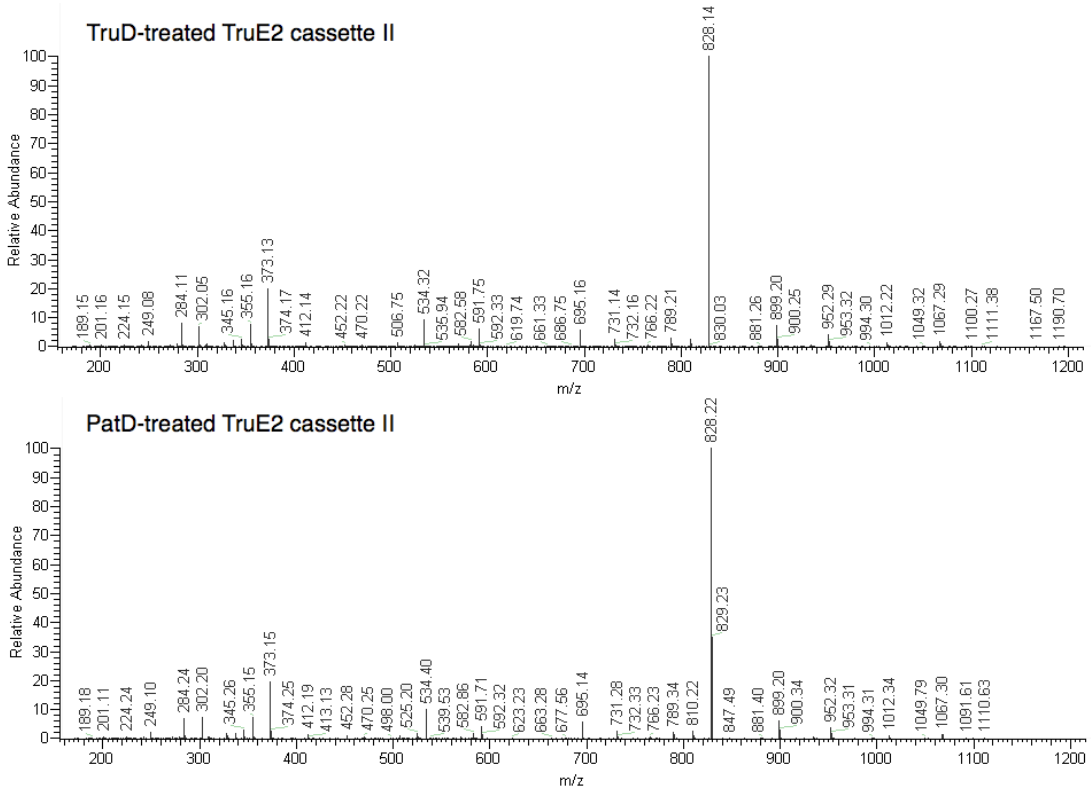

TruD-treated PatEdm cassette I
820.3605
820.8618
821.3630
821.8641
822.3649
822.8654
823.3664
PatD-treated PatEdm cassette I
-4H2O
802.3507
802.8522
803.3530
803.8539
804.3546
804.8552
805.3032
PatD-treated PatEdm cassette
-3H2O
811.3572
811.8589
812.3597
812.8605
813.3410
813.8446
814.3475
Relative Abundance
m/z

Figure S5. MS-MS of modified peptides and summary tables (a) MS-MS of TruD- and PatD-modified TruE2 cassette I (b) MS-MS of and TruD- and PatD-modified TruE2 cassette II (c) Tables summarizing MS-MS of TruE2 cassettes I and II (d) MS-MS of TruD- and PatD-modified PatEdm; one PatD-modified PatEdm cassette I is $-3H_2O$ an due to oxazoline ring-opening; fully modified is $-4H_2O$ (e) Tables summarizing MS-MS of TruD- and PatD-modified (both $-3H_2O$ and $-4H_2O$) PatEdm.

(a)



(b)


TruD-treated TruE2 cassette II
Relative Abundance
m/z
828.14
189.15
201.16
224.15
249.08
284.11
302.05
345.16
355.16
373.13
374.17
412.14
452.22
470.22
506.75
534.32
535.94
582.58
591.75
592.33
619.74
661.33
686.75
695.16
731.14
732.16
766.22
789.21
830.03
881.26
899.20
900.25
952.29
953.32
994.30
1012.22
1049.32
1067.29
1100.27
1111.38
1167.50
1190.70
PatD-treated TruE2 cassette II
Relative Abundance
m/z
828.22
829.23
189.18
201.11
224.24
249.10
284.24
302.20
345.26
355.15
373.15
374.25
412.19
413.13
452.28
470.25
498.00
525.20
534.40
539.53
582.86
591.71
592.32
623.23
663.28
677.56
695.14
731.28
732.33
766.23
789.34
810.22
847.49
881.40
899.20
900.34
952.32
953.31
994.31
1012.34
1049.79
1067.30
1091.61
1110.63

(c)

TruD-treated TruE2 cassette I	B-ions	B-ions (+2)	B°-ions	B°-ions (+2)
TF	249.08	-	-	-
TFP	346.12	-	-	-
TFPV	445.13	-	-	-
TFPVP	-	-	-	-
TFPVPT	643.41	-	-	-
TFPVPTV	-	-	-	-
TFPVPTV(TzIn)	-	-	-	-
TFPVPTV(TzIn)S	914.30	-	-	-
TFPVPTV(TzIn)SY	1077.36	-	-	-
TFPVPTV(TzIn)SYD	-	-	-	-
TFPVPTV(TzIn)SYDG	-	-	-	-
TFPVPTV(TzIn)SYDGV	1348.40	-	-	-
TFPVPTV(TzIn)SYDGVD	1463.43	-	-	-
TFPVPTV(TzIn)SYDGVDA	-	767.96	-	-
TFPVPTV(TzIn)SYDGVDAS	-	-	-	-

PatD-treated TruE2 cassette I	Y-ions	Y-ions (+2)	Y°-ions	Y°-ions (+2)
AS	-	-	-	-
DAS	292.12	-	-	-
VDAS	-	-	-	-
GVDAS	-	-	-	-
DGVDAS	563.13	-	-	-
YDGVDAS	-	-	-	-
SYDGVDAS	-	-	-	-
(TzIn)SYDGVDAS	-	-	-	-
V(TzIn)SYDGVDAS	997.20	-	-	-
TV(TzIn)SYDGVDAS	-	-	-	-
PTV(TzIn)SYDGVDAS	1195.35	1177.32	-	-
VPTV(TzIn)SYDGVDAS	-	-	-	-
PVPTV(TzIn)SYDGVDAS	-	-	-	-
FPVPTV(TzIn)SYDGVDAS	-	-	-	-
TFPVPTV(TzIn)SYDGVDAS	-	-	-	-

Double cleaved TruD-treated TruE2 cassette I	Mixed B, Y	Mixed B, Y (+2)	Mixed B°, Y°	Mixed B°, Y° (+2)
PTV(TzIn)SYDGVD	1019.26	-	-	-
PTV(TzIn)SYDGVDA	1091.37	-	-	-
PVPTV(TzIn)SYDGV	1100.35	-	-	-

PatD-treated TruE2 cassette I	B-ions	B-ions (+2)	B°-ions	B°-ions (+2)
TF	249.08	-	-	-
TFP	346.15	-	-	-
TFPV	445.18	-	-	-
TFPVP	-	-	-	-
TFPVP(OxzIn)	-	-	-	-
TFPVP(OxzIn)V	-	-	-	-
TFPVP(OxzIn)V(TzIn)	-	-	-	-
TFPVP(OxzIn)V(TzIn)S	896.28	-	-	-
TFPVP(OxzIn)V(TzIn)SY	1059.47	-	-	-
TFPVP(OxzIn)V(TzIn)SYD	-	-	-	-
TFPVP(OxzIn)V(TzIn)SYDG	-	-	-	-
TFPVP(OxzIn)V(TzIn)SYDGV	1330.49	-	-	-
TFPVP(OxzIn)V(TzIn)SYDGVD	1445.48	-	-	-
TFPVP(OxzIn)V(TzIn)SYDGVDA	-	758.98	-	-
TFPVP(OxzIn)V(TzIn)SYDGVDAS	-	-	-	-

Double cleaved TruD-treated TruE2 cassette I	Mixed B, Y	Mixed B, Y (+2)	Mixed B°, Y°	Mixed B°, Y° (+2)
P(OxzIn)V(TzIn)SYDGVD	1001.41	-	-	-
P(OxzIn)V(TzIn)SYDGVDA	1072.40	-	-	-
PVP(OxzIn)V(TzIn)SYDGV	1082.53	-	-	-

TruD-treated TruE2 cassette II	B-ions	B-ions (+2)	B°-ions	B°-ions (+2)
TS	189.15	-	-	-
TSI	302.05	-	284.11	-
TSIA	373.13	-	355.16	-
TSIAP	470.22	-	452.22	-
TSIAPF	-	-	-	-
TSIAPF(Tzln)	702.06	-	-	-
TSIAPF(Tzln)S	789.21	-	-	-
TSIAPF(Tzln)SY	952.29	-	-	-
TSIAPF(Tzln)SYD	1067.29	534.31	-	-
TSIAPF(Tzln)SYDD	-	-	-	591.75

	Y-ions	Y-ions (+2)	Y°-ions	Y°-ions (+2)
DD	249.08	-	-	-
YDD	412.14	-	-	-
SYDD	-	-	-	-
(Tzln)SYDD	-	-	-	-
F(Tzln)SYDD	731.15	-	-	-
PF(Tzln)SYDD	828.14	-	810.16	-
APF(Tzln)SYDD	899.20	-	-	-
IAPF(Tzln)SYDD	1012.22	-	-	-
SIAPF(Tzln)SYDD	-	-	-	-
TSIAPF(Tzln)SYDD	-	-	-	591.75

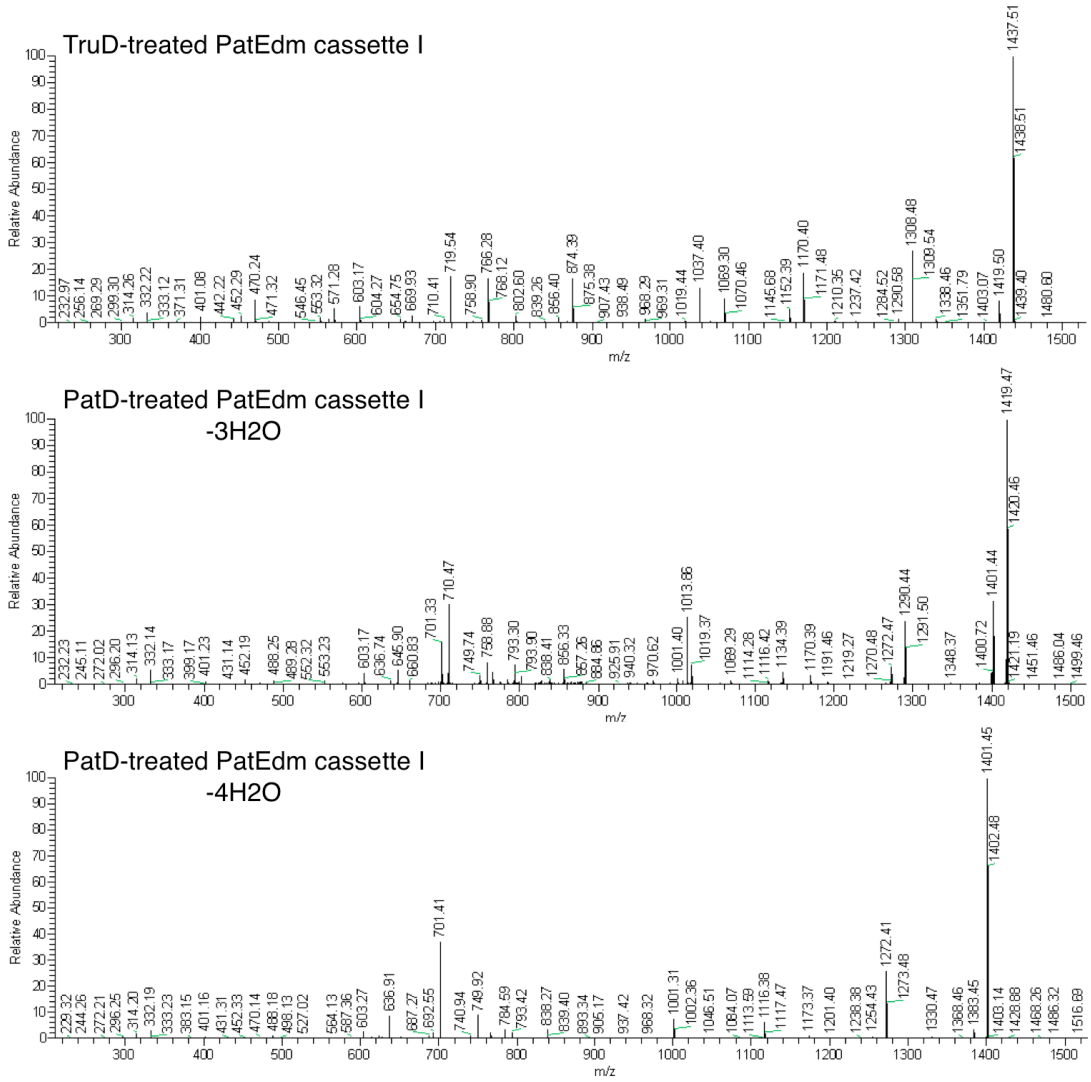
Double cleaved TruD-treated TruE2 cassette II PF(Tzln)SYD	Mixed B, Y	Mixed B, Y (+2)	Mixed B°, Y°	Mixed B°, Y° (+2)
	695.16	-	-	-

PatD-treated TruE2 cassette II	B-ions	B-ions (+2)	B°-ions	B°-ions (+2)
TS	189.18	-	-	-
TSI	302.20	-	284.24	-
TSIA	373.15	-	355.15	-
TSIAP	469.95	-	-	-
TSIAPF	-	-	-	-
TSIAPF(Tzln)	702.41	-	-	-
TSIAPF(Tzln)S	789.32	-	-	-
TSIAPF(Tzln)SY	952.32	-	-	-
TSIAPF(Tzln)SYD	1067.30	534.40	-	525.20
TSIAPF(Tzln)SYDD	-	-	-	591.71

	Y-ions	Y-ions (+2)	Y°-ions	Y°-ions (+2)
DD	249.10	-	-	-
YDD	412.19	-	-	-
SYDD	-	-	-	-
(Tzln)SYDD	-	-	-	-
F(Tzln)SYDD	731.28	-	-	-
PF(Tzln)SYDD	828.22	-	810.22	-
APF(Tzln)SYDD	899.20	-	-	-
IAPF(Tzln)SYDD	1012.34	-	-	-
SIAPF(Tzln)SYDD	-	-	-	-
TSIAPF(Tzln)SYDD	-	-	-	591.71

Double cleaved PatD-treated TruE2 cassette II PF(Tzln)SYD	Mixed B, Y	Mixed B, Y (+2)	Mixed B°, Y°	Mixed B°, Y° (+2)
	695.14	-	-	-

(d)


TruD-treated PatEdm cassette I
Relative Abundance
m/z
PatD-treated PatEdm cassette I
-3H2O
Relative Abundance
m/z
PatD-treated PatEdm cassette I
-4H2O
Relative Abundance
m/z

(e)

TruD-treated PatEdm cassette I	B-ions	B-ions (+2)	B°-ions	B°-ions (+2)
VT	-	-	-	-
VTA	-	-	-	-
VTA(TzIn)	-	-	-	-
VTA(TzIn)I	-	-	452.28 -	-
VTA(TzIn)IT	470.25 -	-	553.31 -	-
VTA(TzIn)ITF	571.29 -	-	-	-
VTA(TzIn)ITF(TzIn)	-	-	-	-
VTA(TzIn)ITF(TzIn)A	874.38 -	-	-	-
VTA(TzIn)ITF(TzIn)AY	1037.40 -	-	-	-
VTA(TzIn)ITF(TzIn)AYD	1152.38 -	-	-	-
VTA(TzIn)ITF(TzIn)AYDG	1210.33 -	-	-	-
VTA(TzIn)ITF(TzIn)AYDGV	1308.47	654.78	1290.58 -	-
VTA(TzIn)ITF(TzIn)AYDGVE	1437.52	719.53	1419.50 -	-
VTA(TzIn)ITF(TzIn)AYDGVEP	-	-	-	802.61
VTA(TzIn)ITF(TzIn)AYDGVEPS	-	-	-	-

PatD-treated PatEdm cassette I (-3H2O)	B-ions	B-ions (+2)	B°-ions	B°-ions (+2)
V(Oxzln)	-	-	-	-
V(Oxzln)A	-	-	-	-
V(Oxzln)A(TzIn)	-	-	-	-
V(Oxzln)A(TzIn)I	452.19 -	-	-	-
V(Oxzln)A(TzIn)IT	553.22 -	-	-	-
V(Oxzln)A(TzIn)ITF	-	-	-	-
V(Oxzln)A(TzIn)ITF(TzIn)	-	-	-	-
V(Oxzln)A(TzIn)ITF(TzIn)A	856.34 -	-	-	-
V(Oxzln)A(TzIn)ITF(TzIn)AY	1019.37 -	-	1001.41 -	-
V(Oxzln)A(TzIn)ITF(TzIn)AYD	-	-	1098.55 -	-
V(Oxzln)A(TzIn)ITF(TzIn)AYDG	-	-	-	-
V(Oxzln)A(TzIn)ITF(TzIn)AYDGV	1290.44	645.92	1272.49 -	-
V(Oxzln)A(TzIn)ITF(TzIn)AYDGVE	1419.47	710.48	1401.47	701.25
V(Oxzln)A(TzIn)ITF(TzIn)AYDGVEP	-	758.88 -	-	749.80
V(Oxzln)A(TzIn)ITF(TzIn)AYDGVEPS	-	-	-	793.29

Double cleaved TruD-treated PatEdm cassette I	Mixed B, Y	Mixed B, Y (+2)	Mixed B°, Y°	Mixed B°, Y° (+2)
DGVE	401.09 -	-	-	-
TA(TzIn)ITF(TzIn)AY	938.48 -	-	-	-
TF(TzIn)AYDGVE	968.29 -	-	-	-
TA(TzIn)ITF(TzIn)AYDGVE	1338.47	669.93 -	-	-

Double cleaved TruD-treated PatEdm cassette I -3H2O	Mixed B, Y	Mixed B, Y (+2)	Mixed B°, Y°	Mixed B°, Y° (+2)
DGVE	401.24 -	-	-	-
TA(TzIn)ITF(TzIn)AY	-	-	-	-
TF(TzIn)AYDGVE	-	-	-	-
TA(TzIn)ITF(TzIn)AYDGVE	-	-	-	-

	Y-ions	Y-ions (+2)	Y°-ions	Y°-ions (+2)
EP	-	-	-	-
EPS	332.22 -	-	314.25 -	-
VEPS	-	-	-	-
GVEPS	-	-	-	-
DGVEPS	603.17 -	-	-	-
YDGVEPS	766.27 -	-	-	-
AYDGVEPS	-	-	-	-
(TzIn)AYDGVEPS	-	-	-	-
F(TzIn)AYDGVEPS	1069.31 -	-	-	-
TF(TzIn)AYDGVEPS	1170.40 -	-	-	-
ITF(TzIn)AYDGVEPS	-	-	-	-
(TzIn)ITF(TzIn)AYDGVEPS	-	-	-	-
A(TzIn)ITF(TzIn)AYDGVEPS	-	-	-	-
TA(TzIn)ITF(TzIn)AYDGVEPS	-	-	-	-
VTA(TzIn)ITF(TzIn)AYDGVEPS	-	-	-	-

	Y-ions	Y-ions (+2)	Y°-ions	Y°-ions (+2)
EP	-	-	-	-
EPS	332.15 -	-	314.16 -	-
VEPS	-	-	-	-
GVEPS	488.23 -	-	-	-
DGVEPS	603.17 -	-	-	-
YDGVEPS	766.25 -	-	-	-
AYDGVEPS	-	-	-	-
(TzIn)AYDGVEPS	-	-	-	-
F(TzIn)AYDGVEPS	1069.29 -	-	-	-
TF(TzIn)AYDGVEPS	1170.38 -	-	-	-
ITF(TzIn)AYDGVEPS	-	-	-	-
(TzIn)ITF(TzIn)AYDGVEPS	-	-	-	-
A(TzIn)ITF(TzIn)AYDGVEPS	-	-	-	-
(Oxzln)A(TzIn)ITF(TzIn)AYDGVEPS	-	-	-	-
V(Oxzln)A(TzIn)ITF(TzIn)AYDGVEPS	-	-	-	-

TruD-treated PatEdm cassette I	B-ions	B-ions (+2)	B°-ions	B°-ions (+2)
VT	-	-	-	-
VTA	-	-	-	-
VTA(Tzln)	-	-	-	-
VTA(Tzln)I	-	-	-	-
VTA(Tzln)IT	470.25	-	452.28	-
VTA(Tzln)ITF	571.29	-	553.31	-
VTA(Tzln)ITF(Tzln)	-	-	-	-
VTA(Tzln)ITF(Tzln)A	874.38	-	-	-
VTA(Tzln)ITF(Tzln)AY	1037.40	-	-	-
VTA(Tzln)ITF(Tzln)AYD	1152.38	-	-	-
VTA(Tzln)ITF(Tzln)AYDG	1210.33	-	-	-
VTA(Tzln)ITF(Tzln)AYDGV	1308.47	654.78	1290.58	-
VTA(Tzln)ITF(Tzln)AYDGVE	1437.52	719.53	1419.50	-
VTA(Tzln)ITF(Tzln)AYDGVEP	-	-	-	-
VTA(Tzln)ITF(Tzln)AYDGVEPS	-	-	-	802.61

PatEdm cassette I (-4H2O)	B-ions	B-ions (+2)	B°-ions	B°-ions (+2)
V(Oxzln)	-	-	-	-
V(Oxzln)A	-	-	-	-
V(Oxzln)A(Tzln)	-	-	-	-
V(Oxzln)A(Tzln)I	-	-	-	-
V(Oxzln)A(Tzln)I(Oxzln)	-	-	-	-
V(Oxzln)A(Tzln)I(Oxzln)F	-	-	-	-
V(Oxzln)A(Tzln)I(Oxzln)F(Tzln)	-	-	-	-
V(Oxzln)A(Tzln)I(Oxzln)F(Tzln)A	-	-	-	-
V(Oxzln)A(Tzln)I(Oxzln)F(Tzln)AY	-	-	-	-
V(Oxzln)A(Tzln)I(Oxzln)F(Tzln)AYD	-	-	-	-
V(Oxzln)A(Tzln)I(Oxzln)F(Tzln)AYDG	-	-	-	-
V(Oxzln)A(Tzln)I(Oxzln)F(Tzln)AYDGV	-	-	-	-
V(Oxzln)A(Tzln)I(Oxzln)F(Tzln)AYDGVE	-	-	-	-
V(Oxzln)A(Tzln)I(Oxzln)F(Tzln)AYDGVEP	-	-	-	-
V(Oxzln)A(Tzln)I(Oxzln)F(Tzln)AYDGVEPS	-	-	-	-

Double cleaved TruD-treated PatEdm cassette I	Mixed B, Y	Mixed B, Y (+2)	Mixed B°, Y°	Mixed B°, Y° (+2)
DGVE	401.09	-	-	-
TA(Tzln)ITF(Tzln)AY	938.48	-	-	-
TF(Tzln)AYDGVE	968.29	-	-	-
TA(Tzln)ITF(Tzln)AYDGVE	1338.47	669.93	-	-

Double cleaved TruD-treated PatEdm cassette I -4H2O	Mixed B, Y	Mixed B, Y (+2)	Mixed B°, Y°	Mixed B°, Y° (+2)
DGVE	401.20	-	-	-
TA(Tzln)ITF(Tzln)AY	-	-	-	-
TF(Tzln)AYDGVE	-	-	-	-
TA(Tzln)ITF(Tzln)AYDGVE	-	-	-	-

EP	-	-	-	-
EPS	332.22	-	314.25	-
VEPS	-	-	-	-
GVEPS	-	-	-	-
DGVEPS	603.17	-	-	-
YDGVEPS	766.27	-	-	-
AYDGVEPS	-	-	-	-
(Tzln)AYDGVEPS	-	-	-	-
F(Tzln)AYDGVEPS	1069.31	-	-	-
TF(Tzln)AYDGVEPS	1170.40	-	-	-
ITF(Tzln)AYDGVEPS	-	-	-	-
(Tzln)ITF(Tzln)AYDGVEPS	-	-	-	-
A(Tzln)ITF(Tzln)AYDGVEPS	-	-	-	-
TA(Tzln)ITF(Tzln)AYDGVEPS	-	-	-	-
VTA(Tzln)ITF(Tzln)AYDGVEPS	-	-	-	-

EP	-	-	-	-
EPS	332.10	-	314.15	-
VEPS	-	-	-	-
GVEPS	488.17	-	470.15	-
DGVEPS	603.27	-	-	-
YDGVEPS	766.25	-	-	-
AYDGVEPS	-	-	-	-
(Tzln)AYDGVEPS	-	-	-	-
F(Tzln)AYDGVEPS	-	-	-	-
(Oxzln)F(Tzln)AYDGVEPS	-	-	-	-
I(Oxzln)F(Tzln)AYDGVEPS	-	-	-	-
(Tzln)I(Oxzln)F(Tzln)AYDGVEPS	-	-	-	-
A(Tzln)I(Oxzln)F(Tzln)AYDGVEPS	-	-	-	-
(Oxzln)A(Tzln)I(Oxzln)F(Tzln)AYDGVEPS	-	-	-	-
V(Oxzln)A(Tzln)I(Oxzln)F(Tzln)AYDGVEPS	-	-	-	-

Figure S6. Shown below are intact ESI analyses of TruE2, demonstrating that one dehydration catalyzed by PatD is an oxazoline ring. After treatment with mild base, PatD-modified TruE2 (a) is -$3H_2O$, while TruD-modified TruE2 (b) is -$2H_2O$. However, after treatment with mild acid, PatD-modified TruE2 (c) is -$2H_2O$, while the mass of TruD-modified TruE2 (d) is unchanged. Thus, two acid- and base-stable dehydrations are catalyzed by both PatD and TruD, based on this and MS-MS analysis these are thiazoline rings. Additionally, PatD catalyzes formation of a further base-stable, acid-labile dehydration, which by MS-MS analysis localized to Thr. Based on the chemical properties of this third dehydration, it can only be oxazoline.

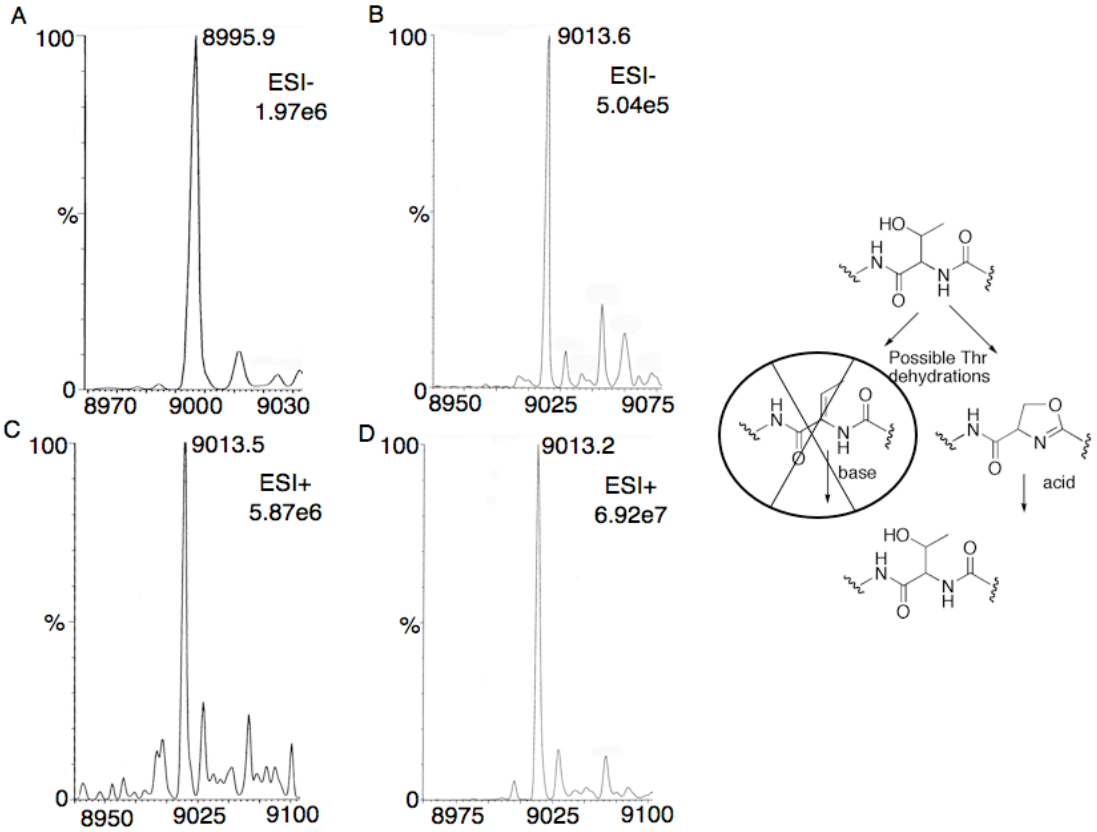


Figure S7. ESI-MS-(-) and SDS-PAGE analyses demonstrating modification of TruE5 by TruD and PatD. (a) From left-to-right unmodified TruE5, TruD-modified TruE5, and PatD-modified TruE5 are shown. The predominant product of both TruD and PatD by MS is a singly dehydrated species that most likely represents a peptide with the Cys in the first cassette cyclized. However, in both TruD and PatD reactions, some amount of fully modified TruE5 is present, corresponding to products that contain, respectively, one and two oxazoline rings in the first cassette. (b) SDS-PAGE gel showing modification of TruE5. From left to right, lanes are: (1) ladder (2) TruE5 standard (3) TruE2 + PatD (4) TruE2 + TruD (5) TruE2 + PatD (6) TruE2 + TruD (7) TruE5 + PatD (8) TruE5 + TruD.

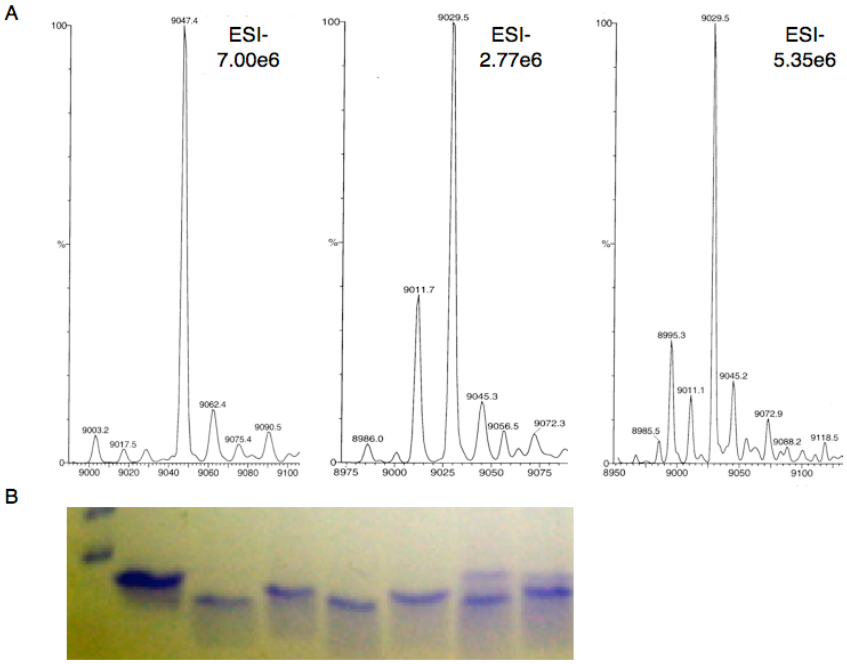
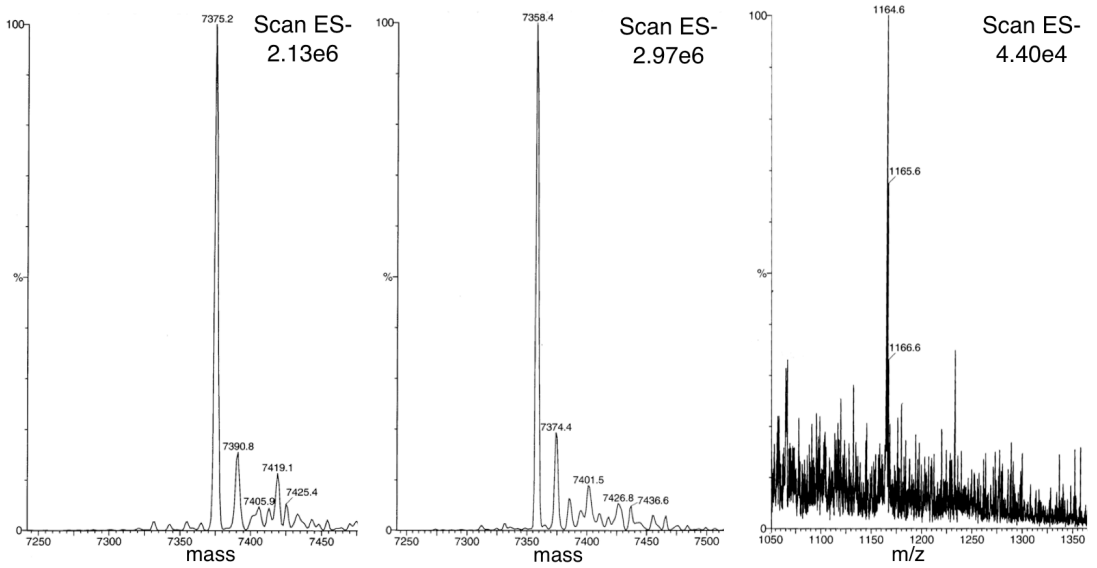


Figure S8. ESI-MS-(-) data demonstrating modification of TruE4 by a single dehydration event localized to the single product cassette that it contains. At left is intact analysis of unmodified TruE4; middle is intact analysis of TruD-modified TruE4, and at right is ESI-MS-(-) of PatD-treated TruE4,



digested with PatA, showing the cassette is -H2O (1164.6 Da).

Table S1. Table of primers used in this study.

	Primer Name	
1	PatE-F	AACATATGGACTTAAATTGACAGGCTTC
2	PatEα-R	GCATCACTTTTTGCGCTTATGATGGTGTGGAGCCATCTCATCACCACCACCATCACCATCACGCTTACGATGGTGAATAA
3	TruE-R	AATTCGGTACCTTAGTCGTCGTAAGAGCAGAG
4	TruD-F	TTCATGCAACCAACCGCCCTCCAAATTAAG
5	TruD-R	AACATATGGACTTAAATTGACAGGCTTC

Experimental

General methods. Isopropyl β-D-1-thiogalactopyranoside (IPTG), dithiothreitol (DTT), leupeptin, pepstatin, 4-(2-aminoethyl) benzenesulfonyl fluoride (AEBSF), and phenylmethanesulfonyl fluoride (PMSF) were purchased from ISC Bioexpress. Metal-free nitric acid (Optima, Fisher Scientific) was purchased from Fisher Scientific. Ultra-pure $MgCl_2$, β,γ-methylene-adenosine triphosphate (β,γ-methylene-ATP) were purchased from Sigma-Aldrich. γ-^{32}P-labeled ATP was purchased from Perkin-Elmer. Ni-NTA resin was purchased from Qiagen. ZipTip C18 pipette tips were purchased from Millipore. All expression vectors were purchased from Novagen. *Escherichia coli* strain DH5α was used for all cloning steps, while *E. coli* strain BL21(DE3)Star was used for all protein expressions. Protein concentrations were determined by absorbance at 280 nm; extinction coefficients were predicted as described elsewhere.[16]

Gene cloning. Genes were obtained from ascidian symbiont metagenomes as previously described.[2] *patD* was cloned into pET15b with an N-terminal His-tag. *TruD* was cloned into pCDF Duet-1. Both *patD* and *truD* have an internal PstI site located 700 bp into the coding region and are 99% identical in this N-terminal region. Thus, to construct *truD*, the His-tagged N-terminus of *patD* was obtained by PstI / NdeI digest and ligated into homologous restriction sites in the *truD* construct. *TruE2* and *truE4* were cloned untagged into pRSF-Duet vector between the Nde1 and Kpn1 restriction sites. The N-terminal His tag was then moved from pET28b into the pRSF vectors through the restriction sites MluI and NdeI. Clones were verified by restriction digests and DNA sequencing. *patE*α was modified from a previously described *patE* variant, *patE2*. *TruE5* was modified from *truE2* using the Quikchange site directed mutagenesis kit (Stratagene). The *patE2* gene was amplified using the primers PatEf and PatEαr, subcloned into pCR2.1-TOPO (Invitrogen), and then cloned into pRSFDuet-1. PatA and PatEdm were cloned as previously described.[17]

Expression of PatD and TruD. Seed cultures were inoculated into 1 L of LB and grown in a Fernbach at 30°C with shaking at 225 rpm until the cells reached 0.4 OD_{600}. Expression was induced

with 0.1 mM IPTG at 15°C and left overnight. Cells were pelleted by centrifugation and resuspended in 40 mL of lysis buffer (1 M NaCl, 10 mM imidazole, 50 mM HEPES pH 7.5, 1 μg/mL leupeptin, 1 μg/mL pepstatin, 1 mM AEBSF). Cells were lysed using a Vibracell sonicator and centrifuged. Filtered lysates were applied to a gravity column containing 10mL of wet Ni-NTA resin. The column was then washed with 10 column volumes of lysis buffer, 10 column volumes of wash buffer (500 mM NaCl, 25 mM Imidazole pH 7.8), and the protein was eluted with elution buffer (750 mM NaCl, 250 mM Imidazole pH 7.8). The purified protein was dialyzed twice against dialysis buffer (200 mM NaCl, 25 mM HEPES pH 7.5) at 4°C, and then against dialysis buffer w/ 10% glycerol. Dialyzed protein was aliquoted, flash frozen with liquid nitrogen, and stored at -80°C.

Expression of precursor peptides. TruE2-expressing cells were grown to density in a manner identical to that for PatD and TruD. Expression of TruE2 was induced by addition of IPTG up to a final concentration of 1 mM followed by overnight growth at 37°C. Cell pellets were resuspended in lysis buffer at a concentration of 4 mL per gram of cell paste. After centrifugation, supernatants were discarded, while the pellets were resuspended in 50 mL of buffer B (8 M urea, 0.1 M NaH_2PO_4, 10 mM tris pH 8.0) and then sonicated to affect further resolubilization. The resuspended material was again centrifuged, and filtered lysates were applied to a Ni-NTA column, which was then washed with 2 x 25 mL of buffer C (8 M urea, 0.1 M NaH_2PO_4, 10 mM tris pH 6.3), 4 x 12.5 mL of buffer D (8 M urea, 0.1 M NaH_2PO_4, 10 mM tris pH 5.9), and eluted with 4 x 12.5 mL of buffer E (8 M urea, 0.1 M NaH_2PO_4, 10 mM tris pH 4.5). The eluents were dialyzed twice against dialysis buffer (200 mM NaCl, 100 mM proline, 1 mM DTT, 10% glycerol, 20 mM HEPES pH 8.0). The eluents were then combined, aliquoted, flash frozen, and stored at -80°C.

Expression of TruE4 was performed in a manner identical to that of TruE2, except that owing to the larger amount of insoluble material obtained, pellets were resuspended in 100 mL of buffer B, and the Ni-NTA column was washed with 100 mL of buffer C.

Expression of PatEα was performed in a manner identical with that of TruE2, with the following exceptions: the lysis buffer used consisted of 500 mM NaCl, 20 mM NaH_2PO_4 pH 7.8. Pellets were

resuspended in 8 M urea rather than buffer B, and the Ni-NTA column was washed using 500 mM NaCl, 25 mM imidazole, and protein was eluted using 3 x 10 mL of elution buffer (750 mM NaCl, 250 mM imidazole pH 7.8). Eluents were combined and dialyzed twice against dialysis buffer (500 mM NaCl, 2 mM DTT, 25 mM HEPES pH 7.8).

Expression of TruE5 was performed in a manner identical with that of TruE2.

PatA and PatEdm were expressed as previously described.[17]

Enzyme reactions. Reaction mixtures were incubated at 34°C in an MJ research minicycler for varying amounts of time. Enzyme, precursor peptide, and ATP concentrations varied, and are described below. The following additives were present in standard reactions but were varied in early optimization experiments: 40 mM tris pH 8.0, 8 mM DTT, and 4 mM $MgCl_2$.

Enzyme reactions generally contained the optimized additive mixture and 0.6 μM PatD or TruD, 8 μM TruE2, or 12 μM TruE4, or 42 μM PatEα, or 22 μM PatEdm, and 0.8 mM ATP. Reactions were run using varying times, from 15 min to 27 h. To confirm modification, 10 μL of the reaction mixture was removed and analyzed by SDS-PAGE. At minimum, at least three separate experiments were performed for each enzyme-substrate concentration.

For reactions requiring cleavage by PatA, mixtures as described above were incubated for 27 h, with and without ATP and with and without PatD / TruD. PatA was then added to 1.7 μM final concentration followed by incubation for a further 17 h. After completion of PatA-containing reactions, reaction mixtures were frozen at -80°C until analyzed by MALDI-TOF, FT-MS, and / or ESI-MS.

SDS-PAGE assays. 18% acrylamide gels were used for all assays. Prior to electrophoresis, samples were brought up in 1X SDS sample buffer diluted from 6X SDS sample buffer and then boiled for 3 min.[18] After electrophoresis, gels were placed in boiled fixing solution (53% H_2O, 40% ethanol, 7% acetic acid), incubated 10-20 min with gentle rocking, and then placed in boiled stain (0.02% w/v Coomassie R250 in 85% H_2O, 10% acetic acid, 5% ethanol), and then destained in (85% H_2O, 10%

acetic acid, 5% ethanol) for several h, and then photographed. For TruE4 inhibition experiments, gels were imaged using a Li-Cor Odyssey infrared scanner.

Mass spectrometry analysis. Samples for MALDI-TOF were prepared by desalting with ZipTip C18 pipette tips according to the manufacturer's instructions. Desalted peptide samples were mixed 1:1 with α-cyano-4-hydroxycinnamic acid resin (10 mg/mL CHCA in 50:50 H_2O:methanol w/ 0.1% trifluoroacetic acid), and spotted. MALDI was performed using a Micromass MALDI micro MX (Waters) instrument using an automated targeting protocol. LC-FTMS, ESI-MS(+), and ESI-MS(-) were run at the University of Utah Mass Spectrometry and Proteomics Core Facility. FT was performed using an LTQ FT Ultra Hybrid Mass Spectrometer (Thermo Scientific). ESI used a Micromass Quattro-II (Waters).

CHAPTER 6

MARINE MOLECULAR MACHINES: HETEROCYCLIZATION IN CYANOBACTIN BIOSYNTHESIS

Manuscript reproduced with permission from:

McIntosh, J. A., Schmidt, E. W. (2010) Marine molecular machines: heterocyclization in cyanobactin biosynthesis, *ChemBioChem 11 (10)*, 1413-1421.

Note: my contribution to this paper was in planning and designing experiments, analyzing the mass spectrometry and kinetic data, and also in carrying out all of the experiments described with the exception of the ICP-OES data generated by Pam Smith and Dennis Winge and mass spectrometry data generated by the University of Utah Mass Spectrometry and Proteomics core facility.

DOI: 10.1002/cbic.201000196

Marine Molecular Machines: Heterocyclization in Cyanobactin Biosynthesis

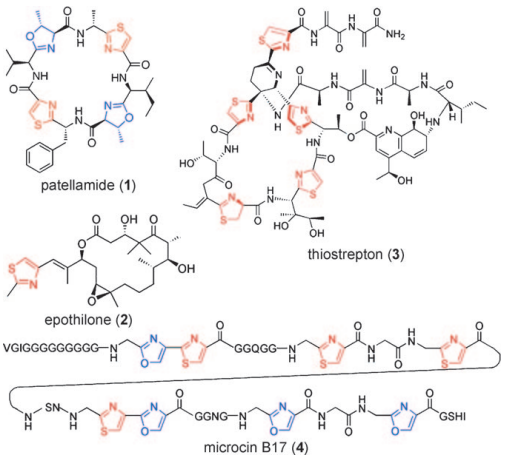
John A. McIntosh and Eric W. Schmidt*[a]

"A heterocycle sounds like a wonderful thing to ride, especially with someone you love."—George Carlin[1]

Natural products that contain amino-acid-derived (Cys, Ser, Thr) heterocycles are ubiquitous in nature, yet key aspects of their biosynthesis remain undefined. Cyanobactins are heterocyclic ribosomal peptide natural products from cyanobacteria, including symbiotic bacteria living with marine ascidians. In contrast to other ribosomal peptide heterocyclases that have been studied, the cyanobactin heterocyclase is a single protein that does not require an oxidase enzyme. Using this simplifying condition, we provide new evidence to support the hypothesis that these enzymes are molecular machines that use ATP in a product binding or orientation cycle. Further, we show that both protease inhibitors and ATP analogues inhibit heterocyclization and define the order of biochemical steps in the cyanobactin biosynthetic pathway. The cyanobactin pathway enzymes, PatD and TruD, are thiazoline and oxazoline synthetases.

Introduction

The heterocyclic thiazole, thiazoline, oxazole, and oxazoline motifs are commonly found in natural products that possess diverse biological activities. These Cys-, Ser-, and Thr-derived heterocycles are present in approved drugs, as well as in drug leads and in toxins produced by human pathogens.[2–5] For example, in the ribosomal peptide group microcin B17 is a DNA gyrase inhibitor produced by *Escherichia coli*,[6] thiostrepton and relatives are potent antibiotics,[7] and cyanobactins are ubiquitous bioactive peptides from cyanobacteria.[8] More recently, heterocyclic relatives of the cyanobactin/microcin B17 group have been shown to occur commonly in diverse bacteria.[5,9] In the nonribosomal peptide group, an epothilone derivative is an FDA-approved anticancer agent, bacitracin is used as an antibiotic, and certain siderophores are key virulence factors.[2,3] In addition, the heterocyclic motif itself is potentially bioactive depending upon the context: thiazol(in)e and oxazol(in)e can interact with nucleic acid, protein, or metal ligands.[3] Thus, routes to their enzymatic synthesis are of keen interest.



Despite extensive and groundbreaking studies of heterocyclization enzymes in both ribosomal and nonribosomal peptide natural products,[10,11] their chemical mechanism(s) remain unknown.[12] Previous biochemical studies of biosynthesis in heterocycle-containing ribosomal peptides involve a complex of three proteins whose activities are intertwined and difficult to separate. In particular, ATPase, oxidase, and zinc-binding domains are invariant requirements for heterocycle biosynthesis in the cases of microcin B17 and streptolysin S.[5,13] These enzymes modify microcin and streptolysin precursor peptides in multiple positions by synthesizing oxidized heterocycles, thiazoles and oxazoles, and not their unoxidized presumed precursors, thiazolines and oxazolines. Of the apparent enzymatic activities (oxidase, ATPase, and heterocyclase), none has been observed in isolation from the others, and while ATPase and oxidase activities are readily attributable to specific proteins, their respective roles in enzymatic heterocyclization are unclear. An oxidase is required, and this explains the lack of observed thiazoline or oxazoline either in vitro or in vivo. It has been proposed that ATP consumption fuels a molecular machine driving heterocyclization, although a direct effect on catalysis has

[a] *J. A. McIntosh, Prof. E. W. Schmidt*
Department of Medicinal Chemistry, College of Pharmacy
University of Utah
Salt Lake City, UT 84112 (USA)
Fax: (+1) 801-585-9119
E-mail: ews1@utah.edu

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201000196.

not been ruled out. These conserved protein domains also appear to be present in thiopeptide biosynthesis,[7] though to our knowledge, they have not yet been studied in vitro. Heterocyclization has also been studied in nonribosomal peptide systems,[10,14] in which the enzymes and resulting chemical mechanisms (excepting the oxidase) are nonhomologous to the ribosomal peptide group.

Here we present results regarding the biosynthesis of heterocycles among the cyanobactins, which are a group of circular, heterocyclic, ribosomally derived peptides from cyanobacteria, including the marine animal symbionts, *Prochloron* spp.[8] Two pathways to *Prochloron*-derived cyanobactins exist: the patellamide (*pat*) pathway, the members of which contain heterocycles derived from Cys, Ser, and Thr, and the trunkamide pathway (*tru*), the members of which contain heterocycles derived only from Cys. Heterocycle oxidation is somewhat variable, with both oxidized (thiazole and oxazole) and unoxidized (thiazoline and oxazoline) heterocycles present in both families. By contrast, the products of the microcin pathway, and presumably the streptolysin pathways, are always oxidized, although the latter has not been completely defined.

The biosynthetic gene clusters of the patellamide and trunkamide pathways have been previously characterized, as have the biochemical basis of circularization and certain aspects of heterocyclization.[12,15] In brief, PatE and TruE are precursor peptides, and each encodes two natural products on a single short peptide. Importantly, PatE and TruE both contain a leader peptide sequence and "enzyme recognition" sequences that flank the N and C termini of the product coding cassettes (Figure S1 in the Supporting Information). PatD and TruD are heterocyclases that operate regioselectively in vivo to modify Cys, Ser, and Thr in the case of PatD, but only Cys in the case of TruD. PatA or TruA proteases cleave N-terminal recognition sequences of cassettes 1 and 2,[15] though it was not clear prior to this study whether PatA/TruA acted prior to or after heterocyclization. Lastly, PatG or TruG proteases cleave the C-terminal recognition sequences in tandem with macrocyclization (Figure S2).[15] All three of these enzyme groups are capable of modifying a diverse set of cassette mutants to yield libraries of natural products.[16]

TruD, which in vivo heterocyclizes only Cys residues, is a didomain protein with an N-terminal region that bears distant similarity to heterocyclase proteins in the streptolysin and microcin systems. The C-terminal region is very distantly similar to the protein proposed to be responsible for peptide binding and ATPase activity in other heterocyclization systems. PatD is >99% identical to TruD in the N-terminal domain, but only 77% identical in the C-terminal domain (Figure 1). No oxidase

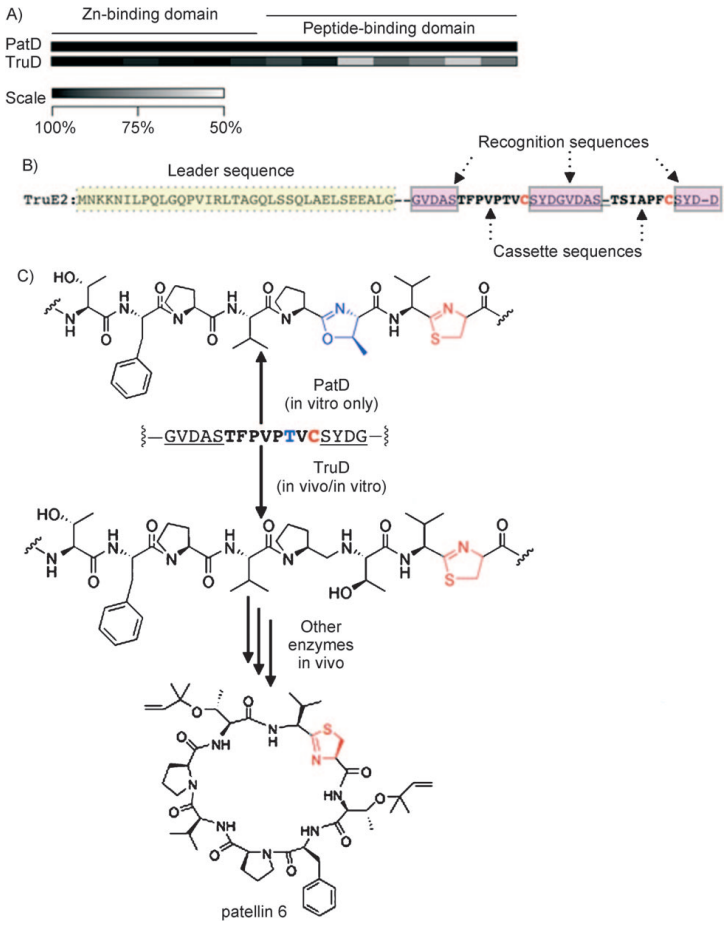


Figure 1. A) Shown are alignments between PatD and TruD. Darker regions indicate regions of higher identity. B) Sequence of TruE2 precursor peptide is shown, with naturally heterocyclized residues highlighted in red. C) A zoomed-in view of the C-terminal cassette in TruE2. In vitro, PatD modifies one Thr and one Cys in this cassette, while TruD modifies one Cys both in vitro and in vivo. In nature, in combination with other biosynthetic enzymes the TruD product shown is converted to the prenylated, heterocyclic natural product patellin 6.

domain is present in TruD or in the *tru* pathway, and unlike the microcin or streptolysin cases, oxidation is not required to reconstitute heterocyclase activity. Thus, TruD is a single protein that fully reconstitutes heterocyclization activity in a regioselective manner. The relative simplicity of this system in comparison to other studied heterocyclases enabled us to obtain new insights into the function of this important enzyme family. Here, we present evidence that strongly supports the molecular-machine hypothesis regarding the role of ATP. We also determined the order of enzymatic steps en route to cyanobactins, and describe inhibitors of heterocyclization.

Results and Discussion

Metal and cofactor requirements of PatD and TruD

The enzymes PatD and TruD, and substrates TruE2 and TruE4 were cloned as described elsewhere.[17] Two point mutations were found in *patD* in these studies, but they lacked any apparent functional consequence. In addition, *patD* was used as a template to clone *truD*; this ensures that both PatD and TruD are 100% identical in their N-terminal catalytic domains.

As previously described, when incubated with substrates, TruD and PatD are fully competent heterocyclase enzymes.[17] A robust SDS-PAGE assay was developed in which heterocyclization can clearly be tracked by mobility shift (Figure S3), as described elsewhere.[17] In general, TruD products migrate more rapidly by SDS-PAGE analysis than unreacted precursor, while PatD products migrate more rapidly still. This assay allowed enzyme requirements and timing to be rigorously defined.

All enzyme reactions described in this work, including determination of enzyme requirements, substrate and product measurements, and kinetic analysis, were performed at minimum in triplicate in independent runs. ATP, $MgCl_2$, and DTT were found to be necessary for the heterocyclization reaction with both PatD and TruD. The minimum Mg^{2+} concentration that could support catalysis was 1 mM, which is somewhat lower than the reported Mg^{2+} requirements of microcin B17 and streptolysin S synthetases (2–20 mM Mg).[5,11] Ultrapure $MgCl_2$ was used in some experiments and was found to support catalysis. Additionally, other additives (Tris, ATP, and DTT) could be passed through Chelex resin without inhibiting catalysis; this indicates that traces of other metals were not required for catalysis.

Like McbB from the microcin B17 pathway,[18] PatD and TruD are both strongly associated with zinc even after extensive dialysis, as indicated by ICP-OES experiments. The apparent binding stoichiometry of the enzyme-metal complex was found to be roughly 1 mol of zinc per mol of enzyme. The role of this zinc is unknown but has been proposed to be structural in the microcin B17 context.[18] No other strong associations of metals with enzyme were found using ICP-OES. Taken together, this indicates that zinc and magnesium are the sole metal cofactors required to catalyze the reaction.

The fate of ATP was probed by HPLC analysis. During the course of the reaction, ATP was shown to be hydrolyzed into ADP, as was found for microcin B17 synthetase (Figure S4).[19] Overall, these requirements are quite similar to those defined in microcin B17 biosynthesis; this indicates that the enzymes could function in a similar manner despite their nearly complete lack of protein sequence similarity. The major differences between microcin B17 biosynthetic enzymes and PatD/TruD are 1) the lack of a requirement for an oxdiase domain in our system, and 2) the presence of putative heterocyclase and peptide-binding domains on a single polypeptide, whereas in the microcin B17 system these domains exist as separate polypeptides.

Order of heterocyclization events in TruE2

Enzymatic reactions with TruE2 and TruD, or PatD generally went to completion within 2 h. Time-course experiments from 15 min to 24 h revealed that Cys reactivity was fast, with TruD-TruE2 reactions being complete within 1 hour. Catalysis of the third dehydration event (Thr cyclization) by using PatD was slower and required up to 2 h for complete modification (Figure S5). No further reaction was observed with extended incubation periods, even with batch addition of further enzyme.

In time-course experiments with TruD, TruE2 was observed to proceed directly to the doubly dehydrated product. We observed an appreciable accumulation of singly dehydrated product only by supplying inadequate amounts of ATP, and even under those conditions the amount of doubly dehydrated product was still greater (Figure S6). These experiments are most consistent with the idea that the substrate can dissociate from the enzyme between heterocyclizations, but the singly heterocyclized product is a much faster reactant than the unmodified precursor peptide.

By contrast, heterocyclization of TruE2 to afford a third heterocycle (oxazoline) by PatD was slower, and intermediates could be captured and readily observed by both SDS-PAGE and mass spectrometry. Quite clearly then, the precursor peptide leaves the enzyme after the second heterocyclization and before the third heterocyclization. One of the most convincing pieces of evidence was found upon incubation of PatD with TruE2. A sample was taken prior to completion of the PatD reaction and showed three bands (listed in order of increasing mobility): 1) unmodified TruE2, 2) a band consistent with two thiazolines, and 3) the fully modified TruE2.

Relative rates have also been determined in microcin B17 biosynthesis, and the results observed in that system are quite similar to what we present here: in both cases, the reaction to form oxygen-containing heterocycles is slower, as would be expected from the reduced nucleophilicity of hydroxy substituents in comparison to thiol.[19]

Timing of biosynthetic steps in cyanobactin synthesis

Previous work has shown that PatA and PatG enzymes, which cleave and circularize PatE, accept broadly different substrates.[15,16] A mystery has been how these enzymes produce only the natural products and not other derivatives. One idea, consistent with work on microcin B17 and previous cyanobac-

tin coexpression experiments, is that the enzymes could form a complex that would sequester substrates. With numerous conditions, however, we could not observe any requirement for complex formation. For example, PatA, PatG, and PatD are all competent catalysts without the addition of other enzymes, and addition of multiple enzymes does not appreciably increase the reaction rate of single steps. Pull-down experiments using various Pat proteins as bait in different conditions were unsuccessful, as were experiments involving coexpression. Thus, there is no evidence that protein complexes are required for cyanobactin biosynthesis.

Nonetheless, when PatA is co-incubated with TruD or PatD and TruE2 or other precursor peptides, we do not detect any PatA cleavage fragments that lack heterocycle modifications. This observation was surprising given that PatA is capable of cleaving unmodified precursor peptides in the absence of PatD or TruD.[15] To further probe this issue, PatA was added to the reaction mixture and allowed to react prior to the addition of PatD or TruD, and vice versa. Under no condition could we observe predicted dehydration products if PatA was allowed to react first; this would cleave the leader sequence, whereas the expected products were obtained if PatD or TruD were added first or if the A/D combinations were co-incubated. Therefore, it appears that the leader sequence, which PatA cleaves, is required in cis for PatD/TruD modification, as observed in the case of microcin B17.[20,21] In other words, the products of PatD/TruD are substrates for PatA, but the reverse is not true. Thus, the fidelity of cyanobactin synthesis is most likely encoded at the level of substrate recognition, and not at the level of protein complexes. Through these experiments, the order of catalytic events in cyanobactin synthesis was shown to be heterocyclization, followed by linearization and N–C circularization. The relative preference of PatD or TruD for heterocyclizing certain positions in competition with the rate of PatA leader-sequence cleavage dictates the structures of the natural products. The precise timing of O-prenylation (as observed in the *tru* pathway) and heterocycle oxidation (as obtained in the *pat* pathways) remains unclear, but oxidation must take place after heterocyclization, and prenylation also seems likely to occur after heterocyclization.

In comparison, these additional protease events are not tied to the microcin B17 gene cluster,[22] nor are microcin B17 or streptolysin S further modified beyond heterocycle synthesis, with the exception of (as yet undefined in the case of streptolysin S) leader sequence cleavage.

Kinetic analysis of TruD

TruD was used for all rate experiments because it is more readily purified compared with PatD, and its reactivity (Cys-only in natural precursor peptides) is substantially simpler than that of PatD. Rates of reaction were analyzed using varying ATP, TruE2, and TruD concentrations. With highly purified TruD, the background hydrolysis of ATP due to enzyme in absence of substrate was virtually nonexistent. The background hydrolysis with TruE2 alone is quite small, but is measurable. By contrast, when substrate and enzyme are co-incubated, ATP hydrolysis occurs at a robust pace. Reactions proceeded at an essentially linear rate for the conditions attempted in the first 40 min. Therefore, reactions were sampled at 0, 20, and 40 min and analyzed by SDS-PAGE for TruE2 turnover and HPLC for ADP formation. Results of this analysis were fitted by using the Solver function in Excel to estimate kinetic constants.

By using variable TruE2 concentrations, from 2.5 to 7.4 μM, TruE2 K_m was estimated as ~1 μM. The apparent k_{cat} for ATP hydrolysis under these conditions was 2.6 min^{-1} (Figures S7 and S8). This study was not designed to specifically measure the k_{cat} of the enzyme for TruE2, which is a complex problem. For example, the clear dissociation of the enzyme-substrate complex—especially with oxazoline synthesis in which the rate of reaction differs—indicates that different intermediates have different k_{cat} values, and possibly even different K_m values. However, it was possible to estimate a turnover number using saturating substrate and enzyme conditions. Under standard reaction conditions with 7.4 μM TruE2 and 140 nM TruD, the reaction proceeded to ~90% completion at $t = 60$ min (Figure 2). Based upon this experiment, a TruE2 turnover number of

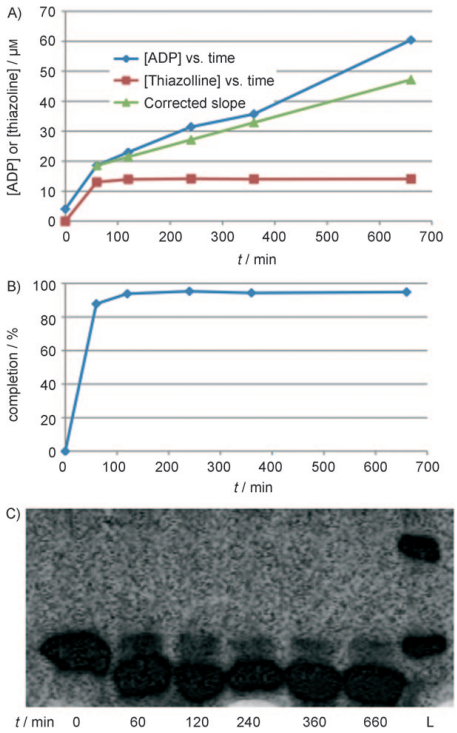


Figure 2. Stoichiometry of heterocycle formation. A) Rates of ADP formation and thiazoline synthesis are overlaid; corrected slope denotes the rate of ATP hydrolysis when corrected for the background hydrolysis. B) %-completion of the heterocyclization reaction as determined by SDS-PAGE gel densitometry. C) SDS-PAGE gel was used to derive the concentration of thiazoline and %-completion.

~0.8 enz^{-1} min^{-1}, or a heterocycle turnover number of ~1.6 enz^{-1} min^{-1}, could be calculated. This is not a true k_{cat} but gives an estimate of the speed of the enzyme under the reaction conditions.

Using variable ATP from 200 to 800 μM, the K_m for ATP was estimated as 300 μM. The calculated k_{cat} value was 2.4 min^{-1}, which is in excellent agreement for that calculated using variable TruE2 concentration. These rates also scaled precisely as enzyme concentration was doubled. At 104 nM TruD, TruE2 reactions were 1.5 times faster than at 69 nM TruD; at 140 nM TruD, reactions were 2.1 times faster than 69 nM TruD.

The kinetic constants described above are strikingly similar to those reported for microcin B17 synthetase,[11,19] despite the greatly different experimental conditions and the absence of both a protein complex and an oxidase. In particular, the relative K_m (1 μM) measured here using kinetic methods is similar to that reported for microcin B17 (2.3 μM),[11] but much higher than that reported for the streptolysin S leader peptide (6.7 nM), which was measured using surface-plasmon resonance.[23] Overall, these studies indicate that the enzymes function similarly, and that results reported here likely are applicable to the distant protein relatives previously studied.

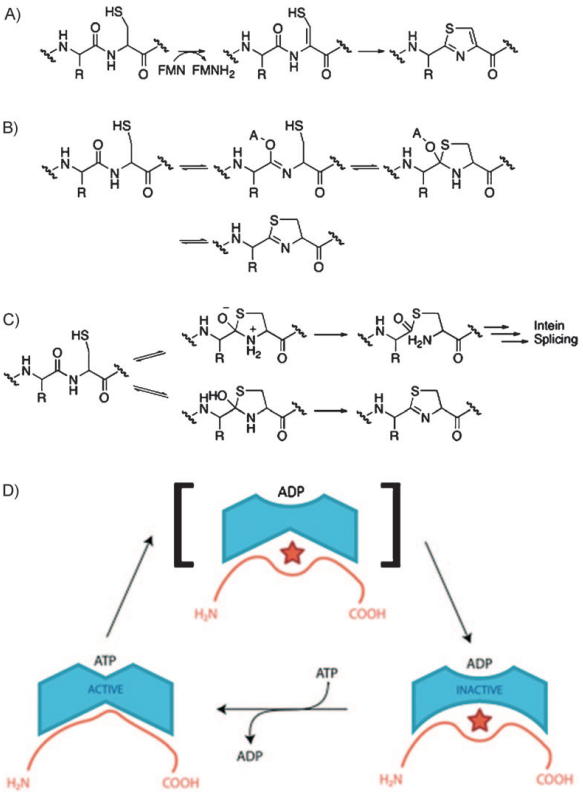
Inhibition of heterocyclization

In the course of our studies, two types of inhibitors were shown to slow heterocyclization; these are protease inhibitors AEBSF and PMSF, and a nonhydrolyzable ATP analogue, β,γ-methylene ATP. As found in the case of microcin B17,[19] addition of the aforementioned nonhydrolyzable ATP analogue to reaction mixtures inhibited heterocycle formation (Figure S11). By contrast, the use of irreversibly acting protease inhibitors could not be anticipated, and this was discovered in a serendipitous manner. In studies of possible AEBSF inhibition of PatA protease, it was observed that the protease cleavage was not blocked, but strikingly heterocyclization was inhibited. This led to further studies demonstrating that AEBSF is a covalent inhibitor of heterocyclization; these studies are described below.

An inhibition curve using 1–10 mM of AEBSF and monitoring TruE2 by SDS-PAGE showed that inhibition was complete at 10 mM for both PatD and TruD (Figure S12). Inhibition was monitored with an appropriate set of controls; this ensured that specific inhibition was responsible and not an effect of solvent or conditions. In principle, because AEBSF could act competitively or allosterically instead of irreversibly, reactions were also incubated with PMSF. At a concentration of 1 mM, the reaction was strongly inhibited by PMSF. These data implicate direct nucleophilic displacement of fluoride in AEBSF and PMSF as the mechanism of inhibition. Moreover, a kinetic analysis was performed in which increasing concentrations of AEBSF were applied to enzyme reactions with TruD; this analysis measured ADP production from ATP (Figure S12). The kinetic profile of reactions containing AEBSF strongly suggests that AEBSF does indeed act as an irreversible inhibitor; that is, ATP hydrolysis is not initially prevented by AEBSF, but as time goes on, virtually all ATP hydrolysis is halted.

Role of ATP in heterocyclization

Upon beginning our study on this reaction, there were two plausible, extant hypotheses regarding the role of ATP in heterocyclization of ribosomal peptides. First, is the idea that ATP could be a necessary cofactor in activating the electrophilic carbonyl carbon for attack by sulfur or oxygen (Scheme 1 B).[11] An alternative hypothesis holds that ATP is used by the heterocyclase enzyme in the manner of a molecular motor or G protein.[13] One of the main lines of evidence in favor of the molecular-machine hypothesis was that ATP was used in super-stoichiometric amounts by microcin B17 synthetase; ATP was not used unless all enzyme components and the substrate were



Scheme 1. Shown above are mechanistic possibilities for heterocyclization. A) Oxidation preceding heterocycle formation. B) Activation of the adjacent carbonyl oxygen, perhaps with phosphate from ATP. C) Intein mechanism for thiazoline formation as a by-product. D) Molecular-machine mechanism for heterocyclase enzymes.

present, but upon incubation with substrate "excess" ATP was consumed. In addition, ATP hydrolysis could be uncoupled from heterocyclization when a large excess of substrate was employed.[19] We therefore performed several experiments to illuminate the role of ATP in the PatD/TruD mechanisms, by using TruD and TruE2.

Virtually no ATP was consumed in the absence of TruE2 substrate. In addition, a negligible ATP background hydrolysis was detected without enzyme. Upon addition of TruE2, a rapid burst in ATP hydrolysis could be observed. In experiments to determine the "minimum" amount of ATP leading to complete TruE2 modification, it was found that addition of 40 μM ATP led to complete modification of 9 μM TruE2 (or 18 μM of heterocycle formed). Although this ATP concentration is well below the K_m, this experiment indicated fewer than 2.5 ATP hydrolysis events are required per heterocycle formed (Figure S9).

Although no ATP is consumed when the TruE2 substrate is absent, we wondered whether ATP would be used when only fully modified TruE2 substrate is present. In a 6 h time-course experiment with 140 nM TruD and 7.4 μM TruE2, heterocycle formation is essentially complete after 1 h; during this initial hour the rate of ADP production is 0.24 μM min^{-1}. During the following five hours, the rate of ADP production is significantly reduced, and remains linear with a rate of 0.06 μM min^{-1} (Figures 2 and S10). When the background rate of hydrolysis due to enzyme only, substrate only and buffer only is accounted for, the "excess" rate of hydrolysis after heterocycle formation is complete remains 0.05 μM min^{-1}; this is a value roughly 20% of the initial fast rate of ATP hydrolysis.

These results strongly support the idea that TruD uses ATP in the manner of a molecular machine, as TruD continues to hydrolyze ATP even after all of the substrate for chemical reaction has been consumed. These results are superficially different than those obtained for the microcin B17 synthetase, in which fivefold more of ATP was used than heterocycles formed (whereas TruD uses closer to a stoichiometric amount of ATP per heterocycle formed). However, the microcin enzymes and substrates are quite different, and stoichiometry experiments in that system involved a precursor peptide that was processed to a tandem-heterocyclic system (McbA 1–46 where GSC becomes G-oxazole-thiazole). The stoichiometry of ATP-used to heterocycles formed was not calculated explicitly for a single-turnover substrate in that system (McbA 1–47 where GSC was mutated to GGC, which is processed to GG-thiazole). However, other experiments showed that the rate of heterocyclization was essentially unchanged compared to wild-type McbA 1–46, but it induced a sevenfold lower level of ATP consumption.[18] Consequently, our results, which show nearly stoichiometric ATP consumption are actually quite consistent with previously reported results, given that there might be something distinctly different about the synthesis of a bisheterocyclic system as found in wild-type McbA 1–46.

We considered several further mechanistic hypotheses consistent with the hydrolysis of ATP in a molecular machine: 1) ATP binding and hydrolysis allows dissociation of a tightly bound enzyme-substrate complex formed after heterocyclization; 2) ATP binding places the enzyme in an active conformation, in which heterocyclization can proceed readily; 3) ATP is used to autophosphorylate the enzyme, thus placing the enzyme into an active conformation. To test hypothesis (1) we conducted an extensive incubation of TruD with the single-cassette precursor TruE4 in the presence and absence of ATP. We used a very large amount of enzyme, such that a single turnover would be apparent in the resulting mass spectrum. No turnover whatsoever was observable in the absence of ATP, even after extended incubation, while the reaction proceeded readily with ATP (Figure S11), a finding that does not support hypothesis (1). To test hypothesis (2) the nonhydrolyzable ATP analogue, β,γ-methylene ATP was added in increasing amounts to the TruD-TruE2 modification reaction; in addition, we attempted to observe TruE4 modification when the nonhydrolyzable analogue was substituted for ATP in the reaction. We found that the nonhydrolyzable ATP analogue strongly inhibited the TruE2 modification reaction, and that no modification to TruE4 could be observed when nonhydrolyzable ATP was employed (Figure S11); this argues for rejection of hypothesis (2). Finally, hypothesis (3) was tested by incubating enzyme, both alone and with substrate with radiolabeled ^{32}P ATP. Reactions were then analyzed by SDS-PAGE and autoradiography. However, we were unable to trap any autophosphorylated enzyme intermediate under any condition attempted. Consequently, it seems that the molecular machine might function through a more complicated mechanism than any proposed mechanism above.

Sequence analysis and comparison

PatD and TruD are didomain proteins that are nearly identical in their N-terminal domains (in the constructs used here, they are 100% identical through residue 323). These N-terminal domains have elsewhere been proposed to directly catalyze heterocyclization.[13] In addition, these enzymes share low similarity with related heterocyclization enzymes from other families such as goadsporin, streptolysin, thiostrepton and relatives, and microcins. The percent identity to microcins is low enough that the sequences are essentially unalignable, but they can be transitively related as the streptolysin enzymes are related to both microcin and cyanobactin enzymes.[5] Although there are highly conserved residues within the cyanobactin genes that could be involved in catalysis, there are no universally conserved sequence features that are shared in common amongst the different heterocyclizing proteins.

By contrast to the N-terminal domains, the C-terminal domains of PatD and TruD are only 77% identical to each other (Figure 1). These domains share homology with "YcaO" domains, which have no known function. In the context of heterocyclization, indirect experimental evidence indicates that they are likely involved largely in substrate binding and ATP hydrolysis, and not directly in catalysis of heterocycle formation.[13,19]

Interestingly, the N- and C-terminal domains of PatD and TruD share short conserved sequence features with MccB, an enzyme involved in adenylation of a peptide intermediate in microcin C7 biosynthesis (though it must be emphasized that microcin C7 does not contain any heterocycles and the biosyn-

thetic purpose of adenylation is distinct from any reaction described in this work). A recent crystal structure of MccB shows that these residues line a substrate-recognition pocket that strongly associates with the C-terminal residues of microcin C7.[24] These microcin C7 residues are very similar to residues found in the C-terminal regions of PatE/TruE. Within MccB, the region of homology ends just before the ATP-binding site and the reaction center. This alignment suggests that the homologous residues in PatD/TruD could bind to PatE/TruE in a similar way, but it does not inform on the role of ATP in heterocyclization. Speculatively, PatE/TruE might be held in place in the region adjacent to ATP binding, and hydrolysis of ATP leads to substrate release. ATP binding and hydrolysis could be either covalent (for example, to the C terminus of the substrate), or noncovalent.

The conserved PatE/TruE leader sequence contains a short region that is predicted to be helical by multiple methods.[25,26] This region is of about the same length as an experimentally determined helical region (in trifluoroethanol) of the microcin B17 precursor peptide. This microcin helix was shown to be critical in interaction with the peptide-binding protein,[27] and it could serve the same function here. Indeed, without the leader sequence PatD and TruD cannot synthesize heterocycles, as shown in this study. The same precursor peptide region has been studied through extensive mutagenesis in the streptolysin S group.[23] It is clear from these studies and others in the ribosomal peptide field that the leader sequence is critical to Cys/Ser/Thr modification in the lantibiotic- and microcin-like peptides.[28]

Mechanistic hypothesis

Our results allow us to rule out a number of alternative hypotheses regarding the mechanism of ribosomal peptide heterocyclases. The apparent requirement for the oxidase in heterocyclization of microcins and streptolysin led to the proposal that oxidation could occur prior to formation of Ser- or Cys-derived heterocycles (Scheme 1 A); this means that the relevant nucleophiles in the heterocyclization reaction would be ene-hydroxyl and ene-thiol.[29] We have shown here that the flavin-containing oxidase is not required for heterocyclization; thus, this demonstrates that this mechanism cannot be correct. Kinetic analysis strongly supports the hypothesis that ATP drives a molecular machine that promotes heterocyclization, as proposed for microcin B17 synthesis (Scheme 1 D). By contrast, a mechanism wherein ATP is used to activate the adjacent carbonyl for attack by sulfur or oxygen (Scheme 1 B) is rendered less plausible by the results presented here.

Recently, it was shown that intein chemistry can yield heterocyclic thiazolines as side products (Scheme 1 C).[30] Thus, it seems possible that heterocyclization chemistry could be quite similar to that required for intein splicing, as well as the chemical strategy employed by many autoproteolyzing enzymes.[31] On this account, the peptide merely needs to be held in the right conformation, with appropriate nearby acids and bases to accelerate nucleophilic attack on the carbonyl adjacent to Cys or Ser/Thr.

Given intein chemistry (which proceeds without added energy, that is, ATP), it seems mysterious why ATP would be required by heterocyclase enzymes. One possible explanation for this discrepancy is found in the fact that inteins are, in effect, single-turnover enzymes. In the case of multiple turnover heterocyclases, the enzyme-substrate complex could be so tightly bound that ATP is needed for product release, or the enzyme-substrate complex might be unable to form without the use of ATP. In addition, this system is strongly reminiscent of AAA-proteases, which require ATP for the detection of misfolded proteins and then hydrolyze these aberrant proteins through separate protease domains.[32] It is noteworthy that there is sequence homology, both in substrate and enzyme, between cyanobactin biosynthetic proteins and microcin C7 proteins, which do not catalyze heterocyclization but instead adenylation. This similarity suggests a possible evolutionary relationship connecting the microcin-group biosynthetic pathways.

Conclusions

We have characterized two heterocyclization enzymes from cyanobactin pathways, PatD and TruD. In both cases, single enzymes were sufficient to recapitulate heterocyclization activity; this makes this the first single-protein reconstitution of heterocyclization activity and the first thiazoline/oxazoline synthetases to be characterized in ribosomal systems. Previous work in the microcin B17 and streptolysin S systems had shown that three protein domains are required for the biosynthesis of ribosomal peptide-derived heterocycles: an ATPase, a zinc-binding domain, and a flavin-containing oxidase.[5,13] By contrast, in the PatD/TruD group, oxidation is not a required component of catalysis, and the heterocyclization activity exists within a single polypeptide. These differences substantially simplified our approaches to gain mechanistic insights into the function and mechanism of heterocyclases in nature; this allowed us to provide confirmatory evidence that ATP is used to drive a molecular-heterocyclization machine. These results also define the order of steps in cyanobactin biosynthetic pathways, which lead to diverse and ubiquitous cyanobacterial natural products.

Experimental Section

General methods: Isopropyl β-D-1-thiogalactopyranoside (IPTG), dithiothreitol (DTT), leupeptin, pepstatin, 4-(2-aminoethyl) benzenesulfonyl fluoride (AEBSF), and phenylmethanesulfonyl fluoride (PMSF) were purchased from ISC Bioexpress. Metal-free nitric acid (Optima) was purchased from Fisher Scientific. Ultrapure $MgCl_2$, and β,γ-adenosine triphosphate (β,γ-ATP) were purchased from Sigma–Aldrich. γ-^{32}P-labeled ATP was purchased from Perkin–Elmer. Ni-NTA resin was purchased from Qiagen. ZipTip C18 pipette tips were purchased from Millipore. Krypton fluorescent protein stain was purchased from Pierce. All expression vectors were purchased from Novagen. *Escherichia coli* strain DH5α was used for all cloning steps, while *E. coli* strain BL21(DE3)Star was used for all protein expressions.

Gene cloning and expression: Genes were obtained from ascidian symbiont metagenomes and cloned as previously described.[16,17] Enzymes were expressed as previously described, except that for

kinetic analysis additional enzyme purification was performed. Nickel-purified TruD was loaded onto a HiPrep 16/10 Q FF column (GE Healthcare), and run on an AKTA purifier FPLC system. The column was washed with 0.5 column volumes of buffer A, a linear gradient from 100% buffer A–0% buffer B to 20% buffer A–80% buffer B over 30 column volumes was run, followed by a five column volume wash at 0% buffer A–100% buffer B. Buffer A consisted of NaCl (0.1 M) buffered to pH 8.0 with HEPES (25 mM), while buffer B consisted of NaCl (1 M) buffered to pH 8.0 with HEPES (25 mM). TruE2 used for kinetic analysis was purified as previously described,[17] except that the protein was stocked in a solution containing NaCl (350 mM), DTT (10 mM), and sucrose (1% *w/v*) buffered to pH 8.0 with HEPES (20 mM).

Enzyme reactions: All enzyme reactions used in this study were performed at least in triplicate in independent runs. Reaction mixtures were incubated at 34 °C in an MJ research minicycler for varying amounts of time. Enzyme, precursor peptide, and ATP concentrations varied, as described below. The following additives were present in standard reactions but were varied in early optimization experiments: tris pH 8.0 (40 mM), DTT (8 mM), and $MgCl_2$ (4 mM). Enzyme reactions generally contained the optimized additive mixture and PatD or TruD (0.6 μM), TruE2 (8 μM), or TruE4 (12 μM), and ATP (0.8 mM). Reactions were run using varying times, from 15 min to 27 h. To confirm modification, the reaction mixture (10 μL) was removed and analyzed by SDS-PAGE. At minimum, at least three separate experiments were performed for each enzyme-substrate concentration.

Sulfonyl fluoride inhibition: AEBSF and PMSF were used with standard concentrations of reagents, including TruE2 and TruD or PatD. Concentrations of sulfonyl fluorides were tested at 1 mM and 10 mM and compared to controls containing equivalent amounts of vehicle only (methanol for PMSF, water for AEBSF). PMSF was tested only against TruD. After 1.5 h reactions, the mixtures were analyzed by SDS-PAGE.

ATP conversion: To ascertain whether PatD and TruD hydrolyze ATP to ADP, or to AMP, PatD or TruD (0.6 μM) were incubated with TruE2 (2 μM), and ATP (100 μM) for 30 min. Reactions were then quenched by adding 10 μL of a saturated solution of urea and analyzed by HPLC as described below.

ATP stoichiometry: All reactions were performed in triplicate, by using standard reaction conditions with TruE2 and PatD or TruD. No enzyme and no substrate controls were performed. Further controls contained AEBSF (10 mM) and were performed with enzyme and substrate or without substrate. In one round, reactions were run with ATP (800 μM) for 0.5 h, then quenched with urea (8 M).

Nonhydrolyzable ATP inhibition: Reactions were performed according to standard conditions except that two sets of reactions were used, one containing ATP (800 μM) and the other ATP (200 μM). β,γ-methylene-ATP was added to each set (800 or 200 μM ATP) at the following concentrations: 0, 100, 200, 400, 800, and 1600 μM. Reactions were allowed to proceed for 0.5 h, and then analyzed by SDS-PAGE.

^{32}P-ATP labeling experiments: γ-^{32}P-ATP (1 μCi) was doped into reactions containing cold ATP (40 μM). Standard reaction conditions were used with TruD and TruE2. No substrate, no enzyme, and AEBSF-inhibited controls were performed alongside these reactions. Reactions were analyzed both by adsorption to Nytran paper followed by scintillation counting, as well as by SDS-PAGE followed by autoradiography.

SDS-PAGE assays: 18% acrylamide gels were used for all assays. Prior to electrophoresis, samples were brought up in 1X SDS sample buffer diluted from 6× SDS sample buffer: tris pH 6.8 (7 mL, 0.5 M) glycerol (3 mL), SDS (1 g), DTT (0.93 g), bromophenol blue (1.2 mg), H_2O (up to 10 mL), and then boiled for 3 min.[33] After electrophoresis, gels were placed in boiled fixing solution consisting of H_2O (53% *v/v*), ethanol (40% *v/v*), and acetic acid (7% *v/v*), incubated for 10–20 min with gentle rocking, and then placed in boiled stain solution consisting of: Coomassie R250 (0.02% *w/v*) in H_2O (85% *v/v*), acetic acid (10% *v/v*), and ethanol (5% *v/v*). Gels were then destained in a solution consisting of H_2O (85% *v/v*), acetic acid (10% *v/v*), and ethanol (5% *v/v*) for several hours, these gels were then photographed. For TruE2 time course experiments, gels were stained using Krypton fluorescent protein stain (Pierce) according to the manufacturer's instructions, and imaged using a Typhoon fluorescence reader (GE).

HPLC analysis: The HPLC method for all ATP-usage and stoichiometric experiments employed a Vydac 302IC4.6 ion-exchange column, and a Hitachi LaChrom Elite HPLC system. A linear gradient proceeding from 100% buffer A, which consisted of formic acid (45 mM), adjusted to pH 4.5 using NaOH, to 100% buffer B, which consisted of NaH_2PO_4 (0.5 M), adjusted to pH 2.5 using formic acid over 12 min was used to effect separation of ATP, ADP, and AMP. The elution profiles of the experimental runs were compared to those of authentic ATP, ADP, and AMP standards. Peaks were quantified by comparison with a calibration curve constructed by injecting known quantities of AMP.

Metal requirements: To assess whether or not the enzymes bound zinc as predicted, PatD and TruD were dialyzed over two days with stirring at 4 °C against a solution (2 L) containing glycerol (5% *v/v*), NaCl (500 mM), and Sepharose chelating resin (10 g) buffered to pH 7.8 using HEPES (25 mM). The purified, dialyzed enzyme samples were then digested using metal-free nitric acid, heated to 95 °C, and read on a Perkin–Elmer Optima 3100 XL ICP-OES instrument.

The potential requirement for other metals (aside from magnesium and zinc) was tested by using ultrapure $MgCl_2$ (Aldrich 255 777) with TruD (1.6 μM), and TruE2 (2 μM), using Chelex-treated ATP and DTT in standard concentrations without tris buffer. ATP and DTT were passed through Chelex 100 (Biorad) resin (700 μL) prior to addition to the reaction. An additional set of controls were performed without using Chelex treatment.

Enzyme kinetics: Reactions were performed in the same manner as the standard conditions described above, except that enzyme, substrate, ATP, and AEBSF concentrations were varied. In the experiments where the concentration of ADP was measured at 0, 20, and 40 min, experiments were performed with variable TruD, TruE2, ATP, and AEBSF. Experiments varying TruD were performed as follows: TruD was added at variable concentrations (140, 104, and 69 nM) while holding constant TruE2 (7.4 μM), ATP (800 μM), and AEBSF (0 μM). Experiments varying TruE2 were performed as follows: TruE2 was added at variable concentrations (7.4, 5, and 2.5 μM) while holding constant TruD (104 nM), ATP (800 μM), and AEBSF (0 μM). Experiments varying ATP were performed as follows: ATP was added at variable concentrations (800, 400, and 200 μM) while holding constant TruD (104 nM), TruE2 (7.4 μM), and AEBSF (0 μM). Experiments varying AEBSF were performed as follows: AEBSF added at variable concentrations (10, 5, and 1 mM) while holding constant TruD (104 nM), TruE2 (7.4 μM), and ATP (800 μM). Controls were run that lacked TruD while holding constant TruE2 (7.4 μM), ATP (800 μM), and AEBSF (0 μM). Additionally, controls

were run that lacked TruE2 while holding constant TruD (104 nM), ATP (800 μM), and AEBSF (0 μM). In experiments measuring concentrations of ADP at 0, 60, 120, 240, and 360 min, the reactions contained TruD (140 nM), TruE2 (7.4 μM), and ATP (800 μM). Controls were run that lacked either enzyme, substrate, or both. All reactions were performed in triplicate. After removal of aliquots at each time point, the reactions were quenched by addition of an equal volume of 8 M urea, and then frozen at −80 °C until analysis. Nucleotide content was assessed by HPLC as described above. Additionally, selected reactions were analyzed by SDS-PAGE to ensure that the heterocyclization reaction was proceeding at the expected pace. Further, to analyze the stoichiometry of ATP-hydrolyzed to heterocycles formed, samples of the 0, 60, 120, 240, and 360 min time points described above were analyzed by SDS-PAGE, stained with Krypton fluorescent protein stain according to the manufacturer's instructions, and then imaged on a Typhoon fluorescence reader. The resulting images were analyzed for band densitometry using the program ImageJ.

Protein quantitation: TruE2 used for kinetic analysis was quantitated through amino-acid analysis. The protein was dialyzed extensively against a solution containing NaCl (0.35 M), DTT (10 mM), sucrose (1% *w/v*), and then subjected to amino-acid analysis. The concentration was calculated based on the nanomoles recovered of Gly, Ala, Leu, Tyr, Phe, Lys, His, and Arg, which were averaged together between two separate runs.

Acknowledgements

This work was funded by NIH GM071425 and by a Willard Eccles Fellowship and an ACS Medicinal Chemistry Fellowship funded by Sanofi-Aventis to J.A.M. We thank Pam Smith and Dennis Winge for ICP-OES assistance and Darrell Davis for help conducting the ATP-usage experiments. We also thank Archana Yerra for technical assistance, Adele Flail for graphical assistance, and Janet Lindsley for helpful discussions. Lastly we thank an anonymous reviewer for suggesting several additional experiments in a previous revision of this paper.

Keywords: heterocycles · microcin · natural products · thiazoline · trunkamide

[1] E. W. Schmidt in *The UCSD Guardian, Vol. 74*, **1991**.
[2] F. Y. F. Lee, R. Borzilleri, C. R. Fairchild, S.-H. Kim, B. H. Long, C. Reventos-Suarez, G. D. Vite, W. C. Rose, R. A. Kramer, *Clin. Cancer Res.* **2001**, *7*, 1429.
[3] R. S. Roy, A. M. Gehring, J. C. Milne, P. J. Belshaw, C. T. Walsh, *Nat. Prod. Rep.* **1999**, *16*, 249.
[4] B. Anderson, D. Hodgkin, M. A. Viswamitra, *Nature* **1970**, *225*, 233.
[5] S. W. Lee, D. A. Mitchell, A. L. Markley, M. E. Hensler, D. Gonzalez, A. Wohlrab, P. C. Dorrestein, V. Nizet, J. E. Dixon, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 5879.
[6] P. Yorgey, J. Lee, J. Kordel, E. Vivas, P. Warner, D. Jebaratnam, R. Kolter, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 4519.
[7] C. Li, W. L. Kelly, *Nat. Prod. Rep.* **2010**, *27*, 153.
[8] M. S. Donia, E. W. Schmidt in *Comprehensive Natural Products II: Chemistry and Biology* (Eds.: L. Mander, H.-W. Liu), Elsevier, **2010**.
[9] M. S. Donia, J. Ravel, E. W. Schmidt, *Nat. Chem. Biol.* **2008**, *4*, 341.
[10] W. L. Kelly, N. J. Hillson, C. T. Walsh, *Biochemistry* **2005**, *44*, 13385.
[11] Y. M. Li, J. C. Milne, L. L. Madison, R. Kolter, C. T. Walsh, *Science* **1996**, *274*, 1188.
[12] J. A. McIntosh, M. S. Donia, E. W. Schmidt, *Nat. Prod. Rep.* **2009**, *26*, 537.
[13] J. C. Milne, R. S. Roy, A. C. Eliot, N. L. Kelleher, A. Wokhlu, B. Nickels, C. T. Walsh, *Biochemistry* **1999**, *38*, 4768.
[14] T. L. Schneider, B. Shen, C. T. Walsh, *Biochemistry* **2003**, *42*, 9722.
[15] J. Lee, J. A. McIntosh, B. J. Hathaway, E. W. Schmidt, *J. Am. Chem. Soc.* **2009**, *131*, 2122.
[16] M. S. Donia, B. J. Hathaway, S. Sudek, M. G. Haygood, M. J. Rosovitz, J. Ravel, E. W. Schmidt, *Nat. Chem. Biol.* **2006**, *2*, 729.
[17] J. A. McIntosh, M S. Donia, E. W. Schmidt, *J. Am. Chem. Soc.* **2010**, 132, 4089.
[18] D. B. Zamble, C. P. McClure, J. E. Penner-Hahn, C. T. Walsh, *Biochemistry* **2000**, *39*, 16190.
[19] J. C. Milne, A. C. Eliot, N. L. Kelleher, C. T. Walsh, *Biochemistry* **1998**, *37*, 13250.
[20] L. L. Madison, E. I. Vivas, Y. M. Li, C. T. Walsh, R. Kolter, *Mol. Microbiol.* **1997**, *23*, 161.
[21] R. Sinha Roy, P. J. Belshaw, C. T. Walsh, *Biochemistry* **1998**, *37*, 4125.
[22] N. Allali, H. Afif, M. Couturier, L. Van Melderen, *J. Bacteriol.* **2002**, *184*, 3224.
[23] D. A. Mitchell, S. W. Lee, M. A. Pence, A. L. Markley, J. D. Limm, V. Nizet, J. E. Dixon, *J. Biol. Chem.* **2009**, *284*, 13004.
[24] C. A. Regni, R. F. Roush, D. J. Miller, A. Nourse, C. T. Walsh, B. A. Schulman, *Embo J.* **2009**, *28*, 1953.
[25] C. Cole, J. D. Barber, G. J. Barton, *Nucleic Acids Res.* **2008**, *36*, W197.
[26] K. Bryson, L. J. McGuffin, R. L. Marsden, J. J. Ward, J. S. Sodhi, D. T. Jones, *Nucleic Acids Res.* **2005**, *33*, W36.
[27] R. S. Roy, S. Kim, J. D. Baleja, C. T. Walsh, *Chem. Biol.* **1998**, *5*, 217.
[28] T. J. Oman, W. A. van der Donk, *Nat. Chem. Biol.* **2010**, *6*, 9.
[29] T. Kupke, F. Götz, *J. Biol. Chem.* **1997**, *272*, 4759.
[30] C. Ludwig, D. Schwarzer, H. D. Mootz, *J. Biol. Chem.* **2008**, *283*, 25264.
[31] C. T. Walsh, *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, 1st ed., Roberts, Greenwood Village, **2006**.
[32] F. Striebel, W. Kress, E. Weber-Ban, *Curr. Opin. Struct. Biol.* **2009**, *19*, 209.
[33] S. Gallagher, *Curr. Protoc. Mol. Biol.* **1998**, *6*.1.1.

Received: March 27, 2010
Published online on June 10, 2010

CHEMBIOCHEM

Supporting Information

Marine Molecular Machines: Heterocyclization in Cyanobactin Biosynthesis

John A. McIntosh and Eric W. Schmidt*[a]

cbic_201000196_sm_miscellaneous_information.pdf

Figure S1. Precursor peptides of trunkamide and patellamide pathways. Shown are amino acid sequences for TruE4, TruE2, and PatE2. The second cassette of both TruE2 and TruE4 encodes the natural product trunkamide, while the first cassette of TruE2 encodes patellin-6. The first, and second cassettes of PatE2 encode patellamide C, and ulithiacyclamide, respectively; PatE2 is not used in this study, and is shown here merely for comparison.

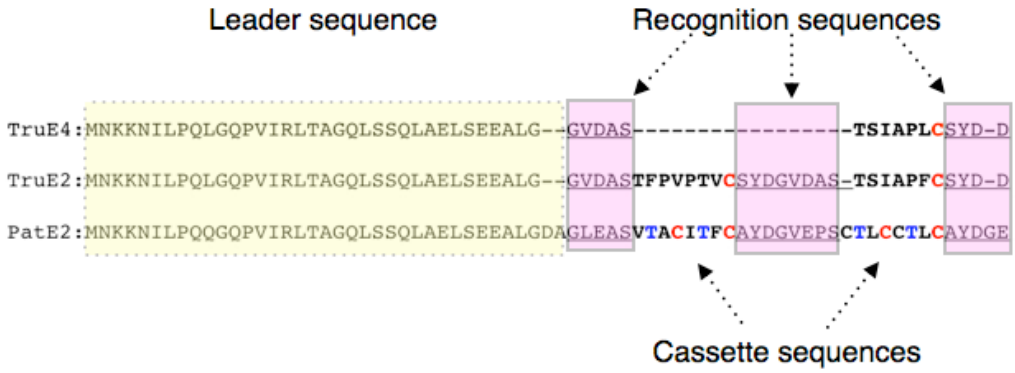


Figure S2. Patellamide pathway biosynthetic scheme. Shown is modification of the patellamide C cassette of the PatE2 precursor shown above in Figure S1. First, PatD catalyzes heterocyclization of Cys or Thr residues, PatA then cleaves the N-terminal side of the cassette, and then finally, PatG cleaves the C-terminal side of the cassette in tandem with macrocyclization; an oxidase domain of PatG also catalyzes oxidation at an unknown juncture during biosynthesis.

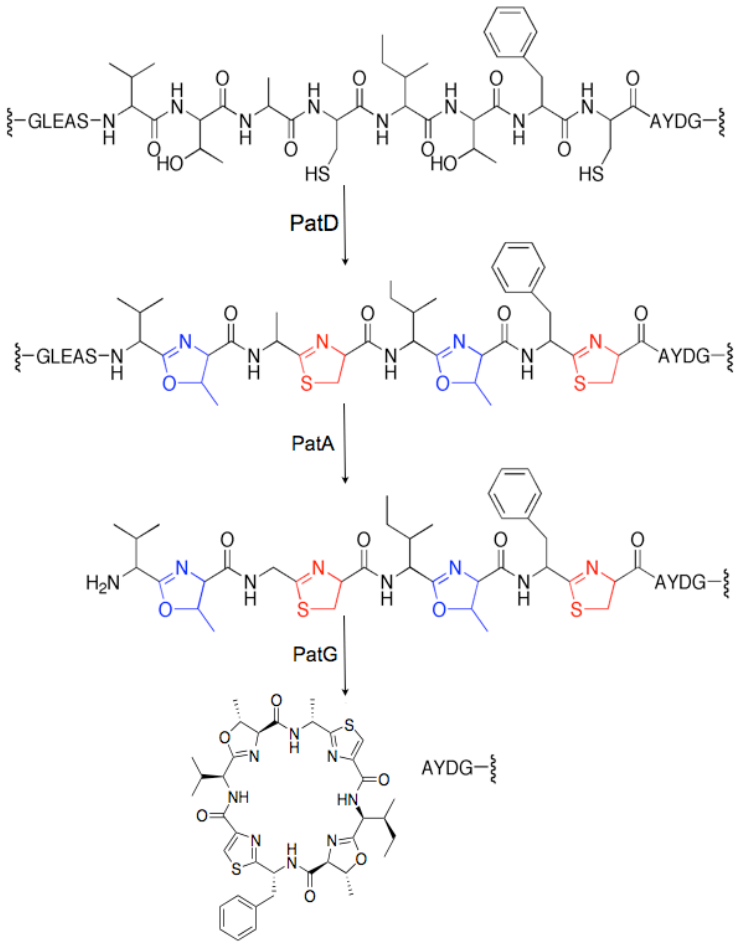


Figure S3. Intact analysis of TruD- and PatD-treated TruE2. A), B), and C) show, intact mass spectrometric analyses of A) unmodified TruE2, B) TruD-treated TruE2, and C) PatD-treated TruE2. Change in mass is consistent with loss of water attendant upon thiazoline/oxazoline ring formation. D) SDS-PAGE showing typical band shifts resulting from TruE2 modification reactions. From left to right, lanes are as follows: (1) TruE2 standard (2) TruD-treated TruE2 (3) TruD-treated TruE2 + TruE2 standard (4) PatD-treated TruE2 (5) PatD-treated TruE2 + TruE2 standard (6) TruE2 standard (7) ladder.

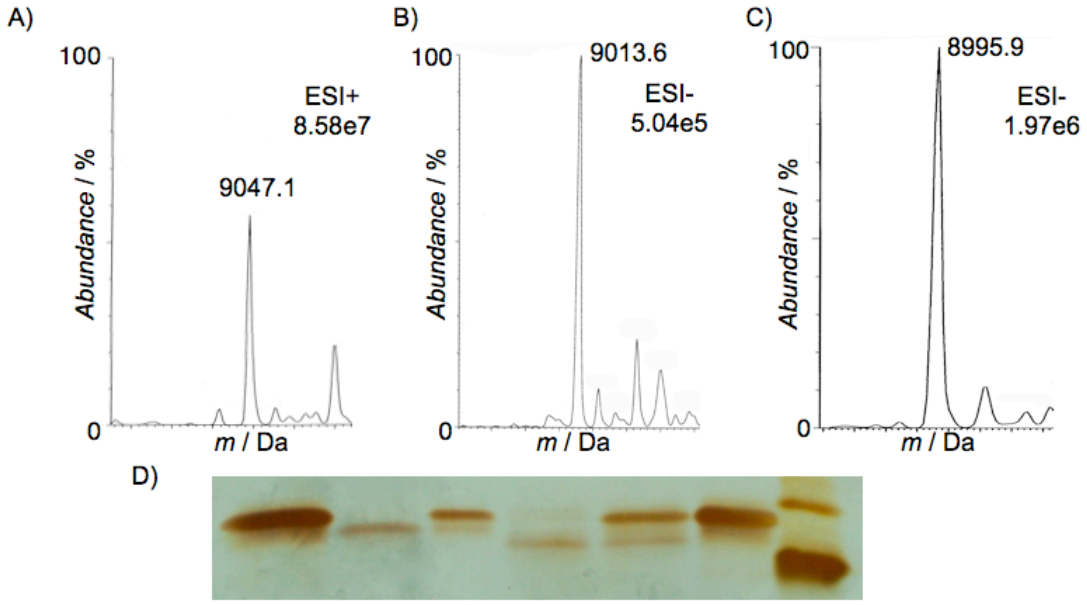


Figure S4. HPLC analysis of adenosine nucleotides produced after 1/2 hour reaction of TruD or PatD with TruE2. Absorbance was monitored at 260 nm.

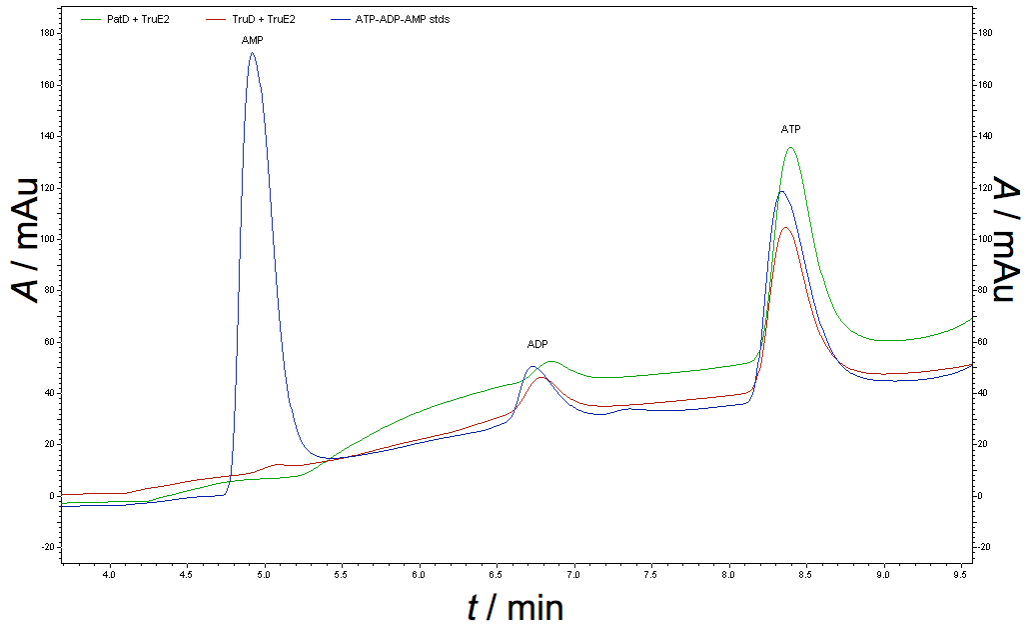


Figure S5. SDS-PAGE gel showing time course of PatD reaction. Letters beside gel bands in lane #2 denote the following: A) unmodified TruE2 B) Doubly modified TruE2 C) Fully modified TruE2. From the left lanes are: (1) TruE2 standard (2) 0.5h PatD + TruE2 (3) 1h PatD + TruE2 (4) 2h PatD + TruE2 (5) 4h PatD + TruE2 (6) 24h PatD + TruE2.

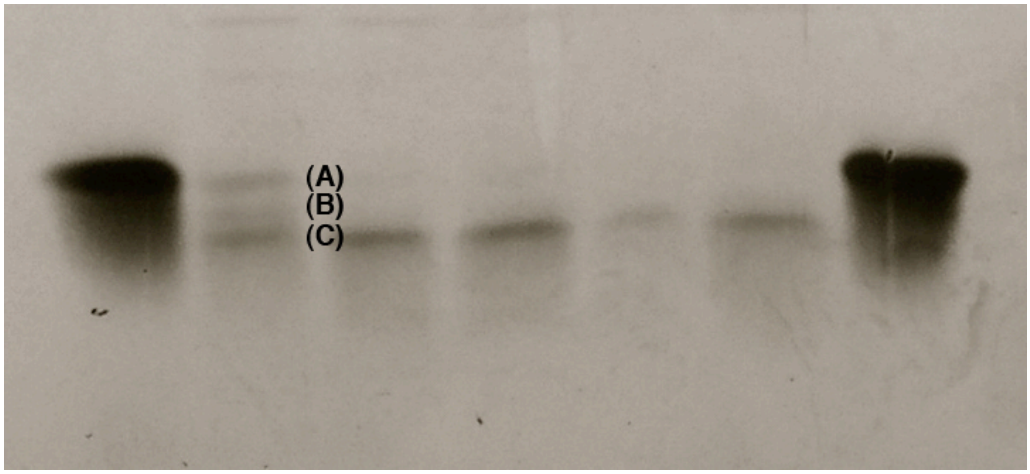


Figure S6. Deconvoluted mass spectrum showing TruE2 modification by TruD when an inadequate amount of ATP was supplied to drive the reaction to completion. Most of the TruE2 remains unmodified (9048.1Da), while another portion is singly dehydrated (9030.6Da), still another portion has a mass consistent with two dehydrations (9013.3Da).

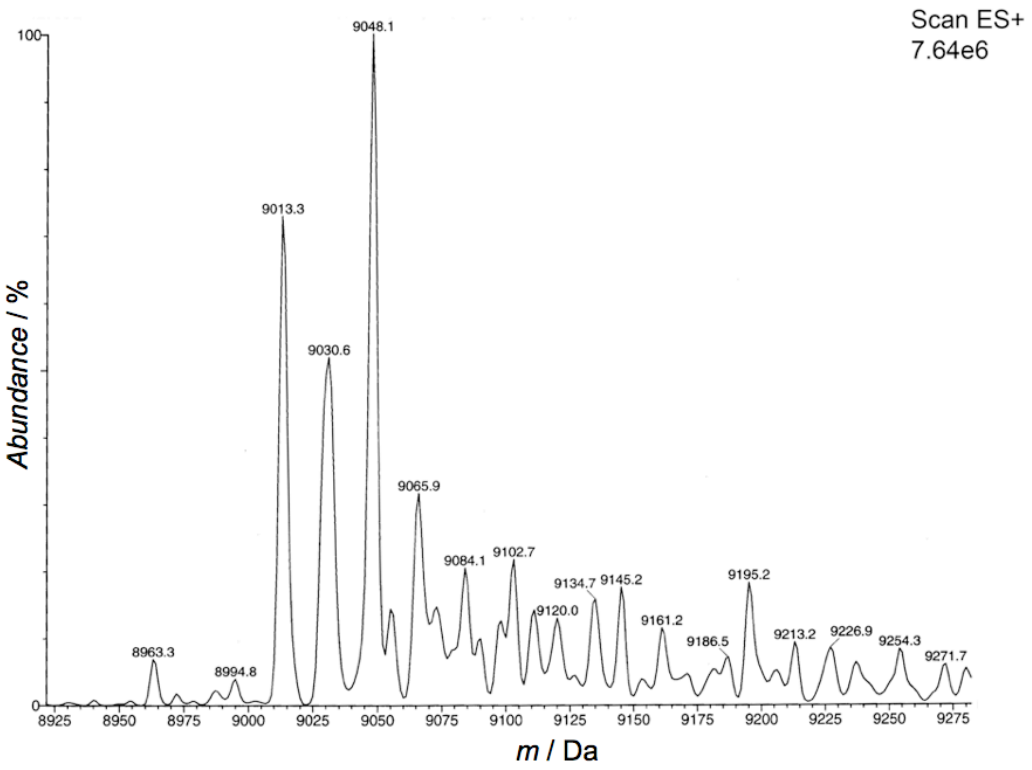
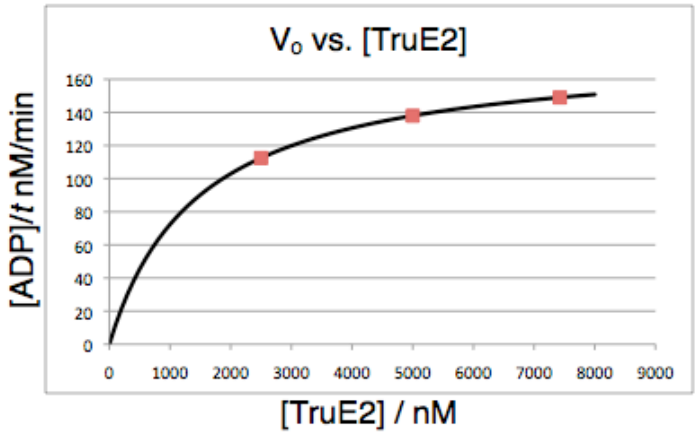


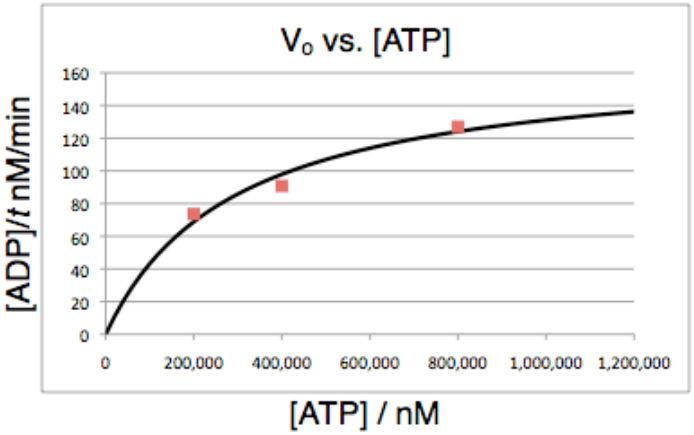
Figure S7. Estimated Michaelis-Menten parameters for TruD, varying either [TruE2] or [ATP].



V_{max}=1.8*10^2 nM/min

K_m=1.5*10^3 nM

k_{cat}=2.6 min^{-1}

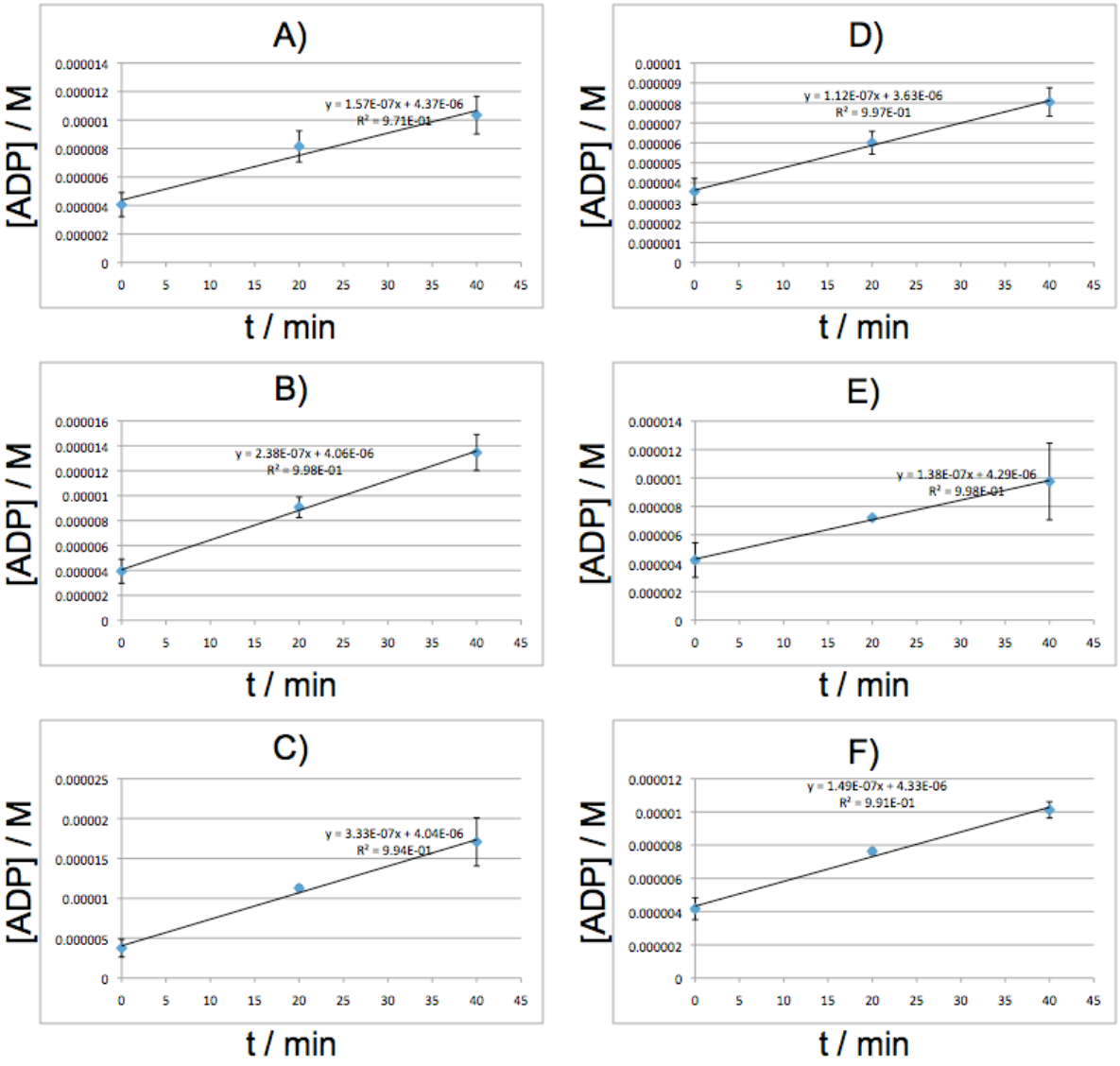


V_{max}=1.7*10^2 nM/min

K_m=3*10^2 µM

k_{cat}=2.4 min^{-1}

Figure S8. Full kinetics data. Error bars represent standard deviations. Y-axes represent [ADP] in molar concentration; X-axes represent time in minutes. A), B), C) Variable TruD (140, 104, and 69 nM), constant TruE2 (7.4 μM) and ATP (800 μM); D), E), F) Variable TruE2 (7.4, 5.0, and 2.5 μM), constant TruD (104 nM) and ATP (800 μM). G), H), I) Variable ATP (800, 400, 200 μM ATP), constant TruE2 (7.4 μM) and TruD (104 nM). J) TruD (104nM), ATP (800 μM), no TruE2 K) ATP (800 μM), TruE2 (7.4 μM).


A)
y = 1.57E-07x + 4.37E-06
R² = 9.71E-01
[ADP] / M
t / min
D)
y = 1.12E-07x + 3.63E-06
R² = 9.97E-01
[ADP] / M
t / min
B)
y = 2.38E-07x + 4.06E-06
R² = 9.98E-01
[ADP] / M
t / min
E)
y = 1.38E-07x + 4.29E-06
R² = 9.98E-01
[ADP] / M
t / min
C)
y = 3.33E-07x + 4.04E-06
R² = 9.94E-01
[ADP] / M
t / min
F)
y = 1.49E-07x + 4.33E-06
R² = 9.91E-01
[ADP] / M
t / min

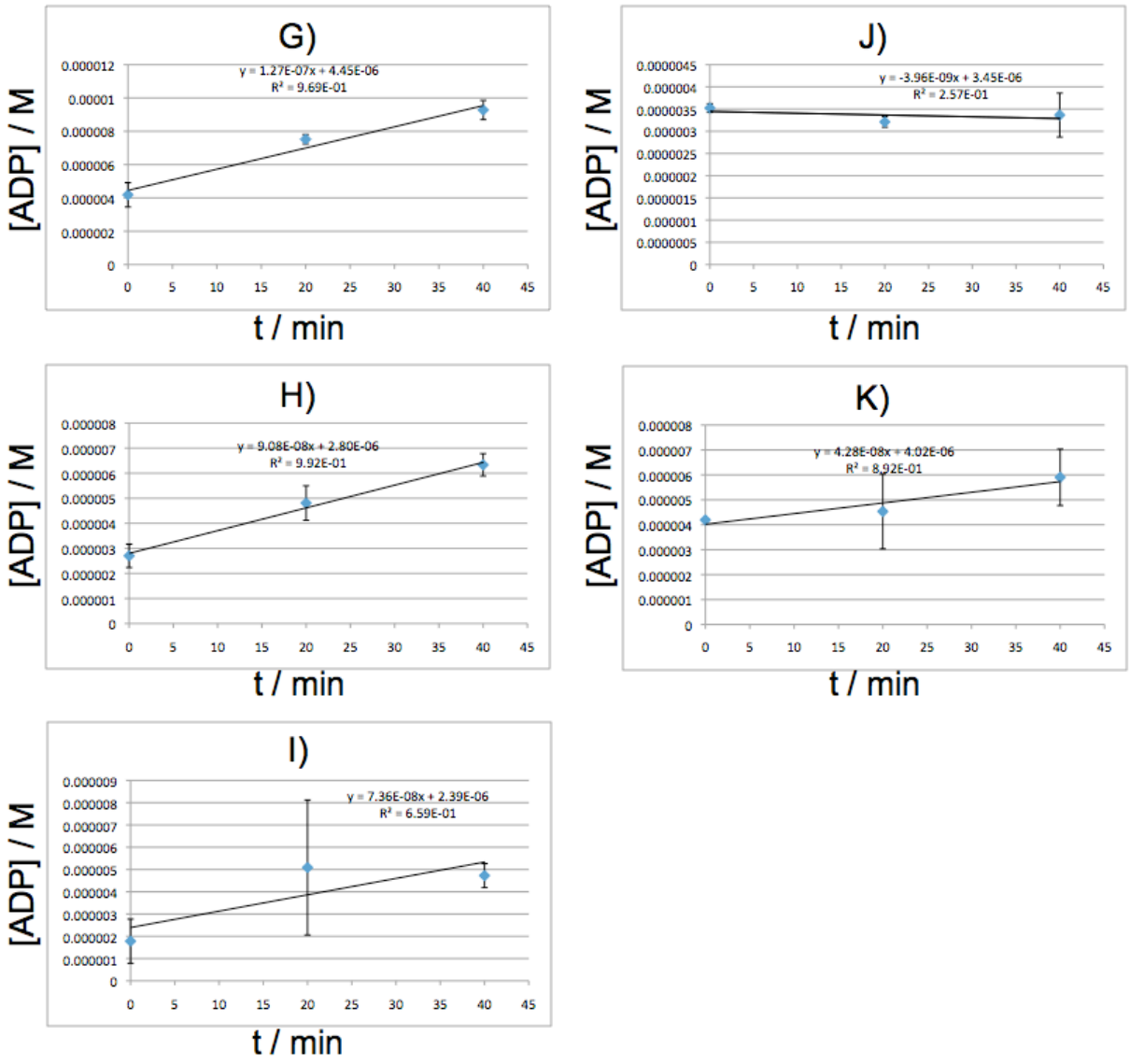
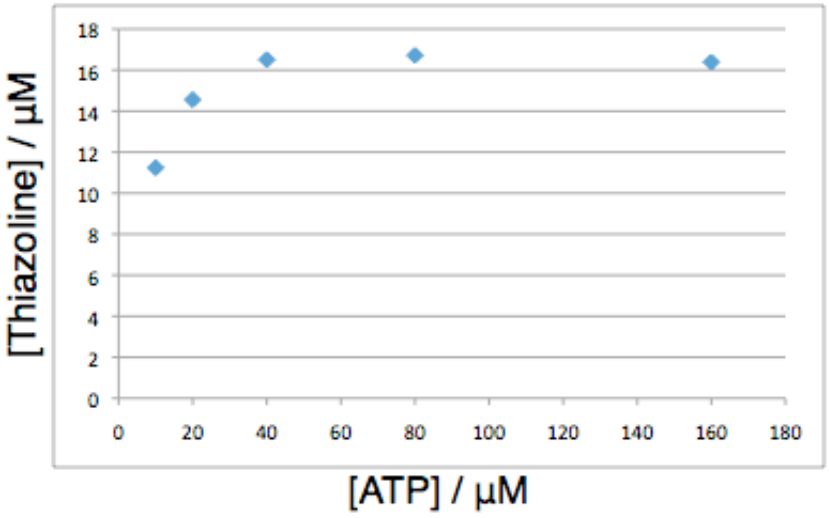

G)
y = 1.27E-07x + 4.45E-06
R² = 9.69E-01
[ADP] / M
t / min
J)
y = -3.96E-09x + 3.45E-06
R² = 2.57E-01
[ADP] / M
t / min
H)
y = 9.08E-08x + 2.80E-06
R² = 9.92E-01
[ADP] / M
t / min
K)
y = 4.28E-08x + 4.02E-06
R² = 8.92E-01
[ADP] / M
t / min
I)
y = 7.36E-08x + 2.39E-06
R² = 6.59E-01
[ADP] / M
t / min

Figure S9. Minimum ATP required for complete heterocyclization. A) Graph showing relation between [ATP] provided and [Thiazoline] present at completion of TruD-TruE2 modification reaction as measured by gel densitometry. B) Gel used to derive [Thiazoline]; from left to right, lanes are as follows: (1) TruE2 standard (2) ATP (10 μM) (3) ATP (20 μM) (4) ATP (40 μM) (5) ATP (80 μM) (6) ATP (160 μM) (7) TruE2 standard

A)



B)

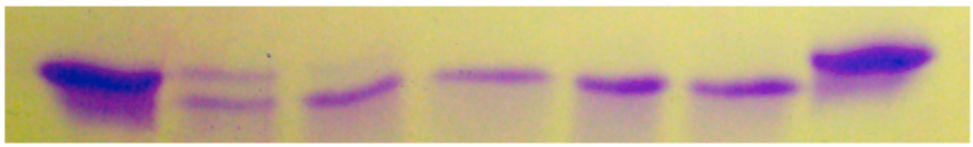


Figure S10. Background rates of ATP hydrolysis are shown under various conditions: A) ATP hydrolysis due to TruE2 with buffer B) ATP hydrolysis due to buffer alone C) ATP hydrolysis due to TruD with buffer D) ATP hydrolysis due to TruE2, TruD and buffer in enzymatic reaction.

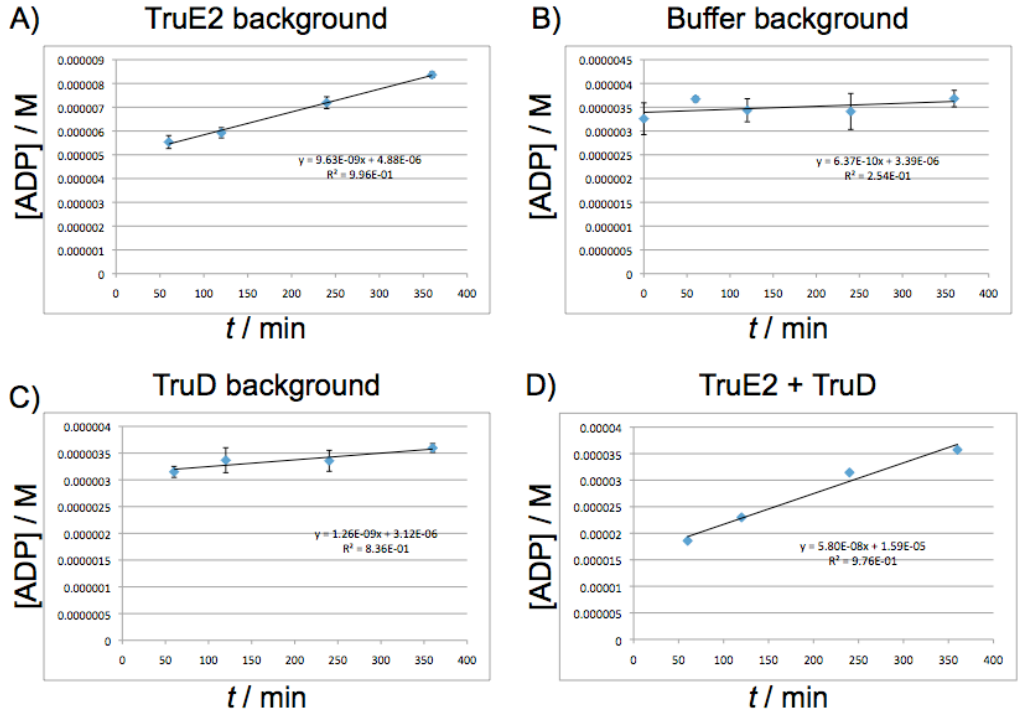


Figure S11. Intact ESI-MS and SDS-PAGE analyses of reactions including or lacking ATP, or including a non-hydrolyzable ATP analogue (β,γ-methylene-ATP). A) TruD-TruE4 reaction with ATP added B) TruE4 reaction lacking ATP, but containing a 1:4 enzyme:substrate ratio C) TruD-TruE4 reaction to which non-hydrolyzable ATP analogue was added D) SDS-PAGE analysis of TruD-TruE2 reactions were titrated with increasing amounts of non-hydrolyzable ATP, and allowed to react for 0.5h. From the left, lanes are: (1) TruE2 standard (2) ATP (800 μM) (3) β,γ-methylene-ATP (100 μM), ATP (800 μM) (4) β,γ-methylene-ATP (200 μM), ATP (800 μM) (5) β,γ-methylene-ATP (400 μM), ATP (800 μM) (6) β,γ-methylene-ATP (800 μM), ATP (800 μM) (7) β,γ-methylene-ATP (1.6 mM); ATP (800 μM) (8) ATP (200 μM) (9) β,γ-methylene-ATP (100 μM), ATP (200 μM) (10) β,γ-methylene-ATP (200 μM), ATP (200 μM) (11) β,γ-methylene-ATP (400 μM), ATP (200 μM) (12) β,γ-methylene-ATP (800 μM), ATP (200 μM) (13) β,γ-methylene-ATP (1.6 mM), ATP (200 μM) (14) TruE2 standard (15) ladder.

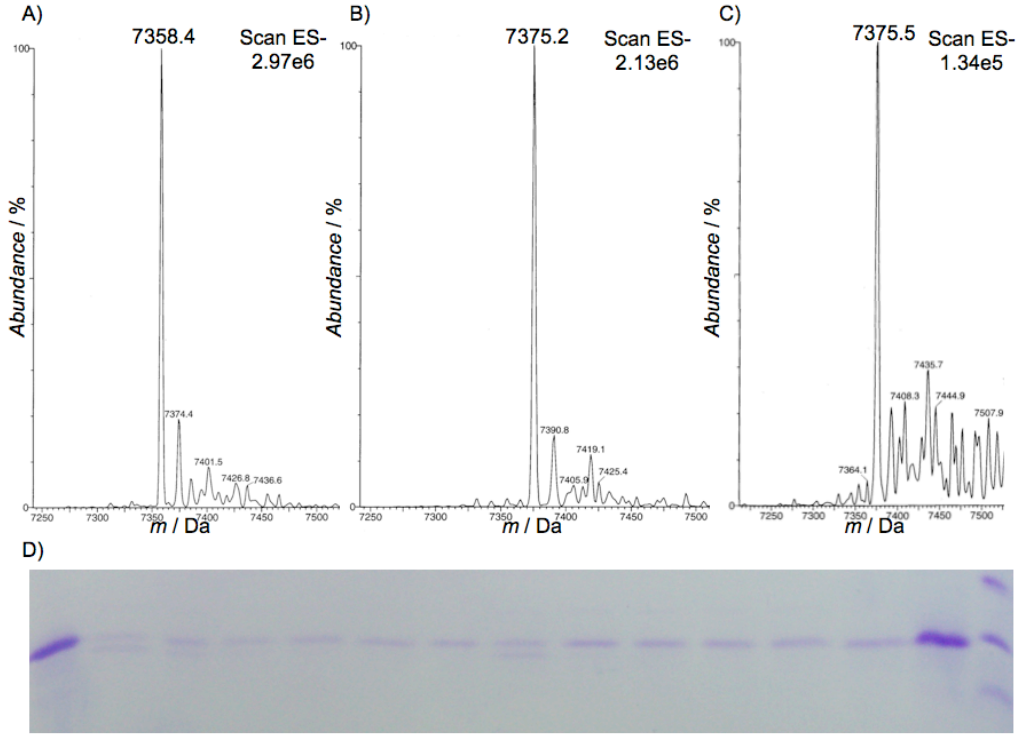
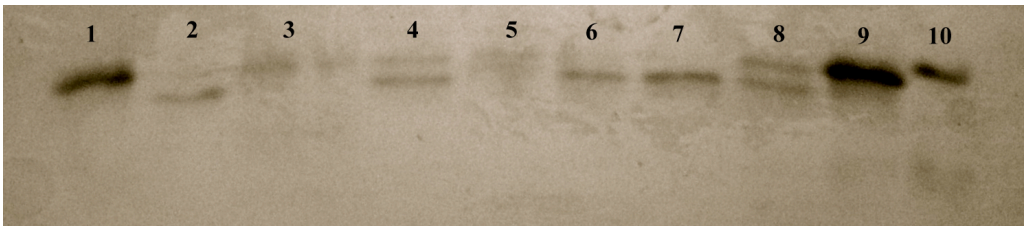
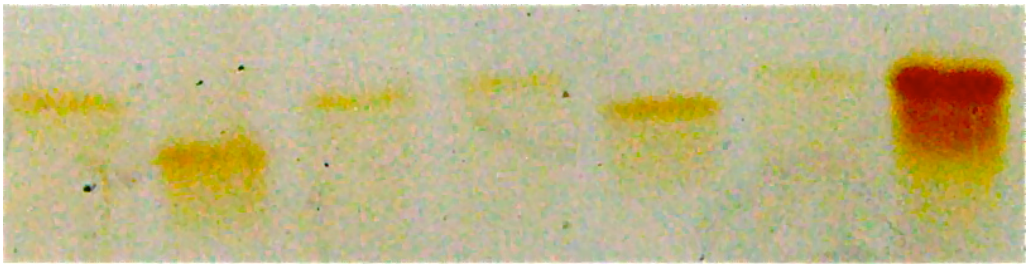


Figure S12. SDS-PAGE gels showing inhibition of TruD and PatD by AEBSF and/or PMSF. A) SDS-PAGE gel showing inhibition of TruD-catalyzed TruE2 modification by AEBSF and PMSF; AEBSF (0.1 M) stock solution was made in water, PMSF (0.1M) stock solution was made in methanol. From the left, lanes are: (1) TruE2 standard (2) TruD+TruE2 (3) TruD + TruE2 + AEBSF (10 mM) (4) TruD + TruE2 + MeOH (10% v/v) (5) TruD + TruE2 PMSF (10 mM) (6) TruD + TruE2 + AEBSF (1 mM) (7) TruD + TruE2 + MeOH (1% v/v) (8) TruD + TruE2 + PMSF (1 mM) (9) TruE2 standard (10) Ladder. B) SDS-PAGE gel showing inhibition of TruD- and PatD-catalyzed TruE2 modification by AEBSF. From the left, lanes are: (1) PatD + TruE2 + 10mM AEBSF (2) PatD + TruE2 (3) TruE2 + 10mM AEBSF (4) TruD + TruE2 + 10mM AEBSF (5) TruD + TruE2 (6) TruE2 + 10mM AEBSF (7) TruE2 standard. C) Kinetic analysis of AEBSF inhibition.

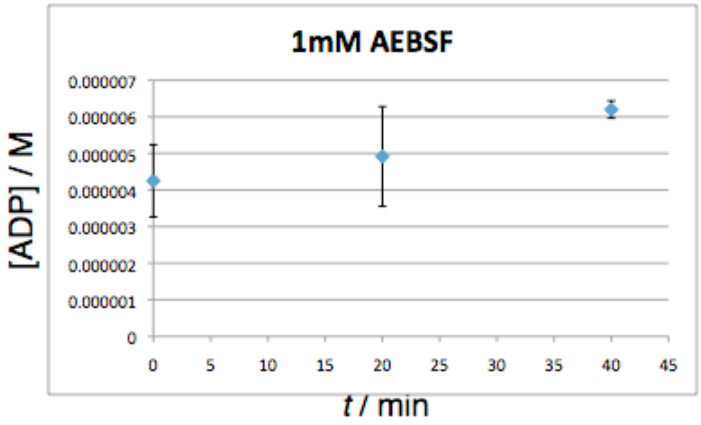
A)

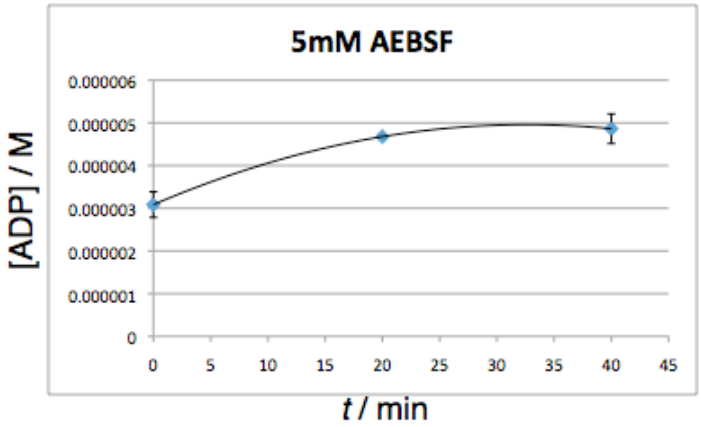


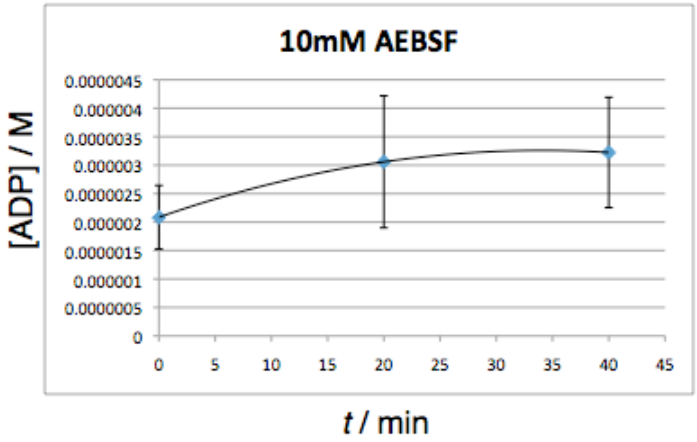
B)



C)


1mM AEBSF
[ADP] / M
0.000007
0.000006
0.000005
0.000004
0.000003
0.000002
0.000001
0
0 5 10 15 20 25 30 35 40 45
t / min


5mM AEBSF
[ADP] / M
0.000006
0.000005
0.000004
0.000003
0.000002
0.000001
0
0 5 10 15 20 25 30 35 40 45
t / min


10mM AEBSF
[ADP] / M
0.0000045
0.000004
0.0000035
0.000003
0.0000025
0.000002
0.0000015
0.000001
0.0000005
0
0 5 10 15 20 25 30 35 40 45
t / min

CHAPTER 7

ENZYMATIC BASIS OF RIBOSOMAL PEPTIDE PRENYLATION IN CYANOBACTERIA

Manuscript reproduced with permission from:

McIntosh, J. A., Donia, M. S., Nair, S. K. Schmidt, E. W. (2011) Enzymatic basis of ribosomal peptide prenylation in cyanobacteria, *J. Am. Chem. Soc.* Epub July 18, 2011.

Note: my contribution to this paper was in designing, performing and analyzing the results of all experiments with the exception of the phylogenetic analysis. Mass spectrometry data was generated by the University of Utah Mass Spectrometry and Proteomics core facility run by Chad Nelson and Krishna Parsawar.

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

pubs.acs.org/JACS

Enzymatic Basis of Ribosomal Peptide Prenylation in Cyanobacteria

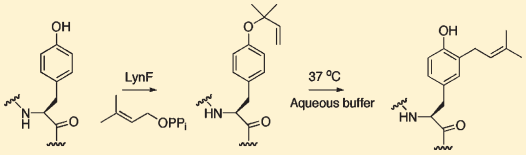
John A. McIntosh,[†] Mohamed S. Donia,[†,§] Satish K. Nair,[‡] and Eric W. Schmidt*[,†]

[†]Department of Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112, United States

[‡]Department of Biochemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: The enzymatic basis of ribosomal peptide natural product prenylation has not been reported. Here, we characterize a prenyltransferase, LynF, from the TruF enzyme family. LynF is the first characterized representative of the TruF protein family, which is responsible for both reverse- and forward-O-prenylation of tyrosine, serine, and threonine in cyclic peptides known as cyanobactins. We show that LynF reverse O-prenylates tyrosine in macrocyclic peptides. Based upon these results, we propose that the TruF family prenylates mature cyclic peptides, from which the leader sequence and other enzyme recognition elements have been excised. This differs from the common model of ribosomal peptide biosynthesis, in which a leader sequence is required to direct post-translational modifications. In addition, we find that reverse O-prenylated tyrosine derivatives undergo a facile Claisen rearrangement at 'physiological' temperature in aqueous buffers, leading to forward C-prenylated products. Although the Claisen rearrangement route to natural products has been chemically anticipated for at least 40 years, it has not been demonstrated as a route to prenylated natural products. Here, we show that the Claisen rearrangement drives phenolic C-prenylation in at least one case, suggesting that this route should be reconsidered as a mechanism for the biosynthesis of prenylated phenolic compounds.



INTRODUCTION

Prenylation is a common biochemical modification that has been studied in detail in numerous systems.[1,2] For example, proteins are often farnesylated or geranylgeranylated on cysteine residues, and numerous peptide natural products are known to be prenylated at diverse positions. Prenylation is key to the biological activity of these molecules.[3,4] Several enzyme families have been described that catalyze prenyl transfer, and indeed whole new prenyltransferase (PT) families continue to be discovered.[5−9] For example, the ABBA family has been shown to catalyze aromatic C-prenylation on a variety of substrates,[10] especially ortho to phenolic oxygen. Several ribosomal peptide natural products, such as ComX and the cyanobactins, are prenylated in unique ways that greatly increase the chemical diversity of the resulting compounds.[11−13] PTs for this growing ribosomal peptide natural product group have yet to be enzymatically characterized.[14]

Cyanobactins are a broadly distributed group of ribosomally derived, macrocyclic peptides whose biosynthetic genes are homologous. These compounds are often highly post-translationally modified. Indeed, many cyanobactins are prenylated by dimethylallyl pyrophosphate (DMAPP) on the oxygen atom of serine, threonine, or tyrosine (Figure 1), all of which are biochemically unprecedented reactions in the context of ribosomal peptides. Among cyanobactins, prenylation is known to occur in the "reverse" position (DMAPP 3-carbon) with serine and threonine and in the "forward" position (DMAPP 1-carbon) with tyrosine. However, no obvious candidate PTs exist in cyanobactin gene clusters.[11] By comparing several cyanobactin pathways discovered by metagenome sequencing, we proposed that the TruF family of proteins might be PTs.[11] Given that the TruF family lacks the typical sequence hallmarks of PTs, we sought to obtain biochemical evidence for this proposal. Consequently, we explored the set of sequenced gene clusters to find a soluble TruF relative for biochemical analysis.[15] Among TruF relatives, one protein, LynF from *Lyngbya aestuarii*,[16,17] could be solubly expressed in *Escherichia coli*. Although the other steps in cyanobactin biosynthesis have been characterized,[18−21] the PT substrate was unknown prior to this study. Consequently, substrate analogues for each possible biochemical step were synthesized, and their products upon reaction with LynF were analyzed (Figure 2). We show here that the LynF/TruF family represents a remarkably broad substrate family of O-prenyltransferases, with LynF prenylating a wide variety of tyrosyl peptides and phenol derivatives. Among biosynthetically relevant substrates, only cyclic peptides are prenylated.

Unexpectedly, reactions catalyzed by LynF led to carbon-prenylated phenolic products. Generally speaking, PTs catalyze electrophilic alkylation of their substrates. In the case of aromatic substrates, reactions are thought to proceed via electrophilic aromatic substitution.[10] However, several alternative mechanisms exist. For one, it has been proposed since at least the early 1970s that C-prenylated phenols might also arise via reverse O-prenylation of phenols followed by a Claisen rearrangement (Figure 1).[22] This chemical proposal was exploited in several elegant 'biomimetic' total

Received: June 13, 2011

   dx.doi.org/10.1021/ja205458h | *J. Am. Chem. Soc.* XXXX, XXX, 000–000

syntheses of complex natural products.[23,24] Despite a resurgence of biochemical studies of aromatic PTs in recent years,[6,8,10,25−30] the Claisen rearrangement proposal has not been revisited. Here, in addition to characterizing a novel PT family, we show that enzymatically synthesized aromatic C-prenylated phenols can indeed arise via the Claisen rearrangement pathway.

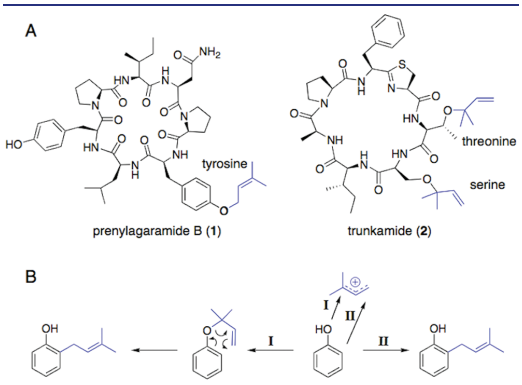


Figure 1. Prenylation in natural products. (A) Representative cyanobactin peptide natural products, showing groups derived from DMAPP in blue. (B) Two possible enzymatic mechanisms of phenol *ortho*-C-prenylation. First, DMAPP is dephosphorylated to yield a cation that can react either at oxygen (pathway I) or at carbon (pathyway II). In principle, a Claisen rearrangement from pathway I could then yield the C-prenylated product. Only pathway II has been previously linked to enzymatic modification.

RESULTS

Expression of LynF and Synthesis of Substrates. We found that several of the TruF-group PTs were difficult to overexpress. Fortuitously, in our search for candidate PTs, LynF (44% amino acid identity with TruF1), from the *lyn* pathway of the cyanobacterium *L. aestuarii*, was readily expressed in *E. coli* in soluble form. We predicted that LynF should prenylate phenols, but there are no known natural products of the *lyn* pathway.[16,17] Therefore, analogues of possible substrates were synthesized, including linear peptides representing putative pathway intermediates as well as *cyclo*[APMPPYP] (**6**), which is similar to the predicted *lyn* pathway product (Figure 2, Table 1). Additionally, reactions with several wholly unnatural phenol derivatives and linear peptides containing tyrosine were attempted. The purity and the identity of the synthetic substrates were assessed spectroscopically.

LynF is a Tyrosine PT. When incubated with DMAPP and $MgCl_2$, LynF catalyzed the prenylation of peptides and a subset of phenols related to tyrosine (Table 1). Mg^{2+} was added because many PTs require it for function.[10] Indeed, LynF does not function in the absence of either Mg^{2+} or Mn^{2+}. Tyrosine prenylation was demonstrated by Fourier-transform ion cyclotron resonance (FT-ICR) and MS/MS fragmentation, which revealed the presence of the expected ions (Table S2, Supporting Information) and also localized the prenylation to tyrosine (Figure S1, Supporting Information).

LynF prenylated several tyrosine-containing substrates and showed a strong preference for reaction with cyclic over linear peptides (Table 1). In contrast to typical ribosomal peptide biosynthesis,[31] LynF did not act on pathway intermediates (**3**−**5**) that still contained a leader sequence or enzyme "recognition elements" (Figures 2 and S1 and S2, Supporting Information). Of biosynthetically relevant

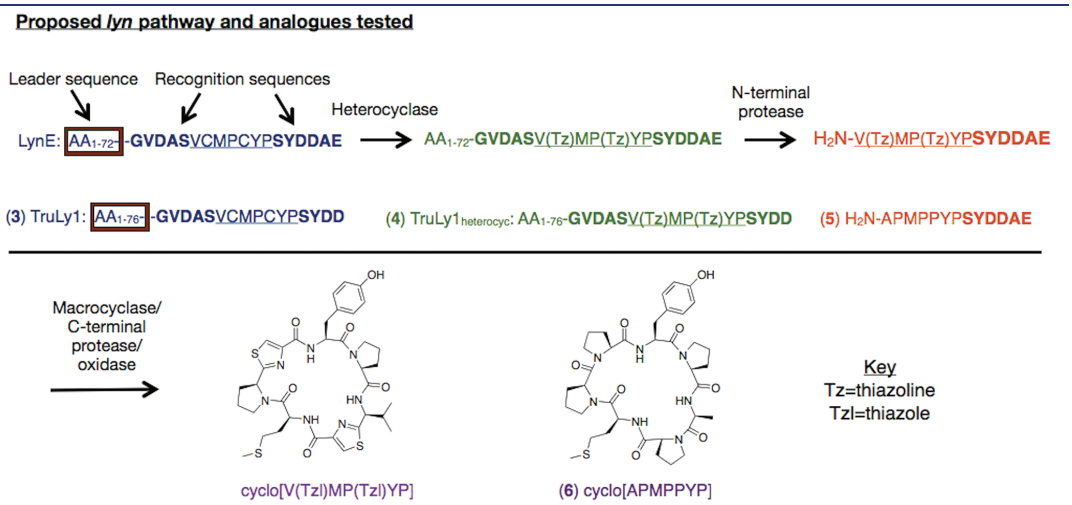


Figure 2. Defining the biosynthetic route to prenylated cyanobactins. Proposed biosynthetic scheme for Lyn pathway showing modification of precursor peptide by heterocyclization, proteolysis, macrocyclization, and prenylation. The first steps in this route are supported by previous enzymological studies, but the timing of prenylation was not known. Enzyme recognition elements are highlighted. Analogues (**3**−**6**) were used to assess prenylation of each possible biosynthetic intermediate and are shown in analogous colors beneath each proposed biosynthetic intermediate. Pro was substituted as an approximate isostere for thiazole in later analogues. No reaction was observed with any intermediate except the final cyclic peptide, showing that prenylation occurs at a late step, after all enzyme recognition elements have been excised.

dx.doi.org/10.1021/ja205458h |*J. Am. Chem. Soc.* XXXX, XXX, 000–000

Table 1. Substrates Assayed[a]

no.	substrate	yield
3	TruLy1 precursor peptide	NR
4	TruLy1 w/3 heterocycles	NR
5	APMPPYPSYDDAE	NR
6	*cyclo*[APMPPYP]	48%
7	APMPPYP	12%
8	*N*-acetyl APMPPYP	10%
9	*cyclo*[APMPPPAPMPPYP]	47%
10	*cyclo*[APMPPYPAPMPPYP]	43%
11	*cyclo*[KKPYILP]	37%
12	*cyclo*[KPYILP]	94%
13	KPYILP	1%
14	boc-L-Tyr	71%
15	boc-D-Tyr	66%
16	boc-4-cyano-L-Phe	NR
17	boc-O-allyl-L-Tyr	NR
18	boc-4-iodo-L-Phe	NR
19	boc-4-methoxy-L-Phe	NR
20	L-Phe	NR
21	*N*-acetyl-L-Tyr	3%
22	L-Tyr	NR
23	dopamine	NR
24	phenol	NR
25	L-Trp	NR
26	*cyclo*[QGGRGDWP]	NR
27	QGGRGDWPAYDGE	NR

[a] All yield quantitations are based on HPLC (for **6**−**8** and **14**−**25**) or MS (**3**−**5** and **9**−**13**) analyses of 24 h reactions and do not represent isolated yields. NR ("no reaction") denotes no detectable prenylation of tyrosine or derivatives. All substrates were run at 100 μM except **3**, **4**, and **9**, **10**, **11**, and **12** (14, 14, 70, 30, 20, and 20 μM, respectively). Both full (100 μM) and reduced concentrations (28, 56 μM) were employed with substrates **6** and **14** in order to compare data for substrates tested at reduced concentration.

substrates, LynF acted only on the mature cyclic peptides, such as **6**. Overall, LynF was capable of prenylating a broad variety of cyclic peptides, including substrates (**9**−**12**) containing 6-, 7-, 13-, and 14-amino acid residues. Surprisingly, LynF also prenylated several tyrosine-containing substrates that are not relevant to the natural biosynthetic pathway. Tyrosine itself was not a substrate, but boc-L-tyrosine and other N-terminally blocked tyrosine derivatives were readily prenylated (**14**, **15** and **21**). Derivatives lacking a free phenolic −OH were not substrates (**16**−**20**). Overall, LynF appears to require a blocked N-terminus but otherwise exhibits relaxed substrate specificity.

LynF Products are *ortho*-C-Prenylated. Based upon the known products of cyanobactin pathways, we expected that LynF would catalyze forward O-prenylation of phenol.[11,32] We performed large-scale enzymatic reactions with long incubation times to generate sufficient quantities of products for NMR analysis. Products of two different LynF substrates, **6** and **14**, were isolated by HPLC and characterized by NMR and high-resolution MS. Surprisingly, by comparison to previously described compounds,[33] we established that both **6** and **14** were forward C-prenylated, ortho to the phenolic hydroxyl group (Figures 3 and S3, Supporting Information).

C-Prenylation is the Result of a Claisen Rearrangement. Initially, we assumed that LynF catalyzed electrophilic aromatic

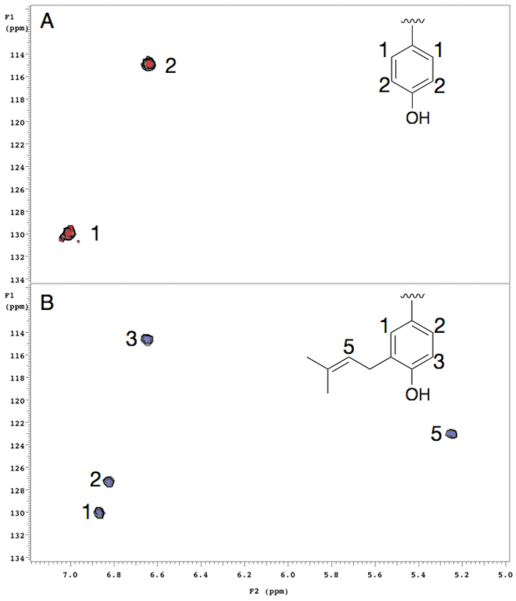
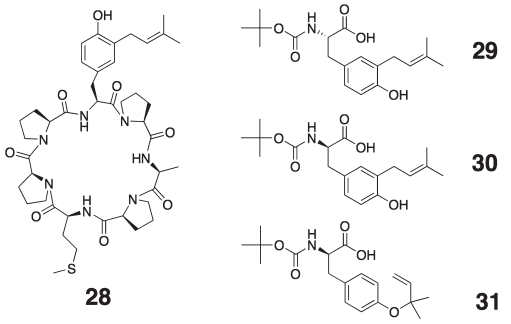


Figure 3. LynF is *ortho*-C-prenylating. The aromatic/olefin region of the heteronuclear quantum coherence NMR spectra is shown for (A) boc-tyrosine (**14**) and (B) its purified enzymatic product with LynF, clearly indicating a single, forward C-prenylation event. Similar spectra were also observed for reactions containing **6**. NMR and MS characterization of compounds is presented in Figures S1, S3, and S5, Supporting Information.

Scheme 1. NMR Characterized Products



substitution at the ortho position in a manner identical with that reported for ABBA PTs. With an eye toward constructing a Hammett plot of reactivity, we assayed the LynF catalyzed prenylation of a series of 4-substituted Tyr derivatives (**16**−**20**) (Table 1). However, all of these reactions failed. If LynF-catalyzed prenylation were to occur via electrophilic aromatic substitution, then one would expect a broader scope of reactivity, as was found for dimethylallyl tryptophan synthase.[34]

An alternative mechanism for the formation of *ortho*-C-prenylated phenols involves reverse O-prenylation followed by Claisen rearrangement of the resulting O-allyl intermediate.[22] In

dx.doi.org/10.1021/ja205458h | *J. Am. Chem. Soc.* XXXX, XXX, 000–000

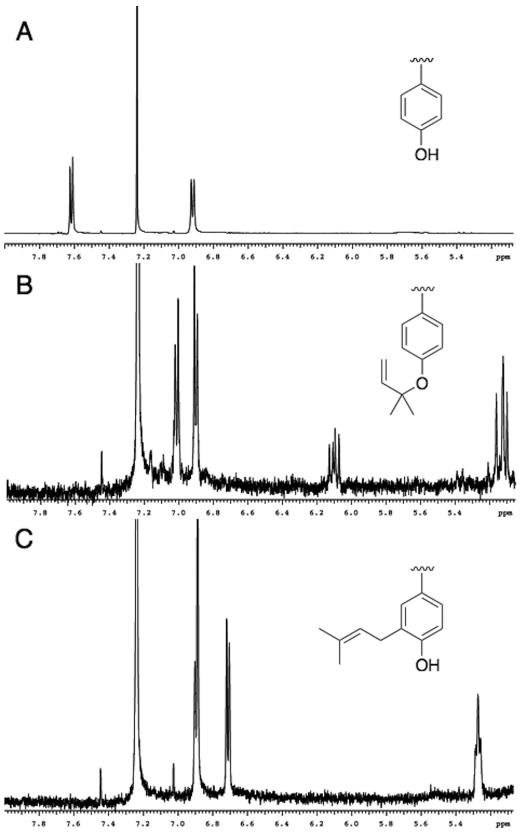


Figure 4. LynF catalyzes reverse O-prenylation of tyrosine. The aromatic/olefin region of ^{1}H NMR spectra are shown (A) for boc-tyrosine (**14**), (B) the HPLC-purified intermediate LynF product (**31**), and (C) the final reaction product (**30**). These spectra clearly indicate that the first product of the LynF reaction is reverse-O-prenyl tyrosine, which subsequently rearranges to give the C-prenylated product.

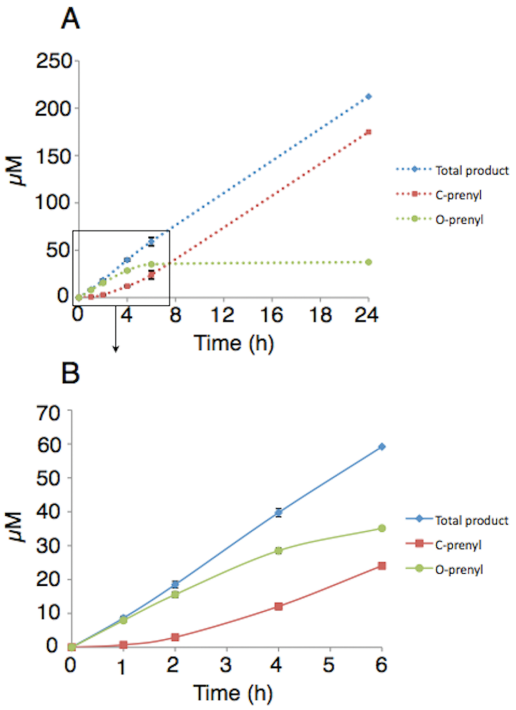


Figure 5. Time course of boc-L-Tyr (**14**) reaction followed by HPLC. The reaction was performed in quadruplicate, with variation indicated by error bars (A) and (B). Initial product of the reaction is almost exclusively O-prenylated as shown at 1, 2, and 4 h time points (B). After 4 h, the level of O-prenylated intermediate reaches steady-state and its levels are constant through 24 h, accompanied by steady increase in the concentration of C-prenylated final product **29**. This allowed kinetic constants for the Claisen rearrangement to be directly determined, since at steady state the concentration of the O-prenyl intermediate can be assumed to be a constant.

that vein, careful examination of LC-FT-ICR analyses of reaction mixtures showed that upon reaction with LynF, all substrates gave rise to two products (Figure S1, Supporting Information). These products were isobaric and prenylated on tyrosine. However, one product was prenylated on carbon and the other on oxygen. Forward carbon prenylation had been established using **28** (Scheme 1), which was purified and characterized by NMR as described above. Of the two products of LynF upon reaction with **6**, purified **28** was found to be identical to the early eluting product (Figure S4, Supporting Information). Moreover, no fragmentation of the C-prenyl moiety on **28** could be observed in MS-MS experiments. Similarly, for all LynF products, we observed that the early eluting compound was prenylated on tyrosine but did not lose isoprene in MS-MS, indicating C-prenylation.

In contrast, all late-eluting products evinced prominent loss of isoprene (C_5H_8) in their MS-MS spectra. Loss of C_5H_8 from prenylated phenols is diagnostic of O-prenylation, as shown in previous studies.[35−37] This reaction was more difficult to characterize by NMR, owing to the apparent instability of the O-prenylated products. Fortuitously, we were able to isolate one of these compounds, resulting from reaction of **15** (Figures 4 and S5, Supporting Information). NMR analysis of the purified material confirmed that the product (**31**) was O-prenylated and conclusively demonstrated that O-prenylation occurred in the reverse orientation. Purified **31** was then added to aqueous buffer at 37 °C (for buffer composition see Materials and Methods Section), and it rapidly and spontaneously rearranged to form the forward C-prenylated product **30** which was identical to the previously NMR-characterized product, **29**.

Reverse O-prenylated phenols are known to undergo the Claisen rearrangement to yield forward *ortho*-C-prenylated products.[22,24,38,39] Thus, we realized that if the sole enzymatic reaction catalyzed by LynF were reverse O-prenyltransfer on Tyr, then this would lead to the mixture of products we had consistently observed with all substrates. Alternatively, we supposed that LynF might carry out reverse O-prenyltransfer in addition to direct electrophilic aromatic substitution in the forward direction on carbon. To distinguish between these two possibilities, a kinetic analysis for reactions containing **14** was performed in which C- and O-prenyl products were followed over 24 h in reactions performed in quadruplicate (Figure 5). The O-prenyl product

dx.doi.org/10.1021/ja205458h |*J. Am. Chem. Soc.* XXXX, XXX, 000–000

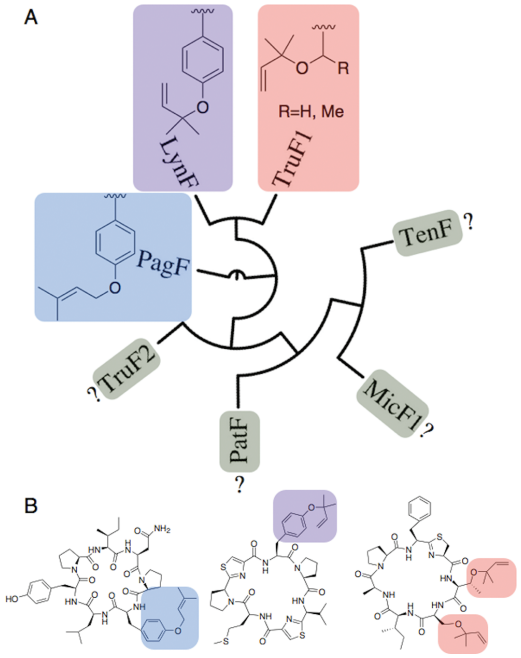


Figure 6. Phylogenetic analysis of the TruF/LynF family. (A) Tru F-group proteins cluster according to whether natural products are prenylated (top) or nonprenylated (bottom). Chemical products are shown for each TruF-like protein, where they exist. No TruF-like sequence relatives can be identified outside of cyanobactin gene clusters using either BLAST searching or sequence alignments with other PT family proteins, indicating that this is a novel group of PTs. (B) Actual product structures of prenylagaramide (left) and trunkamide (right) pathways; predicted enzymatic product of *lyn* pathway (middle) prior to Claisen rearrangement.

appeared first, with a delayed onset of **29**. After 4 h of reaction time, a steady state was reached in which the rate of prenylation was equal to the rate of the Claisen rearrangement. Interpretation of the kinetic data, which assumed a unimolecular mechanism and steady-state levels of O-prenyl intermediate, yielded a rate of rearrangement of 8.3 μM/h and a rate constant for the Claisen rearrangement (k where rate of Claisen = k*[O-prenyl intermediate]) of 0.23 h^{-1}. Taken together with the absence of reactivity observed with analogues lacking a free phenolic −OH, these data show that the initial enzymatic reaction is reverse O-prenylation, followed by slower conversion to a forward C-prenylated phenol

Having shown that the O-prenylated compound is the product of initial prenyltransfer, we sought to determine whether the rearrangement to the forward C-prenylated compound was enzyme catalyzed. To do so, purified O-prenylated **31** was added to enzyme, buffer, or boiled enzyme. Under all three conditions, **31** was efficiently converted into C-prenylated **30** without any appreciable enzymatic acceleration (Figure S6, Supporting Information).

This spontaneous Claisen rearrangement might seem surprising given that in the synthetic literature, reverse prenylated phenols require elevated temperatures for rearrangement.[24,38] However, these synthetic reactions take place in organic solvents, while we have employed aqueous solvents. Indeed, the speed with which we have observed reverse O-prenylated phenols to rearrange is unsurprising in light of the known aqueous acceleration of the Claisen rearrangement.[40−45] Here we show by experiment that the rearrangement goes to completion at 37 °C in aqueous buffers. Consequently, our conditions may provide a particularly mild reaction condition for the Claisen rearrangement for use in future synthetic studies.

To further examine the rapid rearrangement observed with **31**, we purchased O-allyl-boc-L-tyrosine and examined it to see if its rearrangement might be accelerated with enzyme or the buffer conditions employed with **31**. In contrast to the prenylated substrates, this compound did not undergo the Claisen rearrangement under the aforementioned conditions. Based upon these results, it seems that the geminal methyl groups adjacent to the phenolic oxygen are required to promote the Claisen rearrangement at relatively low temperatures. This effect can be rationalized as an example of the gem substituent effect, which is believed to accelerate the Claisen rearrangement.[46] Indeed, it has been shown that the presence of bulky substituents α- to oxygen can accelerate the Claisen rearrangement and similar reactions,[47,48] though to the best of our knowledge this is the first direct comparison of these substrates in the aromatic Claisen rearrangement.

Phylogenetic analysis of LynF and homologues nicely rationalizes the observed pattern of reactivity, where LynF is most closely related to TruF1 and PagF (Figure 6). In light of the above biochemical evidence, TruF1 can be assigned the role of a reverse O-prenyltransferase acting on Ser and Thr, while PagF can be assigned the role of a forward O-prenyltransferase acting on Tyr. Thus, that LynF would carry out reverse O-prenylation of Tyr is unsurprising in light of its phylogenetic profile.

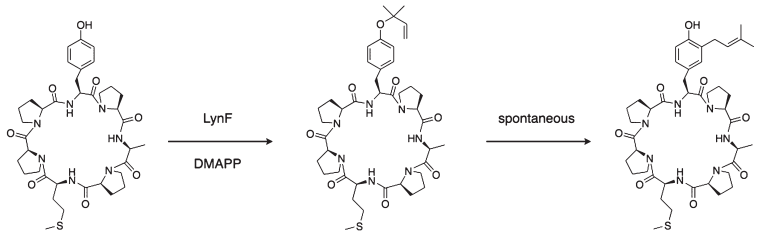
Kinetic Measurements. LynF catalyzed reaction rates were measured in triplicate using two different substrates: *cyclo*-[APMPPYP] (**6**), and an unnatural substrate, boc-L-tyrosine (**14**) (Figure S7, Supporting Information). Using HPLC analysis, the turnover numbers for **6** and **14** were similar (14 and 63 h^{-1}, respectively), as were K_m values (4 and 14 mM, respectively). These rates are slower than those typically reported for prenyltransferases.[27,49] However, the apparent slowness of reactions catalyzed by LynF is not unusual when compared with rates observed with other cyanobactin biosynthetic enzymes. Some of these reactions are quite slow, a fact that has been attributed to their extremely broad substrate tolerance and by extension their relatively low affinities for any given substrate.[18−21]

DISCUSSION

To the best of our knowledge, LynF represents the first enzymatically characterized PT leading to the synthesis of ribosomal peptide natural products. Further, serine and threonine O-PTs have not been previously described nor have tyrosine O-PTs acting on ribosomal peptides.[14] Although these post-translational modifications are currently known only in the cyanobactin family of natural products, cyanobactins are present in perhaps ~30% of all cyanobacteria on Earth and therefore constitute a major fraction of bioactive natural products globally.[50] Another salient feature of this enzyme group is that it clearly acts on polypeptide products, while most other natural product DMAPP transferases act on starting amino acids or on small dipeptides. For example, forward O-prenylated tyrosine has recently been characterized in sirodesmin diketopiperazine biosynthesis.[49,51] However, in this case tyrosine itself is the substrate for prenylation, and the product does not result from ribosomal synthesis.

dx.doi.org/10.1021/ja205458h |*J. Am. Chem. Soc.* XXXX, XXX, 000–000

Scheme 2. Claisen Rearrangement Pathway



Phenols themselves are C- or O-prenylated in many small molecule natural products.[10] In all cases that have been characterized so far, it is thought that C-prenylated phenols arise from direct electrophilic aromatic substitution.[10] Although the Claisen rearrangement was long-predicted from 'biomimetic' chemistry, a biochemical demonstration of its relevance as a route to C-prenylated phenols was lacking. Here, we show that the Claisen rearrangement route can indeed occur to afford C-prenylated products (Scheme 2). This route could easily be missed, since the O-prenylated intermediates are short-lived and not easily detected by commonly used analytical methods. For example, the intermediates have extremely weak absorption at $\lambda = 280$ nm, and their fluorescence spectra are different than for unsubstituted or C-prenylated phenols. The rearrangement is relatively rapid and continues even after enzymes have been denatured or inactivated. Given aqueous acceleration of the Claisen rearrangement and the acceleration provided by the geminal methyl groups of the reverse prenylation, the ease with which the Claisen rearrangement might occur in a cellular context has perhaps been underestimated. Overall, forward prenylation via electrophilic aromatic substitution and reverse O-prenylation followed by the Claisen rearrangement will be indistinguishable under many conditions.

We initially expected that the Claisen rearrangement might be enzymatically accelerated. In synthetic chemistry, several guanidinium-based synthetic catalysts of the Claisen rearrangement have been reported.[38,52] Additionally, in the premier example of a biological Claisen rearrangement, chorismate mutase has been calculated to provide rate enhancements of $>10^6$.[53] However, it is clear from our results using purified reverse-O-prenylated **31** that in this case the reaction is spontaneous. Given the spontaneous nature of this transformation and the similarity to other reported Claisen rearrangements, the simplest hypothesis is that **31** proceeds to **30** via the Claisen rearrangement and not via some other, more complicated mechanism.

We have previously shown that proteins in this group were involved in pathways to very sequence-diverse prenylated natural products in vivo,[16] and here we show that purified LynF accepts many different cyclic substrates. This broad specificity is especially remarkable in that LynF substrates share no common sequence features that would provide robust enzyme recognition elements. In ribosomal peptide natural product synthesis, enzymes commonly recognize conserved motifs in a leader peptide, which is subsequently cleaved and discarded, allowing the enzymes to modify diverse sequences.[31] However, in this case the reaction proceeds after the leader sequence and the recognition elements have already been removed.

The structure and catalytic mechanism of this new family of PTs remains to be determined. Although the proteins bear no homology to any other characterized protein outside of cyanobactin gene clusters, it remains possible that they are structurally related to known PTs. However, no putative ABBA-like required residues are present in the correct places in these proteins (Figure S8, Supporting Information). Since PTs are often deeply divergent and sequence similarity is completely lacking for this protein class, a final comparison will await structural study. It is also unknown why nonprenylating cyanobactin clusters usually contain (and even require) LynF-like proteins.[16,50] In these nonprenylating cases, all enzymatic roles have been assigned, so that LynF homologues serve no obvious enzymatic function.[20,21] However, removal of the LynF homologue from heterologous expression of the nonprenylating pat pathway in *E. coli* abolishes compound production.[16] Possible roles include a chaperone function or perhaps interaction with the leader sequence.

In conclusion, we show that the TruF/LynF group of proteins represents a new family of PTs that catalyze unprecedented enzymatic reactions and that are quite distinct from previously characterized proteins. LynF represents the first ribosomal peptide natural product prenyltransferase to be characterized, opening the door to the study of prenylated ribosomal peptide natural products.

MATERIALS AND METHODS

For detailed methods, see Materials and Methods in the Supporting Information.

Substrates. Dimethylallyl pyrophosphate (DMAPP) was synthesized following previously established procedures.[54−56] Excepting **11**, **12**, and **26** whose synthesis and characterization has been reported elsewhere,[15,19] peptide substrates were synthesized at the University of Utah DNA/peptide synthesis core facility. Synthesis of boc-protected 4-iodo-L-phenylalanine and 4-methoxy-L-phenylalanine was performed according to previously established procedures.[57] Boc-L-tyrosine, sodium hydrogen pyrophosphate, dimethylallyl bromide, tetrabutylammonium hydroxide, and dopamine HCl were purchased from Sigma. *N*-acetyl-L-tyrosine and phenol were purchased from Fisher Scientific. All other Tyr and Phe derivatives were purchased from ChemImpex.

Genes and Cloning. A codon-optimized version of *lynF* was synthesized and cloned into pET28 in frame with the N-terminal his-tag sequence using NdeI and EcoRI (Genscript). TruLy1 was cloned via modification of a previously described vector,[15] which was subsequently cloned into pET28b using NdeI and BamHI.

Protein Expression and Purification. LynF was expressed in BL21(DE3) cells, purified initially by Ni-NTA chromatography, which was followed by size-exclusion chromatography to yield homogeneous protein. Purification of TruLy1 was likewise performed by Ni-NTA chromatography, with the main difference being that rather than attempting to isolate soluble protein, TruLy1 was strongly overexpressed with the intent of driving the protein into inclusion bodies, after which time purification under denaturing conditions was performed.

dx.doi.org/10.1021/ja205458h |*J. Am. Chem. Soc.* XXXX, XXX, 000–000

Enzyme Assays. Enzyme reactions typically contained enzyme (3.8 μM) and variable substrate concentration (100 μM for most substrates; higher concentrations, i.e., 1 mM, were occasionally employed with boc-protected amino acid derivatives. Exceptions include substrates **11** and **12**, which were used at 20 μM final concentration as well as substrates **9** and **10**, which were used at 70 and 30 μM, respectively). Several additives (1 M of NaCl, 40 mM of glycylglycine pH of 9.0, 12 mM of $MgCl_2$, 3 mM of tris(2-carboxyethyl) phosphine (TCEP), and 1 mM of DMAPP) were added to all reactions. Reactions were incubated at 37 °C for 24 h in a DNA Engine Peltier thermocycler (Bio-Rad). Enzyme reactions with full-length precursor peptide contained TruLy1 (28 μM), ATP (0.8 mM), with or without heterocyclase enzyme TruD (90 nM), and additives as above. Controls were run to ensure that LynF was active in the presence of TruD and TruLy1 and vice versa. Products were characterized by MS or diode array (λ = 220 and 280 nm) and fluorescence (λ = 271 nm excitation and 303 nm emission) HPLC. Reactions assessing the rate of rearrangement of purified **31** were performed at 37 °C with time points taken at 0 and 8 h and included the standard additives described above. For descriptions of specific assays, see Materials and Methods in the Supporting Information.

Phylogenetic Tree Construction. The amino acid sequences of LynF homologues from the functionally characterized cyanobactin pathways were aligned using CLUSTALX. Maximum likelihood analysis with molecular clock PROMLK (PHYLIP) using the bootstrap test method (1000 replicates) was performed to assess the phylogenetic relationship between the different homologues. The same tree branches were also supported using other phylogenetic experiments such as maximum parsimony (MEGA 4.0) using 1000 bootstrap replicates.

General Methods. ESI-MS and FT-ICR analyses were performed at the University of Utah Mass Spectrometry and Proteomics core facility. MALDI-MS analyses were performed on a Micromass MALDI micro MX instrument (Waters). HPLC separations were performed on a LaChrom Elite system (Hitachi). NMR spectra were collected on either 400 or 500 MHz spectrometers (Varian). CD spectra were collected on a Jasco J-815 spectrometer, and data were plotted in Excel.

■ ASSOCIATED CONTENT

Supporting Information. Additional mass spectrometry, NMR, kinetic, protein purification, sequence, and substrate data, and full methods. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
ews1@utah.edu

Present Addresses
[§]Department of Bioengineering and Therapeutic Sciences and California Institute for Quantitative Biosciences, University of California, San Francisco, San Francisco, CA 94158

■ ACKNOWLEDGMENT

This work was supported by NIH GM071425. We thank C. Dale Poulter, Jeffrey Rudolph, Gary E. Keck, and John Heemstra for helpful discussions; Chad Nelson, Krishna Parsawar, and Jim Muller for mass spectrometry assistance; Scott Endicott and Robert Schackmann for peptide synthesis; Seth Lilavivat for circular dichroism assistance; and Jack Skalicky, Jay Olsen, Dai Tianero, and Zhenjian Lin for NMR assistance. We dedicate this paper to the late Prof. D. John Faulkner for his pioneering work on the Claisen rearrangement and marine natural products.

■ REFERENCES

(1) Brandt, W.; Bräuer, L.; Günnewich, N.; Kufka, J.; Rausch, F.; Schulze, D.; Schulze, E.; Weber, R.; Zakharova, S.; Wessjohann, L. *Phytochemistry* **2009**, *70*, 1758–1775.
(2) Nguyen, U. T.; Goody, R. S.; Alexandrov, K. *ChemBioChem* **2010**, *11*, 1194–1201.
(3) Gibbs, R. A. *Nat. Chem. Biol.* **2005**, *1*, 7–8.
(4) Casey, P. J.; Seabra, M. C. *J. Biol. Chem.* **1996**, *271*, 5289–5292.
(5) Sacchettini, J. C.; Poulter, C. D. *Science* **1997**, *277*, 1788–1789.
(6) Edwards, D. J.; Gerwick, W. H. *J. Am. Chem. Soc.* **2004**, *126*, 11432–11433.
(7) Reiss, Y.; Goldstein, J. L.; Seabra, M. C.; Casey, P. J.; Brown, M. S. *Cell* **1990**, *62*, 81–88.
(8) Kuzuyama, T.; Noel, J. P.; Richard, S. B. *Nature* **2005**, *435*, 983–987.
(9) Gebler, J. C.; Poulter, C. D. *Arch. Biochem. Biophys.* **1992**, *296*, 308–313.
(10) Saleh, O.; Haagen, Y.; Seeger, K.; Heide, L. *Phytochemistry* **2009**, *70*, 1728–1738.
(11) Donia, M. S.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2008**, *4*, 341–343.
(12) Carroll, A. R.; Coll, J. C.; Bourne, D. J.; MacLeod, J. K.; Zabriskie, T.; Ireland, C. M.; Bowden, B. F. *Aust. J. Chem.* **1996**, *49*, 659–667.
(13) Okada, M.; Sato, I.; Cho, S. J.; Iwata, H.; Nishio, T.; Dubnau, D.; Sakagami, Y. *Nat. Chem. Biol.* **2005**, *1*, 23–24.
(14) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *Nat. Prod. Rep.* **2009**, *26*, 537–559.
(15) Donia, M. S.; Schmidt, E. W. *Chem. Biol.* **2011**, *18*, 508–519.
(16) Donia, M. S.; Hathaway, B. J.; Sudek, S.; Haygood, M. G.; Rosovitz, M. J.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2006**, *2*, 729–735.
(17) Leikowski, N.; Fewer, D. P.; Sivonen, K. *Appl. Environ. Microbiol.* **2009**, *75*, 853–857.
(18) McIntosh, J. A.; Schmidt, E. W. *ChemBioChem* **2010**, *11*, 1413–1421.
(19) McIntosh, J. A.; Robertson, C. R.; Vinayak, A.; Satish, N. K.; Bulaj, G. W.; Schmidt, E. W. *J. Am. Chem. Soc.* **2010**, *132*, 15499–15501.
(20) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *J. Am. Chem. Soc.* **2010**, *132*, 4089–4091.
(21) Lee, J.; McIntosh, J. A.; Hathaway, B. J.; Schmidt, E. W. *J. Am. Chem. Soc.* **2009**, *131*, 2122–2214.
(22) Quillinan, A. J.; Scheinmann, F. *J. Chem. Soc. D* **1971**, 966–967.
(23) Tisdale, E. J.; Slobodov, I.; Theodorakis, E. A. *Org. Biomol. Chem.* **2003**, *1*, 4418–4422.
(24) Nicolaou, K. C.; Li, J. *Angew. Chem., Int. Ed.* **2001**, *40*, 4264–4268.
(25) Jost, M.; Zocher, G.; Tarcz, S.; Matuscheck, M.; Xie, X.; Li, S. M.; Stehle, T. *J. Am. Chem. Soc.* **2010**, *132*, 17849–17858.
(26) Haagen, Y.; Unsöld, I.; Westrich, L.; Gust, B.; Richard, S. B.; Noel, J. P.; Heide, L. *FEBS Lett.* **2007**, *581*, 2889–2893.
(27) Ding, Y.; de Wet, J. R.; Cavalcoli, J.; Li, S.; Greshock, T. J.; Miller, K. A.; Finefield, J. M.; Sunderhaus, J. D.; McAfoos, T. J.; Tsukamoto, S.; Williams, R. M.; Sherman, D. H. *J. Am. Chem. Soc.* **2010**, *132*, 12733–12740.
(28) Schultz, A. W.; Lewis, C. A.; Luzung, M. R.; Baran, P. S.; Moore, B. S. *J. Nat. Prod.* **2010**, *73*, 373–377.
(29) Balibar, C. J.; Howard-Jones, A. R.; Walsh, C. T. *Nat. Chem. Biol.* **2007**, *3*, 584–592.
(30) Pojer, F.; Wemakor, E.; Kammerer, B.; Chen, H.; Walsh, C. T.; Li, S. M.; Heide, L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 2316–2321.
(31) Oman, T. J.; van der Donk, W. A. *Nat. Chem. Biol.* **2010**, *6*, 9–18.
(32) Murakami, M.; Itou, Y.; Ishida, K.; Shin, H. J. *J. Nat. Prod.* **1999**, *62*, 725–755.
(33) Ahmed, F.; Ohtsuki, T.; Aida, W.; Ishibashi, M. *J. Nat. Prod.* **2008**, *71*, 1963–1966.
(34) Gebler, J. C.; Woodside, A. B.; Poulter, C. D. *J. Am. Chem. Soc.* **1992**, *114*, 7354–7360.
(35) Donia, M. S.; Schmidt, E. W. *Chem. Biol.* **2011**, *18*, 508–519.

(36) Sobolev, V. S.; Potter, T. L.; Horn, B. W. *Phytochem. Anal.* **2006**, *17*, 312–322.
(37) Miranda, C. L.; Stevens, J. F.; Ivanov, V.; McCall, M.; Frei, B.; Deinzer, M. L.; Buhler, D. R. *J. Agric. Food Chem.* **2000**, *48*, 3876–3884.
(38) Uyeda, C.; Jacobsen, E. N. *J. Am. Chem. Soc.* **2008**, *130*, 9228–9229.
(39) Hoarau, C.; Pettus, R. R. T. *Synlett* **2003**, 127–137.
(40) Bagnell, L.; Cablewski, T.; Strauss, C. R.; Trainor, R. W. *J. Org. Chem.* **1996**, *61*, 7355–7359.
(41) Narayan, S.; Muldoon, J.; Finn, M. G.; Fokin, V. V.; Kolb, H. C.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2005**, *44*, 3275–3279.
(42) Majumdar, K. C.; Alam, S.; Chattopadhyay, B. *Tetrahedron* **2008**, *64*, 597–643.
(43) Knowles, J. R.; Copley, S. D. *J. Am. Chem. Soc.* **1987**, *109*, 5008–5013.
(44) Grieco, P. A. *Aldrichimica Acta* **1991**, *24*, 59–66.
(45) White, W. N.; Wolfarth, E. U. *J. Org. Chem.* **1970**, *35*, 2196–2199.
(46) Jung, M. E.; Piizzi, G. *Chem. Rev.* **2005**, *105*, 1735–1766.
(47) Harfenist, M.; Thom, E. *J. Org. Chem.* **1972**, *37*, 841–848.
(48) Goering, H. L.; Jacobson, R. R. *J. Am. Chem. Soc.* **1958**, *80*, 3277–3285.
(49) Zou, H. X.; Xie, X.; Zheng, X. D.; Li, S. M. *Appl. Microbiol. Biotechnol.* **2011**, *89*, 1443–1451.
(50) Donia, M. S.; Schmidt, E. W.*Comprehensive Natural Products Chemistry II*; Mander, L., Liu, H.-W., Eds.; Elsevier: Oxford, U.K., 2008; Vol. 2, pp 539558.
(51) Kremer, A.; Li, S. M. *Microbiology* **2010**, *156*, 278–286.
(52) Uyeda, C.; Rötheli, A. R.; Jacobsen, E. N. *Angew. Chem., Int. Ed.* **2010**, *49*, 9753–9756.
(53) Andrews, P. R.; Smith, G. D.; Young, I. G. *Biochemistry* **1973**, *12*, 3492–3498.
(54) Woodside, A. B.; Huang, Z.; Poulter, C. D. *Org. Synth.* **1988**, *66*, 211–215.
(55) Woodside, A. B.; Huang, Z.; Poulter, C. D. *Org. Synth.* **1993**, *8*, 616–620.
(56) Davisson, V. J.; Woodside, A. B.; Neal, T. R.; Stremler, K. E.; Muehlbacher, M.; Poulter, C. D. *J. Org. Chem.* **1986**, *51*, 4768–4779.
(57) Stankova, I. G.; Videnov, G. I.; Golovinsky, E. V.; Jung, G. *J. Peptide Sci.* **1999**, *5*, 392–398.

dx.doi.org/10.1021/ja205458h |*J. Am. Chem. Soc.* XXXX, XXX, 000–000

Supporting Online Material for:

Enzymatic basis of ribosomal peptide prenylation in cyanobacteria

John A McIntosh,† Mohamed S. Donia,† Satish K. Nair,# Eric W. Schmidt†*

† Department of Medicinal Chemistry, University of Utah, Salt Lake City, UT 84112 USA
# Department of Biochemistry, University of Illinois at Urbana-Champaign, IL 61801 USA

*ews1@utah.edu

This file contains:

Figures S1-S9

Materials and methods

Tables S1, S2

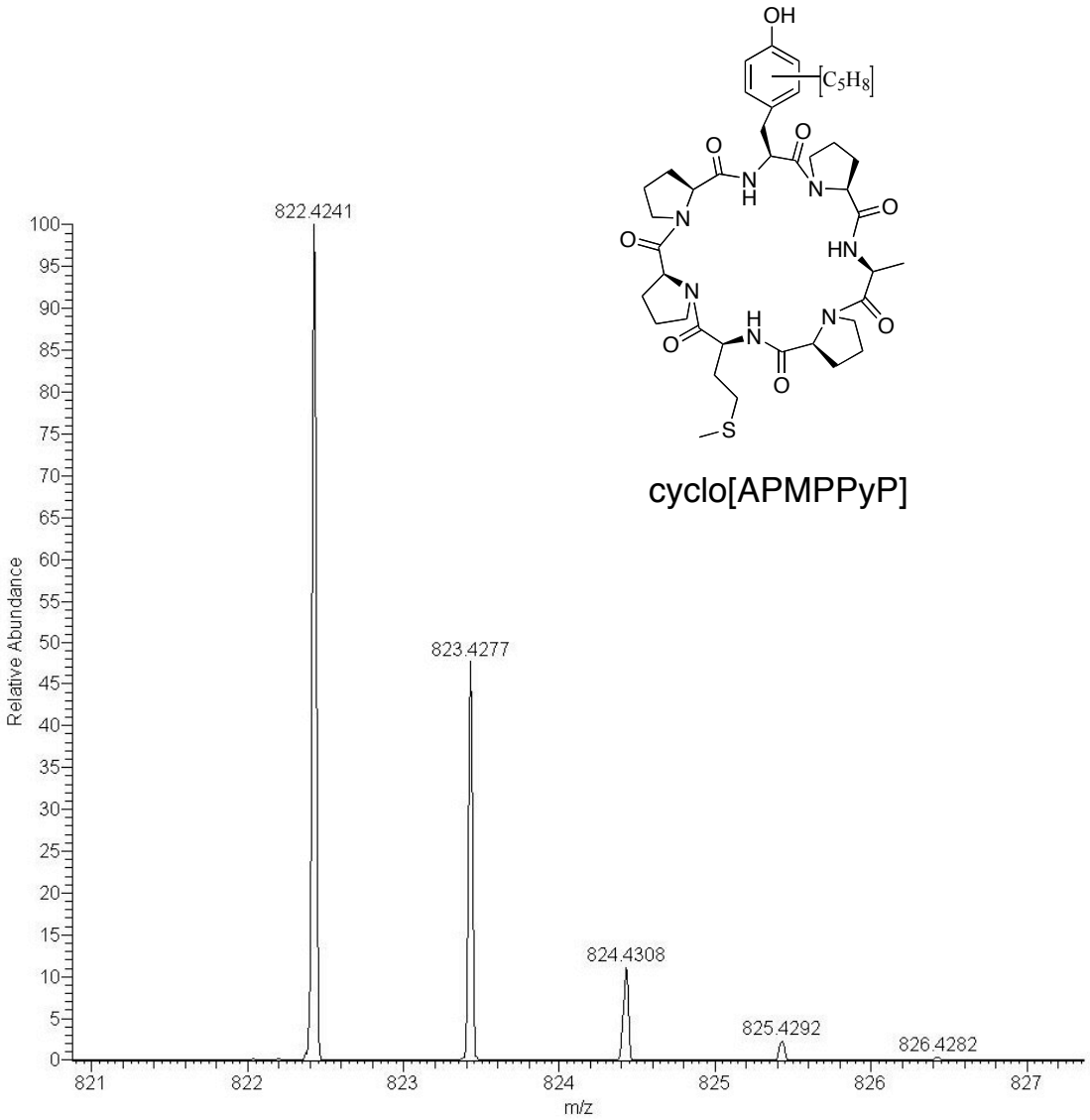
Supporting references

Table of Contents

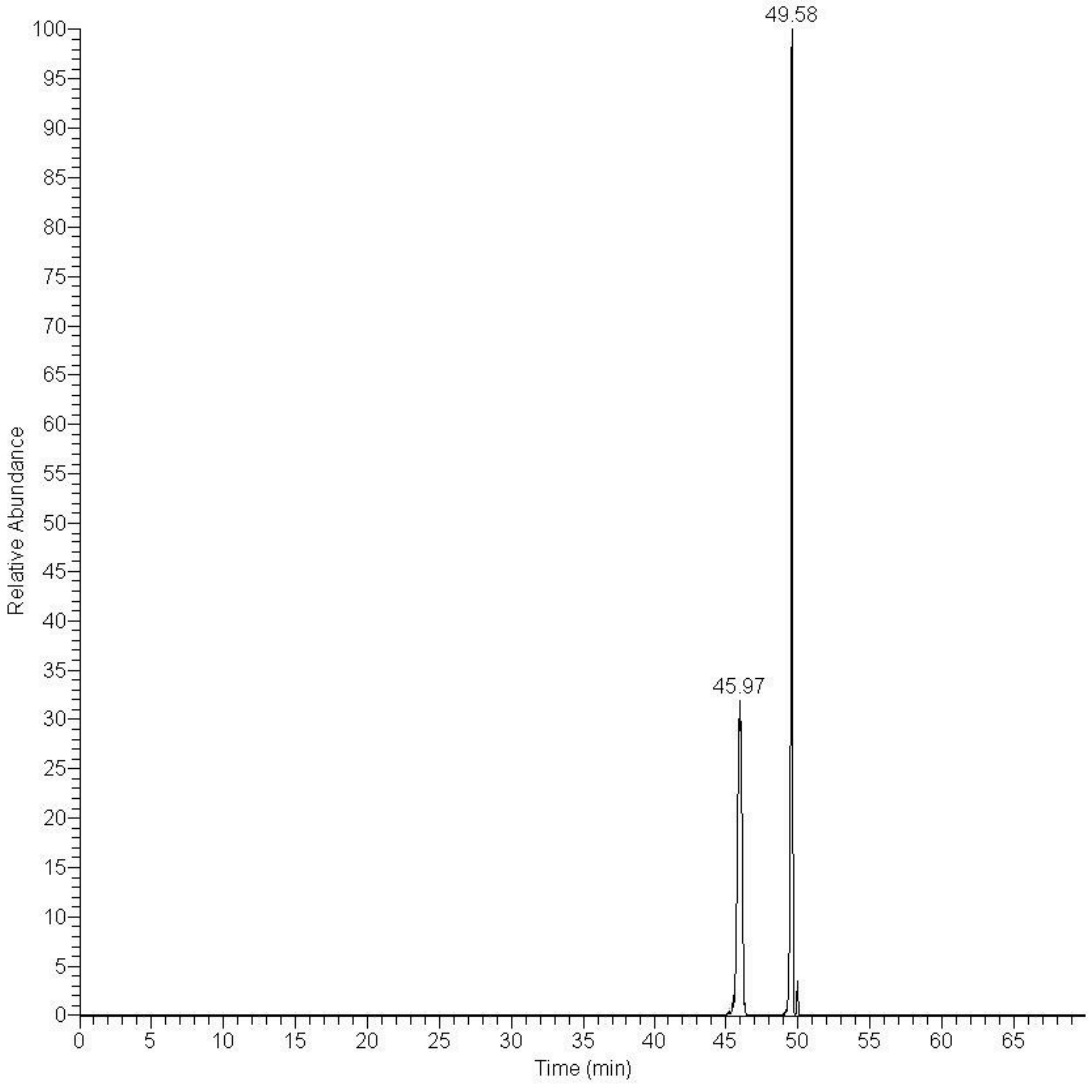
Table of Contents [continued]

Figure S1. LC-FT-ICR and MS-MS, or ESI-MS characterization of products of reactions containing: (a-d) cyclo[APMPPYP] (e-h) APMPPYP (i-l) cyclo[APMPPYPAPMPPP] (m-p), (q-t) cyclo[APMPPYPAPMPPYP] (u-x) cyclo[KKPYILP] (y-bb) cyclo[KPYILP] (cc-ff) KPYILP (gg-hh) boc-L-Tyr (ii) N-acetyl-L-Tyr. Note that prenylated tyrosine is denoted by a lower-case 'y', while unmodified tyrosine is denoted by an upper-case 'Y'

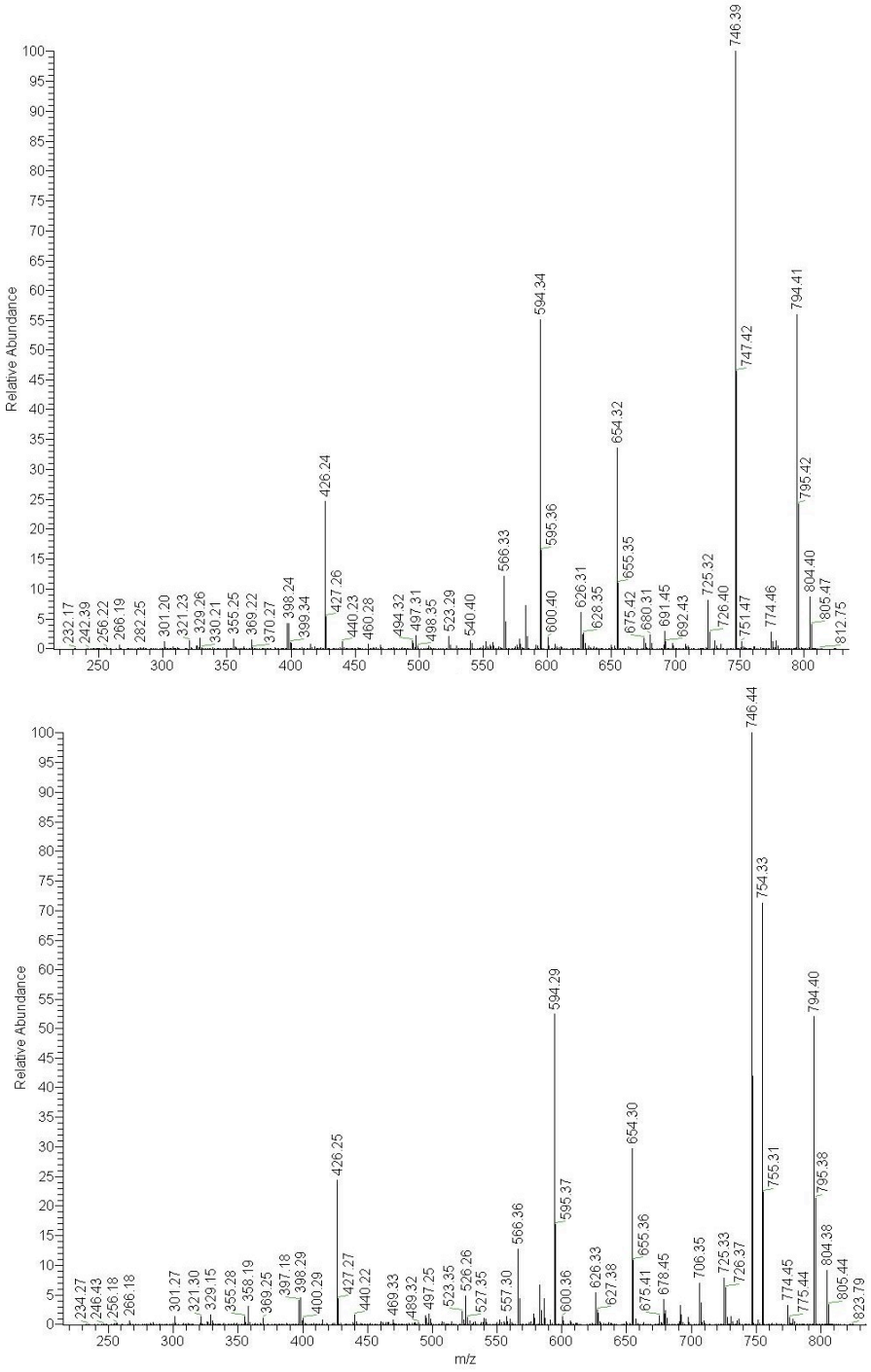
(a) FT-ICR data showing ion whose mass (+/- 2 ppm) corresponds to prenylated cyclo[APMPPYP]



(b) LC-FT-ICR chromatogram selected for mass of prenylated cyclo[APMPPYP]



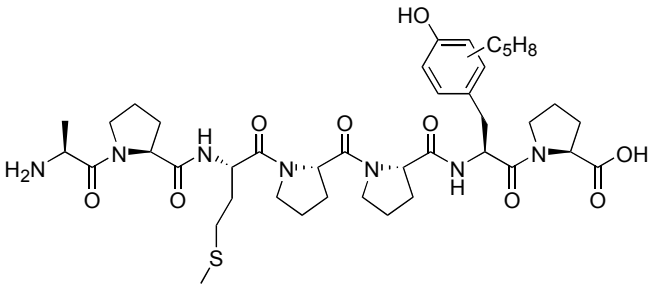
(c) MS-MS spectra for early (top) and late (bottom) eluting products of the reaction of LynF with cyclo[APMPPYP] (see chromatogram on previous page)



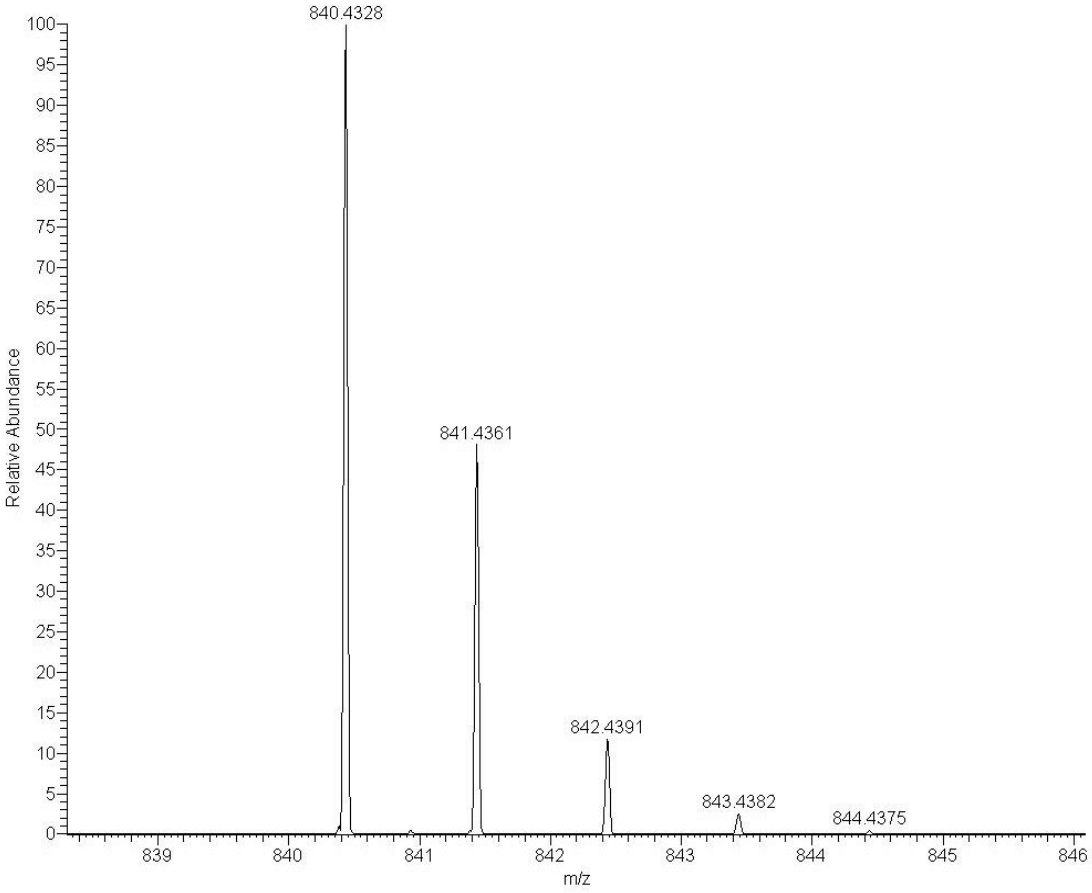
(d) MS-MS assignments for prenylated cyclo[APMPPYP] early (top) and late (bottom) eluting isomers

cyclo[APMPPYP] 1st peak (C-prenyl)					
y- and b- cleavages (+1)	Expected	Observed	y- and b- cleavages (+1)	Expected	Observed
AP	169.10	-	yP	329.19	329.26
APM	300.14	-	yPA	400.22	400.29
APMP	397.19	397.16	yPAP	497.28	497.31
APMPP	494.24	494.32	yPAPM	628.32	628.35
APMPPy	725.37	725.32	yPAPMP	725.37	725.32
PM	229.10	-	PA	169.10	-
PMP	326.15	-	PAP	266.15	266.19
PMPP	423.21	-	PAPM	397.19	-
PMPPy	654.33	654.32	PAPMP	494.24	494.32
PMPPyP	751.39	751.47	PAPMPP	591.30	591.35
MP	229.10	-			
MPP	326.15	-	Miscellaneous Ions		
MPPy	557.28	557.43	-H2O	804.41	804.40
MPPyP	654.33	654.32	-Acylium (C=O)	794.42	794.41
MPPyPA	725.37	725.32	-Methyl sulfide	774.42	774.46
PP	263.18	-	-Met side-chain	746.39	746.39
PPy	426.24	426.24	PyPAP y- a- cleavage	566.33	566.33
PPyP	523.29	523.29			
PPyPA	594.33	594.34			
PPyPAP	691.38	691.45			
Py	329.19	329.26			
PyP	426.24	426.24			
PyPA	497.28	497.31			
PyPAP	594.33	594.34			
PyPAPM	725.37	725.32			
cyclo[APMPPYP] 2nd peak (O-prenyl)					
y- and b- cleavages (+1)	Expected	Observed	y- and b- cleavages (+1)	Expected	Observed
AP	169.10	-	yP	329.19	329.15
APM	300.14	-	yPA	400.22	400.29
APMP	397.19	397.18	yPAP	497.28	497.25
APMPP	494.24	494.28	yPAPM	628.32	628.34
APMPPy	725.37	725.33	yPAPMP	725.37	725.33
PM	229.10	-	PA	169.10	-
PMP	326.15	326.19	PAP	266.15	266.18
PMPP	423.21	-	PAPM	397.19	397.18
PMPPy	654.33	654.30	PAPMP	494.24	494.28
PMPPyP	751.39	751.35	PAPMPP	591.30	591.34
MP	229.10	-			
MPP	326.15	326.19	Miscellaneous Ions		
MPPy	557.28	557.30	-H2O	804.41	804.38
MPPyP	654.33	654.30	-Acylium (C=O)	794.42	794.40
MPPyPA	725.37	725.33	-Methyl sulfanyl radical	774.42	774.45
PP	263.18	-	-Prenyl	754.35	754.33
PPy	426.24	426.25	-Met side-chain	746.39	746.44
PPyP	523.29	523.34	PyPAP y- a- cleavage	566.33	566.36
PPyPA	594.33	594.29	PPyPA -Prenyl	526.27	526.26
PPyPAP	691.38	691.44	PPy, PyP -Prenyl	358.18	358.19
Py	329.19	329.15			
PyP	426.24	426.25			
PyPA	497.28	497.25			
PyPAP	594.33	594.29			
PyPAPM	725.37	725.33			

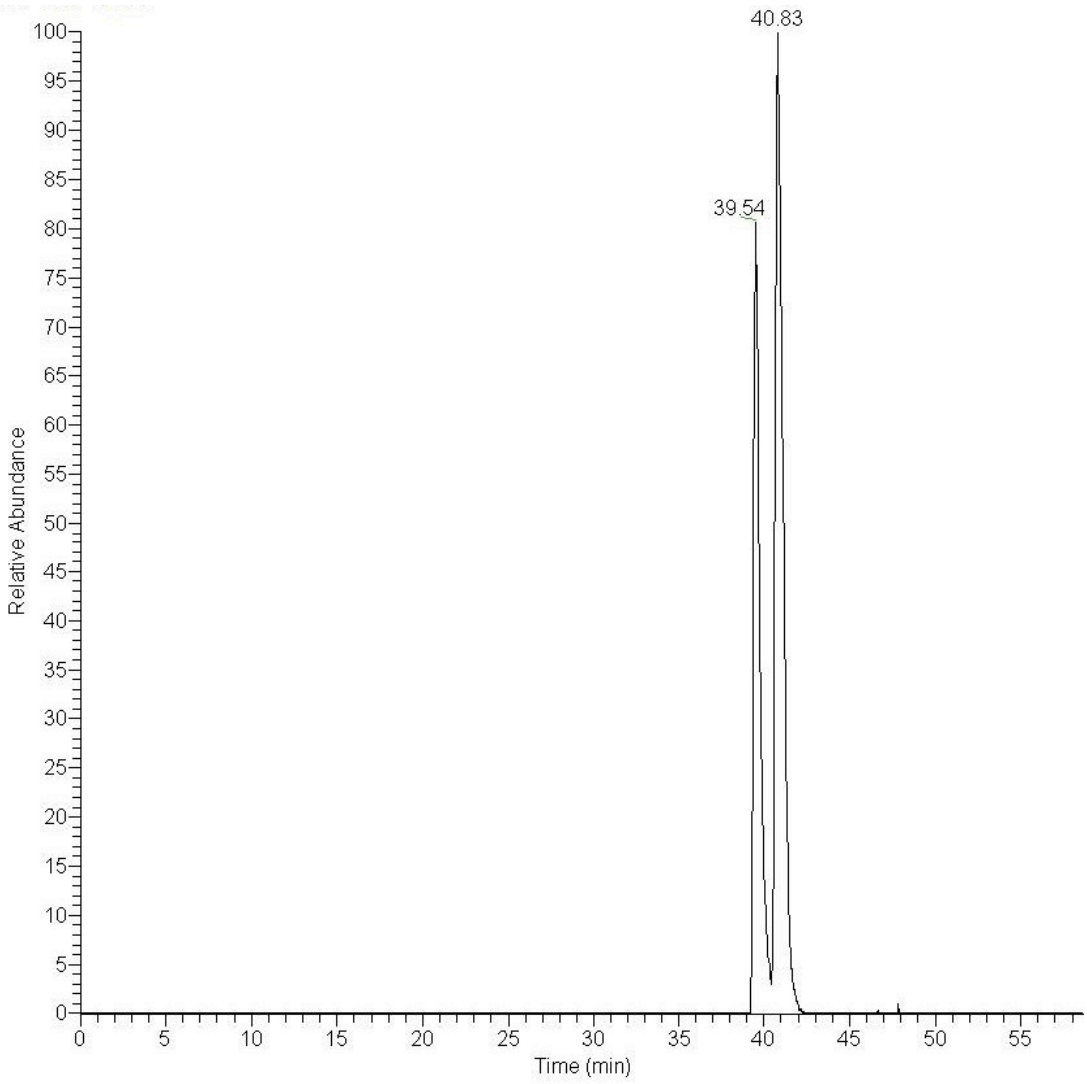
(e) FT-ICR data showing ion whose mass (+/- 2 ppm) corresponds to prenylated APMPPYP



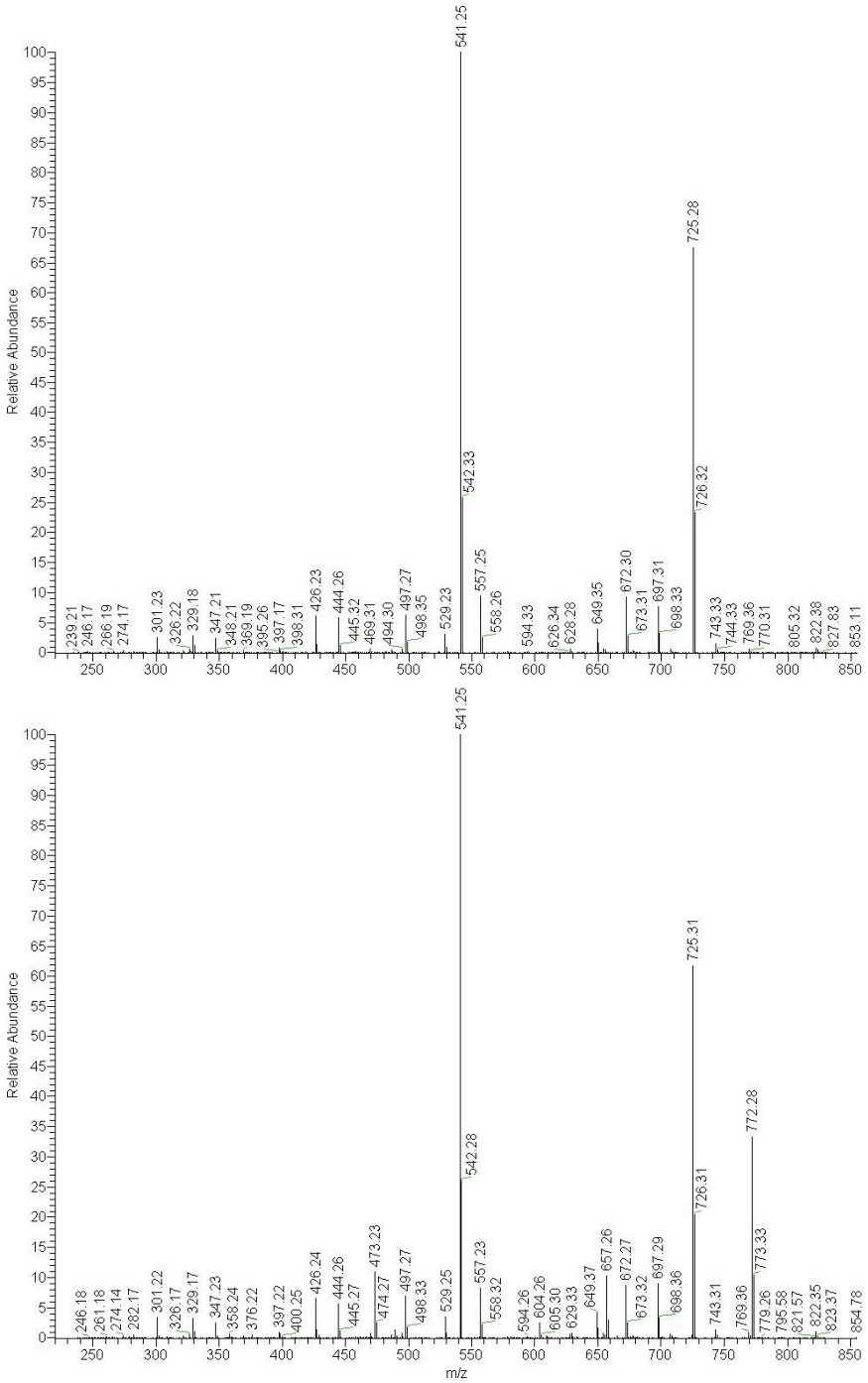
APMPPyP



(f) LC-FT-ICR chromatogram selected for mass of prenylated APMPPYP



(g) MS-MS spectra for early (top) and late (bottom) eluting products of the reaction of LynF with APMPPYP (see chromatogram on previous page)

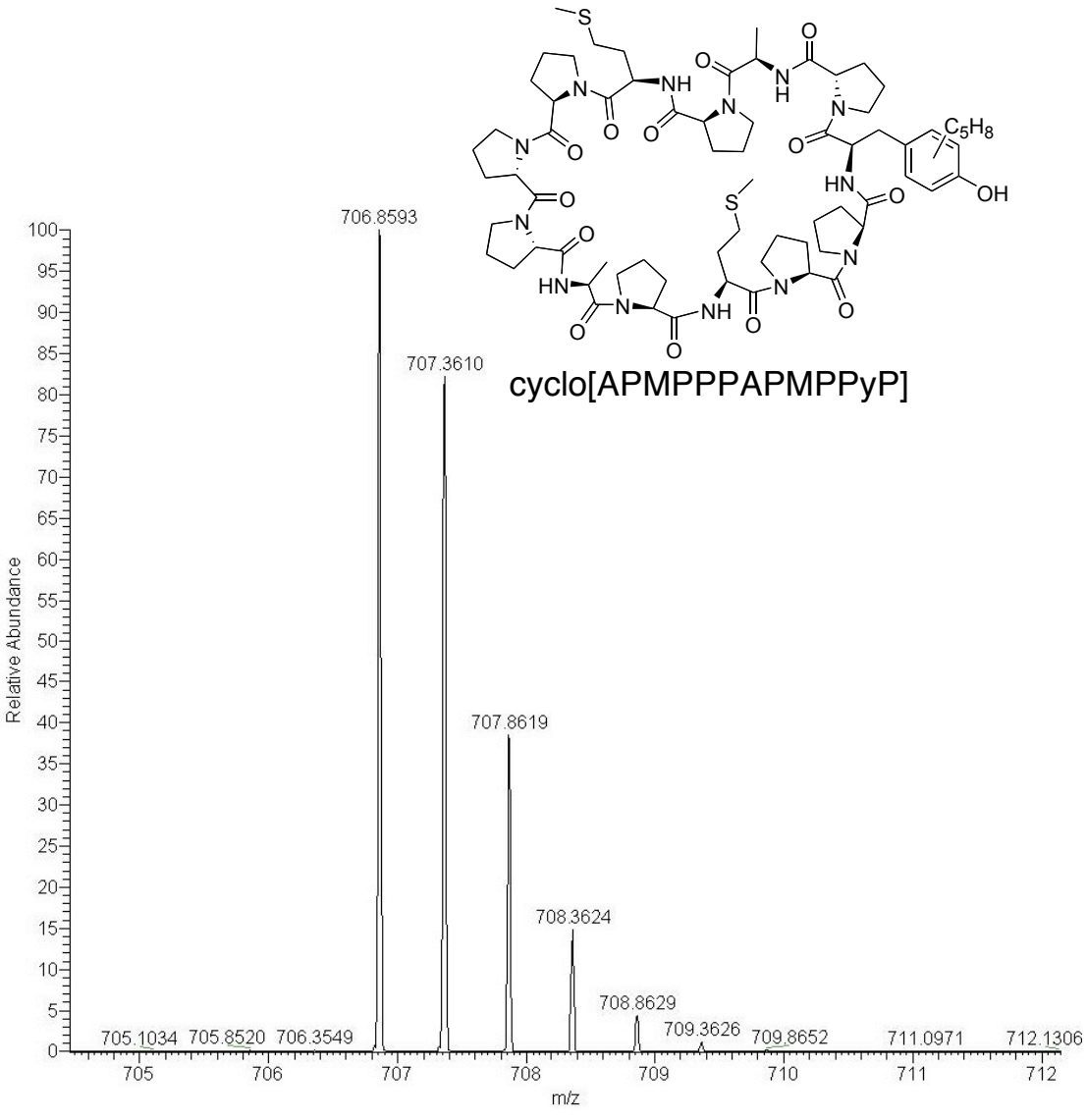


(h) MS-MS assignments for prenylated APMPPYP early (top) and late (bottom) eluting isomers

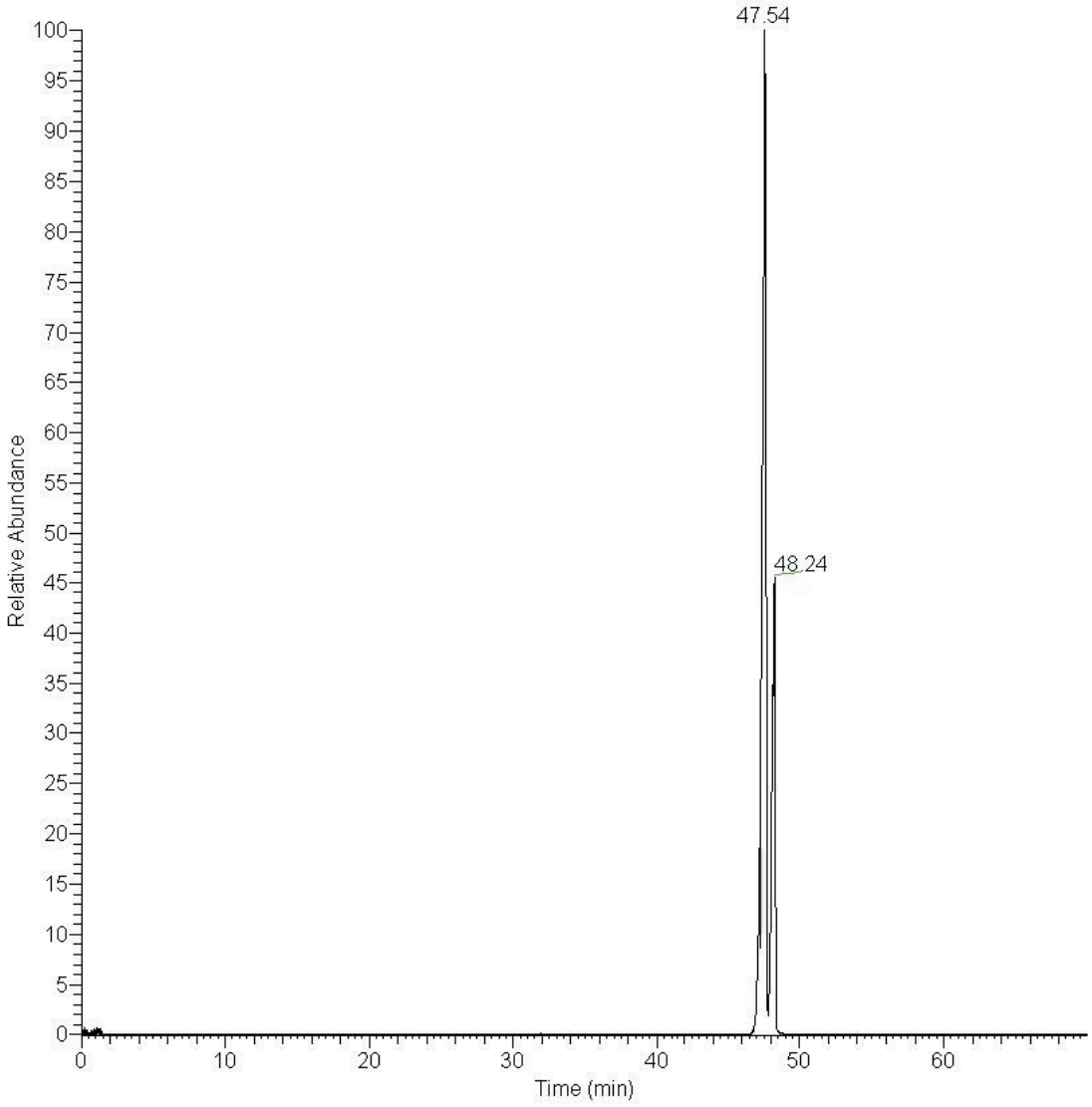
APMPPYP 1st peak (C-prenyl)		
b-ions (+1)	**Expected**	**Observed**
AP	169.10	-
APM	300.14	-
APMP	397.19	397.22
APMPP	494.24	494.33
APMPPy	725.37	725.31
y-ions (+1)		
PMPPyP	769.39	769.36
MPPyP	672.34	672.27
PPyP	541.30	541.25
PyP	444.25	444.26
Py	347.19	347.23
y- b- cleavages (+1)		
Py	329.19	329.17
PPy	426.24	426.24
MPPy	557.28	557.25
y- a- cleavages (+1)		
Py	301.19	301.22
Miscellaneous Ions		
APMPPy a-cleavage +1	697.37	697.31

APMPPYP 2nd peak (O-prenyl)		
b-ions (+1)	**Expected**	**Observed**
AP	169.10	-
APM	300.14	-
APMP	397.19	397.22
APMPP	494.24	-
APMPPy	725.37	725.31
y-ions (+1)		
PMPPyP	769.39	769.36
MPPyP	672.34	672.27
PPyP	541.30	541.25
PyP	444.25	444.26
Py	347.19	347.23
y- b- cleavages (+1)		
Py	329.19	329.17
PPy	426.24	426.24
MPPy	557.28	557.23
y- a- cleavages (+1)		
Py	301.19	301.22
Miscellaneous Ions		
All -prenyl	772.37	772.28
APMPPy a-cleavage +1	697.37	697.29
APMPPy -prenyl	657.31	657.26
PPyP -prenyl	473.24	473.23

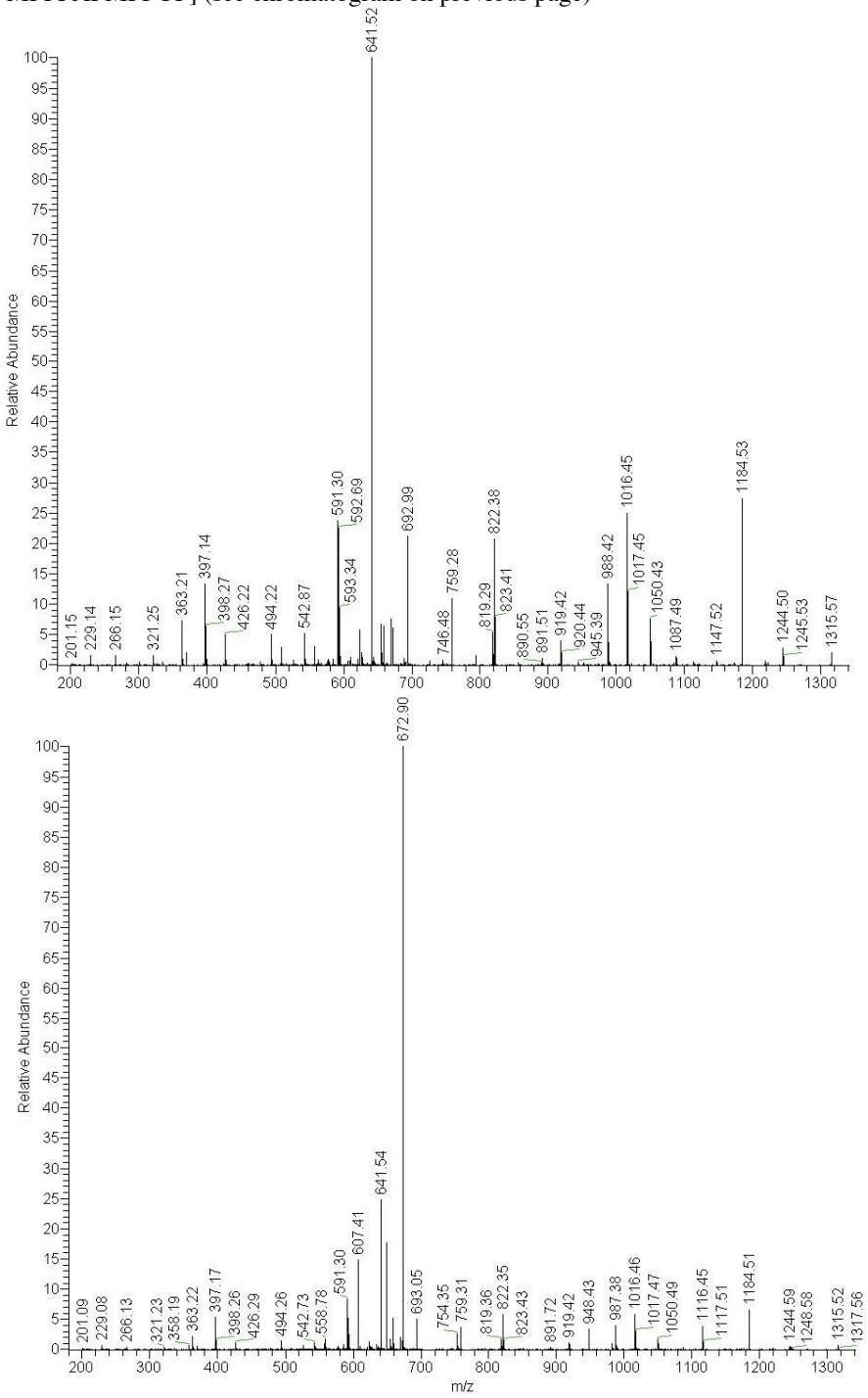
(i) FT-ICR data showing ion whose mass (+/- 2 ppm) corresponds to prenylated cyclo[APMPPPAPMPPYP]



(j) LC-FT-ICR chromatogram selected for mass of prenylated cyclo[APMPPPAPMPPYP]



(k) MS-MS spectra for early (top) and late (bottom) eluting products of the reaction of LynF with cyclo[APMPPPAPMPPYP] (see chromatogram on previous page)



(l) MS-MS assignments for prenylated cyclo[APMPPPAPMPPYP] early (this page) and late (next page) eluting isomers

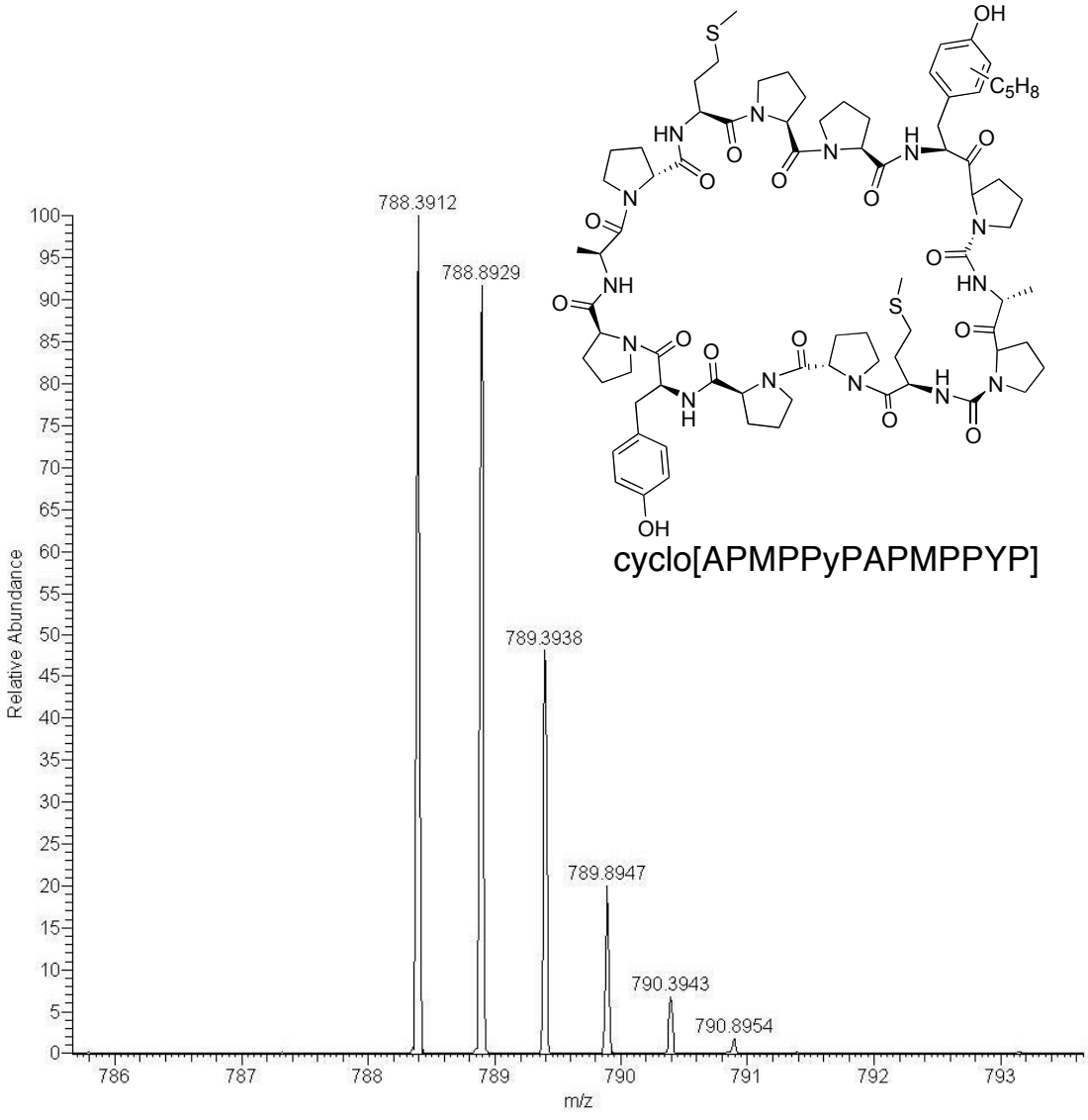
cyclo[APMPPPAPMPPYP] 1st peak (C-prenyl)								
y- b- cleavages (+1)	Expected	Observed	y- b- cleavages (+1)	Expected	Observed	y- b- cleavages (+1)	Expected	Observed
APM	300.14	300.22	MPPPAPMPPy	1147.57	1147.52	PAPMPPy	822.42	822.38
APMP	397.19	397.14	MPPPAPMPPyP	1244.62	1244.51	PAPMPPyP	919.47	919.42
APMPP	494.24	494.22	MPPPAPMPPyPA	1315.66	1315.57	PAPMPPyPAP	1087.56	1087.49
APMPPP	591.30	591.3	PPPA	363.20	363.21	PAPMPPyPAPM	1218.60	1218.53
APMPPPAP	759.39	759.28	PPPAP	460.26	460.28	PAPMPPyPAPMP	1315.66	1315.57
APMPPPAPM	890.43	890.54	PPPAPM	591.30	591.3	APM	300.14	300.22
APMPPPAPMP	987.48	987.42	PPPAPMP	688.35	688.39	APMP	397.19	397.14
APMPPPAPMPP	1084.53	1084.5	PPPAPMPP	785.40	785.47	APMPP	494.24	494.22
APMPPPAPMPPy	1315.66	1315.57	PPPAPMPPy	1016.53	1016.45	APMPPy	725.37	725.38
PM	229.10	229.14	PPPAPMPPyP	1113.58	1113.52	APMPPyP	822.42	822.38
PMP	326.15	326.1	PPPAPMPPyPA	1184.62	1184.52	APMPPyPAPMP	1218.60	1218.53
PMPP	423.21	423.22	PPA	266.15	266.15	APMPPyPAPMPP	1315.66	1315.57
PMPPP	520.26	520.28	PPAP	363.20	363.21	PM	229.10	229.14
PMPPPA	591.30	591.3	PPAPM	494.24	494.22	PMP	326.15	326.1
PMPPPAP	688.35	688.39	PPAPMP	591.30	591.3	PMPP	423.21	423.22
PMPPPAPM	819.39	819.3	PPAPMPP	688.35	688.39	PMPPy	654.33	654.31
PMPPPAPMP	916.44	916.46	PPAPMPPy	919.47	919.42	PMPPyP	751.38	751.41
PMPPPAPMPPy	1244.62	1244.51	PPAPMPPyP	1016.53	1016.45	PMPPyPA	822.42	822.38
MP	229.10	229.14	PPAPMPPyPA	1087.56	1087.49	PMPPyPAP	919.47	919.42
MPP	326.15	326.1	PPAPMPPyPAP	1184.62	1184.52	PMPPyPAPM	1050.51	1050.43
MPPP	423.21	423.22	PPAPMPPyPAPM	1315.66	1315.57	PMPPyPAPMP	1147.57	1147.52
MPPPA	494.24	494.22	PAP	266.15	266.15	PMPPyPAPMPP	1244.62	1244.51
MPPPAP	591.30	591.3	PAPM	397.19	397.14	MP	229.10	229.14
MPPPAPMP	819.39	819.3	PAPMP	494.24	494.22	MPP	326.15	326.1
MPPPAPMPP	916.44	916.46	PAPMPP	591.30	591.3	MPPy	557.28	557.36

y- b- cleavages (+1)	Expected	Observed	y- b- cleavages (+1)	Expected	Observed	Miscellaneous Ions	Expected	Observed
MPPyP	654.33	654.31	yPAPMPP	822.42	822.38	All, -acylium, +2	692.86	692.99
MPPyPA	725.37	725.38	yPAPMPPP	919.47	919.42	PPPAPMPPY, y- a- cleavage +1	988.53	988.42
MPPyPAP	822.42	822.38	yPAPMPPPAP	1087.56	1087.49	PAPM -Met sidechain	321.16	321.25
MPPyPAPM	953.46	953.48	yPAPMPPPAPM	1218.60	1184.52			
MPPyPAPMP	1050.51	1050.43	yPAPMPPPAPMP	1315.66	1315.57			
MPPyPAPMPP	1147.57	1147.52	PAP	266.15	266.15			
MPPyPAPMPPP	1244.62	1244.51	PAPM	397.19	397.14			
MPPyPAPMPPPA	1315.66	1315.57	PAPMP	494.24	494.22			
PPy	426.24	426.22	PAPMPP	591.30	591.3			
PPyPA	594.33	594.33	PAPMPPP	688.35	688.39			
PPyPAPM	822.42	822.38	PAPMPPPA	759.39	759.28			
PPyPAPMP	919.47	919.42	PAPMPPPAP	856.44	856.33			
PPyPAPMPP	1016.53	1016.45	PAPMPPPAPM	987.48	987.42			
PPyPAPMPPP	1113.58	1113.52	PAPMPPPAPMP	1084.53	1084.5			
PPyPAPMPPPA	1184.62	1184.52						
PyP	426.24	426.22	y- b- cleavages (+2)					
PyPAP	594.33	594.33	PAPMPPPAP	428.72	428.83			
PyPAPM	725.37	725.38	PyPAPMPPP	508.77	508.83			
PyPAPMP	822.42	822.38	PMPPyPAPM	525.76	525.76			
PyPAPMPP	919.47	919.42	APMPPPAPMPP	542.77	542.87			
PyPAPMPPP	1016.53	1016.45	PPAPMPPyPAP	592.82	592.69			
PyPAPMPPPA	1087.56	1087.49	PAPMPPyPAPM	609.81	609.84			
PyPAPMPPPAP	1184.62	1184.52	PMPPyPAPMPP	622.81	623.00			
PyPAPMPPPAPM	1315.66	1315.57	PPyPAPMPPPAP	641.34	641.52			
yPAPMP	725.37	725.38	PAPMPPyPAPMP	658.34	658.31			

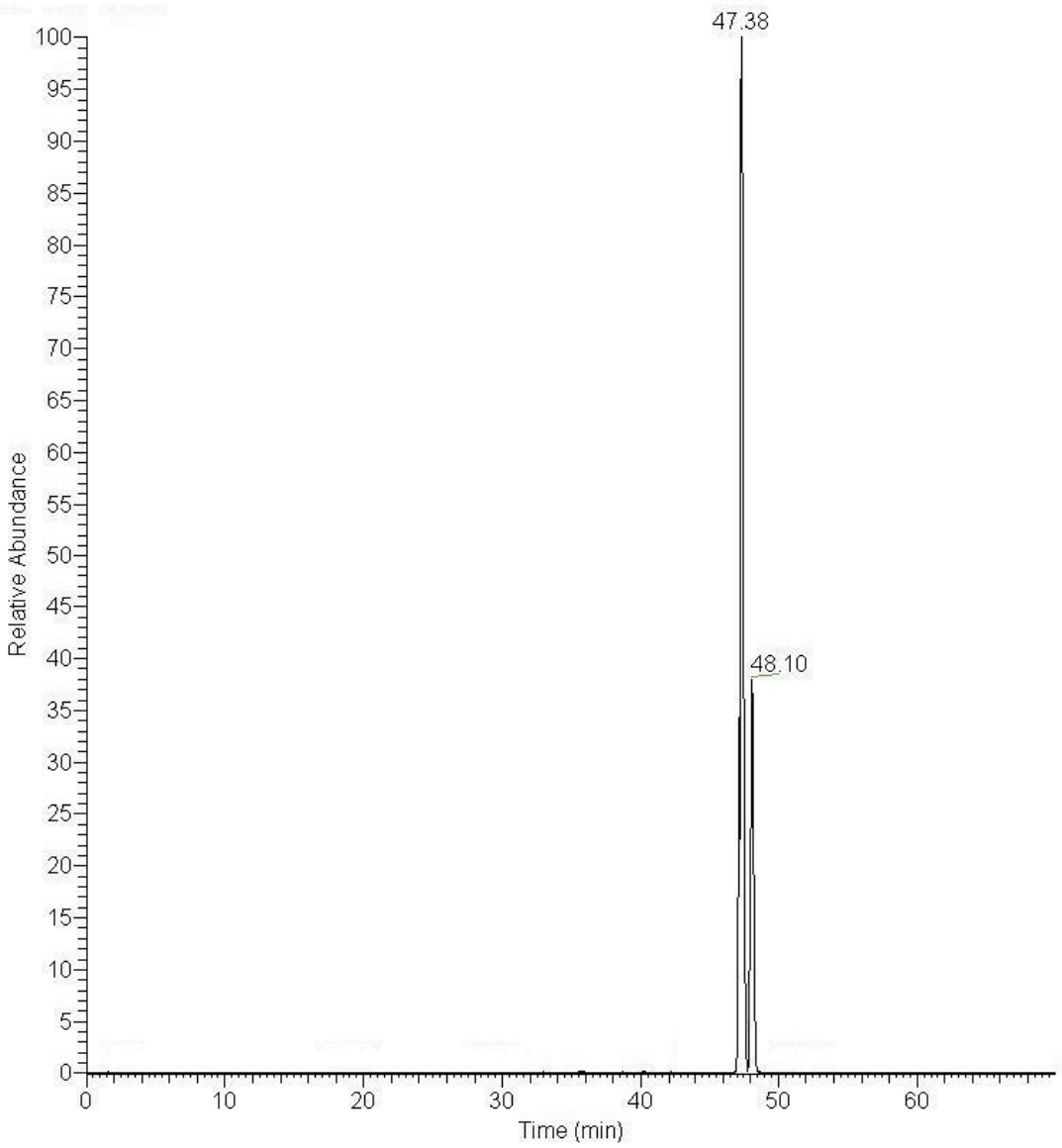
cyclo[APMPPPAPMPPYP] 2nd peak (O-prenyl)								
y- b- cleavages (+1)	Expected	Observed	y- b- cleavages (+1)	Expected	Observed	y- b- cleavages (+1)	Expected	Observed
APMP	397.19	397.17	MPPPAPMPPyP	1244.62	1244.59	PAPMPPyPA	990.51	990.48
APMPP	494.24	494.26	MPPPAPMPPyPA	1315.66	1315.52	PAPMPPyPAP	1087.56	1087.50
APMPPP	591.30	591.30	PPPA	363.20	363.22	PAPMPPyPAPM	1218.60	1218.69
APMPPPAP	759.39	759.31	PPPAP	460.26	460.42	PAPMPPyPAPMP	1315.66	1315.52
APMPPPAPM	890.43	890.45	PPPAPM	591.30	591.30	APMP	397.19	397.17
APMPPPAPMP	987.48	987.38	PPPAPMP	688.35	688.32	APMPP	494.24	494.26
APMPPPAPMPP	1084.53	1084.64	PPPAPMPP	785.40	785.52	APMPPy	725.37	725.32
APMPPPAPMPPy	1315.66	1315.52	PPPAPMPPy	1016.53	1016.46	APMPPyP	822.42	822.35
PM	229.10	229.08	PPPAPMPPyPA	1184.62	1184.51	APMPPyPAP	990.51	990.48
PMP	326.15	326.18	PPA	266.15	266.13	APMPPyPAPM	1121.55	1121.48
PMPP	423.21	432.19	PPAP	363.20	363.22	APMPPyPAPMP	1218.60	1218.69
PMPPPA	591.30	591.30	PPAPM	494.24	494.26	APMPPyPAPMPP	1315.66	1315.52
PMPPPAP	688.35	688.32	PPAPMP	591.30	591.30	PM	229.10	229.08
PMPPPAPM	819.39	819.36	PPAPMPP	688.35	688.32	PMP	326.15	326.18
PMPPPAPMP	916.44	916.43	PPAPMPPy	919.47	919.42	PMPP	423.21	423.19
PMPPPAPMPPy	1244.62	1244.59	PPAPMPPyP	1016.53	1016.46	PMPPyPA	822.42	822.35
MP	229.10	229.08	PPAPMPPyPA	1087.56	1087.50	PMPPyPAP	919.47	919.42
MPP	326.15	326.18	PPAPMPPyPAP	1184.62	1184.51	PMPPyPAPM	1050.51	1050.49
MPPP	423.21	423.19	PPAPMPPyPAPM	1315.66	1315.52	PMPPyPAPMP	1147.57	1147.44
MPPPA	494.24	494.26	PAP	266.15	266.13	PMPPyPAPMPP	1244.62	1244.59
MPPPAP	591.30	591.30	PAPM	397.19	397.17	MP	229.10	229.08
MPPPAPM	722.34	722.36	PAPMP	494.24	494.26	MPP	326.15	326.18
MPPPAPMP	819.39	819.36	PAPMPP	591.30	591.30	MPPy	557.28	557.53
MPPPAPMPP	916.44	916.43	PAPMPPy	822.42	822.35	MPPyPA	725.37	725.42
MPPPAPMPPy	1147.57	1147.44	PAPMPPyP	919.47	919.42	MPPyPAP	822.42	822.35

y- b- cleavages (+1)	Expected	Observed	y- b- cleavages (+1)	Expected	Observed	Miscellaneous Ions	Expected	Observed
MPPyPAPMP	1050.51	1050.49	yPAPMPPPAPM	1218.60	1218.69	+1, -Prenyl	1344.65	1344.53
MPPyPAPMPP	1147.57	1147.44	yPAPMPPPAPMP	1315.66	1315.52	APMPPPAPMPPy -Prenyl	1315.66	1315.52
MPPyPAPMPPP	1244.62	1244.59	PAP	266.15	266.13			
MPPyPAPMPPPA	1315.66	1315.52	PAPM	397.19	397.17			
PPy	426.24	426.29	PAPMP	494.24	494.26			
PPyP	523.29	523.34	PAPMPP	591.30	591.30			
PPyPA	594.33	594.35	PAPMPPP	688.35	688.32			
PPyPAPM	822.42	822.35	PAPMPPPA	759.39	759.31			
PPyPAPMP	919.47	919.42	PAPMPPPAP	856.44	856.39			
PPyPAPMPP	1016.53	1016.46	PAPMPPPAPM	987.48	987.38			
PPyPAPMPPPA	1184.62	1184.51	PAPMPPPAPMP	1084.53	1084.64			
PyP	426.24	426.29						
PyPAP	594.33	594.35	y- b- cleavages (+2)					
PyPAPM	725.37	725.42	PMPPPAPMPP	507.25	507.41			
PyPAPMP	822.42	822.35	APMPPPAPMPP	542.77	542.73			
PyPAPMPP	919.47	919.42	PPPAPMPPyPA	592.81	592.62			
PyPAPMPPP	1016.53	1016.46	PPyPAPMPPPAP	641.34	641.54			
PyPAPMPPPA	1087.56	1087.50						
PyPAPMPPPAP	1184.62	1184.51	Miscellaneous Ions					
PyPAPMPPPAPM	1315.66	1315.52	PPPAPMPPyPA, -Prenyl, +2	607.30	607.41			
yPAPMP	725.37	725.42	+2 -Prenyl, -Methyl Sulfide	648.83	648.90			
yPAPMPP	822.42	822.35	+2, -Prenyl, -Acylium	658.83	658.68			
yPAPMPPP	919.47	919.42	+2, -Prenyl	672.82	672.90			
yPAPMPPPA	990.51	990.48	PPPAPMPPy -Prenyl	948.46	948.43			
yPAPMPPPAP	1087.56	1087.50	PPPAPMPPyPA -Prenyl	1116.55	1116.45			

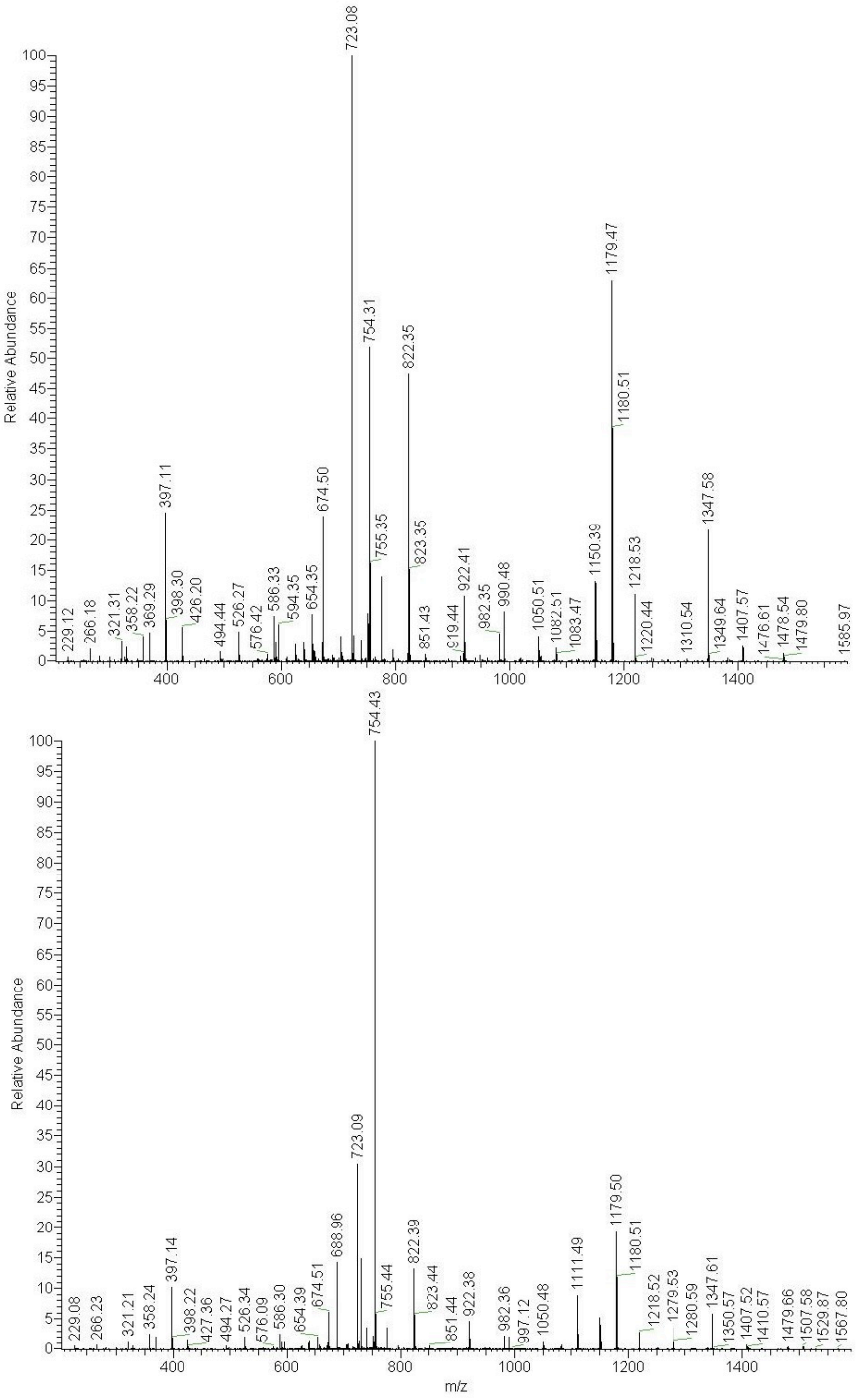
(m) FT-ICR data showing ion whose mass (+/- 2 ppm) corresponds to singly prenylated cyclo[APMPPYPAPMPPYP]



(n) LC-FT-ICR chromatogram selected for mass of singly prenylated cyclo[APMPPYPAPMPPYP]



(o) MS-MS spectra for early (top) and late (bottom) eluting singly prenylated products of the reaction of LynF with cyclo[APMPPYPAPMPPYP] (see chromatogram on previous page)



(p) MS-MS assignments for singly prenylated cyclo[APMPPYPAPMPPYP] early (this page) and late (next page) eluting isomers

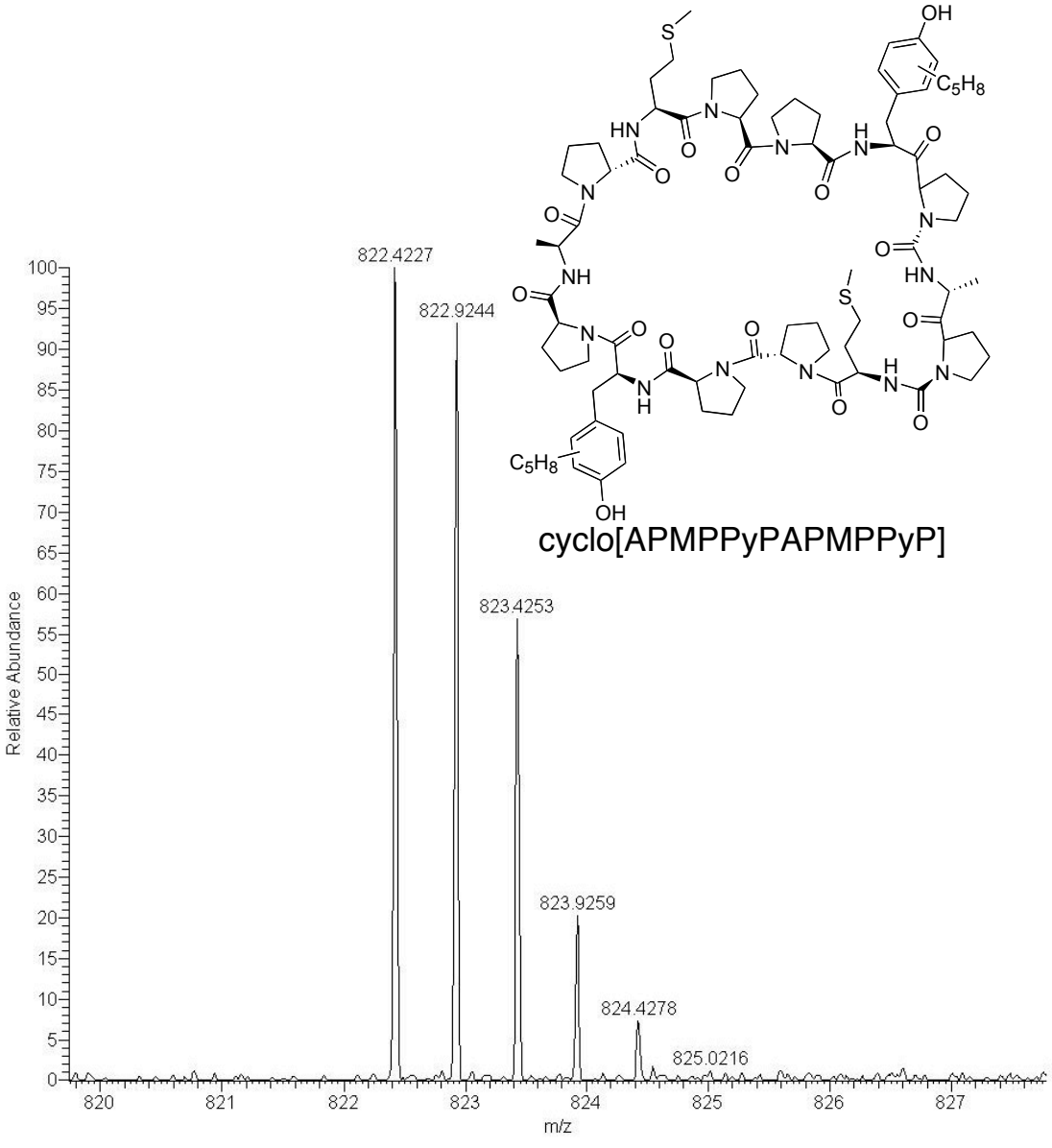
cyclo[APMPPYPAPMPPYP] 1st peak (C-prenyl)								
y- b- cleavages (+1)	**Expected**	**Observed**	**y- b- cleavages (+1)**	**Expected**	**Observed**	**y- b- cleavages (+1)**	**Expected**	**Observed**
APMP	397.19	397.11	**MPPYPAPMPPyPA**	1478.72	1478.54	**PAPMPPyP**	919.47	919.44
APMPP	494.24	494.44	**PPY**	358.18	358.22	**PAPMPPyPA**	990.51	990.48
APMPPYP	754.36	754.31	**PPYPA**	526.27	526.27	**PAPMPPyPAPM**	1218.60	1218.53
APMPPYPAP	922.45	922.41	**PPYPAP**	623.32	623.26	**APMP**	397.19	397.11
APMPPYPAPMP	1150.54	1150.39	**PPYPAPM**	754.36	754.31	**APMPP**	494.24	494.44
APMPPYPAPMPP	1247.60	1247.45	**PPYPAPMP**	851.41	851.43	**APMPPyP**	822.42	822.35
APMPPYPAPMPPy	1478.72	1478.54	**PPYPAPMPP**	948.47	948.45	**APMPPyPA**	893.46	893.51
PM	229.10	229.12	**PPYPAPMPPy**	1179.59	1179.47	**APMPPyPAP**	990.51	990.48
PMP	326.15	326.25	**PPYPAPMPPyPA**	1347.68	1347.58	**APMPPyPAPM**	1121.55	1121.39
PMPPY	586.27	586.33	**PYP**	358.18	358.22	**APMPPyPAPMP**	1218.60	1218.53
PMPPYP	683.32	683.39	**PYPAP**	526.27	526.27	**APMPPyPAPMPPY**	1478.72	1478.54
PMPPYPA	754.36	754.31	**PYPAPMP**	754.36	754.31	**PM**	229.10	229.12
PMPPYPAP	851.41	851.43	**PYPAPMPP**	851.41	851.43	**PMP**	326.15	326.25
PMPPYPAPM	982.45	982.35	**PYPAPMPPy**	1082.54	1082.51	**PMPPy**	654.33	654.35
PMPPYPAPMP	1079.51	1079.44	**PYPAPMPPyP**	1179.59	1179.47	**PMPPyPA**	822.42	822.35
PMPPYPAPMPPy	1407.68	1407.57	**PYPAPMPPyPAP**	1347.68	1347.58	**PMPPyPAP**	919.47	919.44
MP	229.10	229.12	**PYPAPMPPyPAPM**	1478.72	1478.54	**PMPPyPAPM**	1050.51	1050.51
MPP	326.15	326.25	**YPAPMPP**	754.36	754.31	**PMPPyPAPMP**	1147.57	1147.53
MPPYP	586.27	586.33	**YPAPMPPy**	985.48	985.32	**PMPPyPAPMPPY**	1407.68	1407.57
MPPYPAP	754.36	754.31	**YPAPMPPyP**	1082.54	1082.51	**MP**	229.10	229.12
MPPYPAPMP	982.45	982.35	**YPAPMPPyPAPMP**	1478.72	1478.54	**MPP**	326.15	326.25
MPPYPAPMPP	1079.51	1079.44	**PAP**	266.15	266.18	**MPPyP**	654.33	654.35
MPPYPAPMPPy	1310.63	1310.53	**PAPM**	397.19	397.11	**MPPyPAP**	822.42	822.35
MPPYPAPMPPyP	1407.68	1407.57	**PAPMP**	494.24	494.44	**MPPyPAPM**	953.46	953.57
MPPYPAPMPPyPA	1478.72	1478.54	**PAPMPPy**	822.42	822.35	**MPPyPAPMP**	1050.51	1050.51

y- b- cleavages (+1)	**Expected**	**Observed**	**y- b- cleavages (+1)**	**Expected**	**Observed**
MPPyPAPMPP	1147.57	1147.33	**PAPMP**	494.24	494.44
MPPyPAPMPPY	1310.63	1310.54	**PAPMPPY**	754.36	754.31
MPPyPAPMPPYP	1407.68	1407.57	**PAPMPPYP**	851.41	851.43
MPPyPAPMPPYPA	1478.72	1478.54	**PAPMPPYPA**	922.45	922.41
PPy	426.24	426.20	**PAPMPPYPAP**	1019.50	1019.39
PPyPA	594.33	594.35	**PAPMPPYPAPM**	1150.54	1150.39
PPyPAPM	822.42	822.35	**PAPMPPYPAPMP**	1247.60	1247.45
PPyPAPMP	919.47	919.44			
PPyPAPMPP	1016.53	1016.55	**y- b- cleavages (+2)**		
PPyPAPMPPY	1179.59	1179.47	**PPYPAPMPPyPAP**	722.87	723.08
PPyPAPMPPYPA	1347.68	1347.58	**PMPPYPAPMPPy**	704.35	704.36
PyP	426.24	426.20	**PyPAPMPPYPAP**	674.34	674.50
PyPAP	594.33	594.35	**PPyPAPMPPYP**	638.82	638.82
PyPAPM	725.37	725.31	**PAPMPPYPAPMP**	624.30	624.46
PyPAPMP	822.42	822.35			
PyPAPMPP	919.47	919.44	**Miscelaneous Ions**		
PyPAPMPPY	1082.54	1082.51	**PAPM y- a- cleavage**	369.20	329.29
PyPAPMPPYP	1179.59	1179.47	**PAPM -Met sidechain**	321.16	321.31
PyPAPMPPYPAP	1347.68	1347.58			
PyPAPMPPYPAPM	1478.72	1478.54			
yPAPMPP	822.42	822.35			
yPAPMPPYP	1082.54	1082.51			
yPAPMPPYPAPMP	1478.72	1478.54			
PAP	266.15	266.18			
PAPM	397.19	397.11			

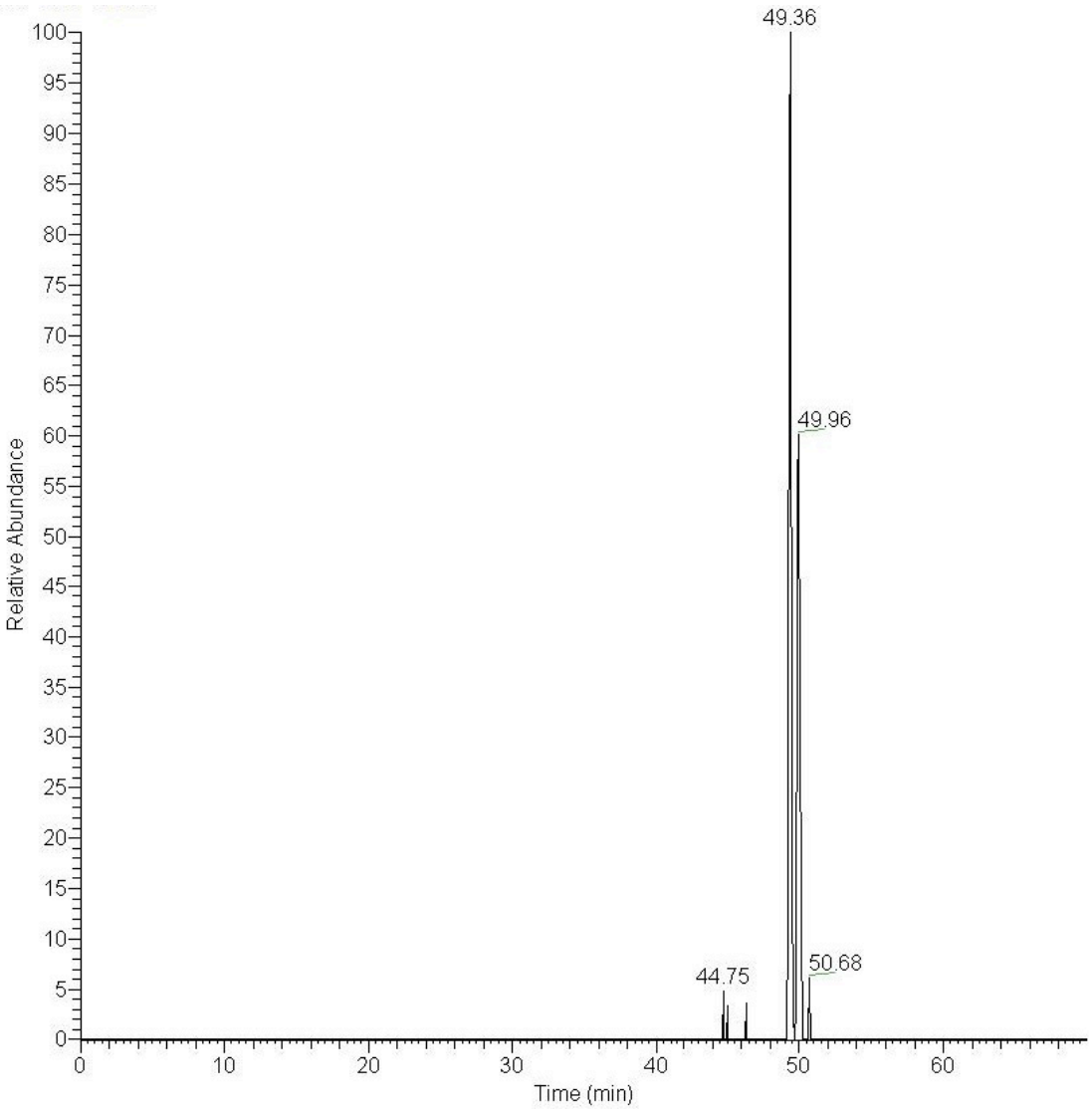
cyclo[APMPPYPAPMPPYP] 2nd peak (O-prenyl)								
y- b- cleavages (+1)	Expected	Observed	y- b- cleavages (+1)	Expected	Observed	y- b- cleavages (+1)	Expected	Observed
APM	300.14	299.94	PPYP	455.23	455.40	YPAPMPPyPAPM	1,381.67	1,381.54
APMP	397.19	397.14	PPYPA	526.27	526.33	YPAPMPPyPAPMP	1,478.72	1,478.47
APMPP	494.24	494.28	PPYPAP	623.32	623.43	PAP	266.15	266.23
APMPPYP	754.36	754.43	PPYPAPM	754.36	754.43	PAPM	397.19	397.14
APMPPYPAP	922.45	922.38	PPYPAPMP	851.41	851.42	PAPMP	494.24	494.28
APMPPYPAPMP	1,150.54	1,150.45	PPYPAPMPP	948.47	948.54	PAPMPPy	822.42	822.39
APMPPYPAPMPP	1,247.60	1,247.60	PPYPAPMPPy	1,179.59	1,179.49	PAPMPPyPA	990.51	990.48
APMPPYPAPMPPy	1,478.72	1,478.47	PPYPAPMPPyP	1,276.64	1,276.48	PAPMPPyPAPM	1,218.60	1,218.51
PM	229.10	229.08	PPYPAPMPPyPA	1,347.68	1,347.60	APM	300.14	299.94
PMPPY	586.27	586.30	PYP	358.18	358.23	APMP	397.19	397.14
PMPPYP	683.32	683.13	PYPA	429.21	429.36	APMPP	494.24	494.28
PMPPYPA	754.36	754.43	PYPAP	526.27	526.33	APMPPyP	822.42	822.39
PMPPYPAP	851.41	851.42	PYPAPMP	754.36	754.43	APMPPyPAP	990.51	990.48
PMPPYPAPM	982.45	982.37	PYPAPMPP	851.41	851.42	APMPPyPAPM	1,121.55	1,121.51
PMPPYPAPMP	1,079.51	1,079.40	PYPAPMPPy	1,082.54	1,082.45	APMPPyPAPMP	1,218.60	1,218.51
PMPPYPAPMPPy	1,407.68	1,407.52	PYPAPMPPyP	1,179.59	1,179.49	APMPPyPAPMPP	1,315.66	1,315.55
MP	229.10	229.08	PYPAPMPPyPA	1,250.63	1,250.73	APMPPyPAPMPPY	1,478.72	1,478.47
MPPYP	586.27	586.30	PYPAPMPPyPAP	1,347.68	1,347.60	PM	229.10	229.08
MPPYPAP	754.36	754.43	PYPAPMPPyPAPM	1,478.72	1,478.47	PMPPy	654.33	654.38
MPPYPAPMP	982.45	982.37	YPA	332.16	332.38	PMPPyPA	822.42	822.39
MPPYPAPMPP	1,079.51	1,079.40	YPAP	429.21	429.36	PMPPyPAPM	1,050.51	1,050.47
MPPYPAPMPPy	1,310.63	1,310.66	YPAPMPP	754.36	754.43	PMPPyPAPMP	1,147.57	1,147.46
MPPYPAPMPPyP	1,407.68	1,407.60	YPAPMPPyP	1,082.54	1,082.45	PMPPyPAPMPPY	1,407.68	1,407.52
MPPYPAPMPPyPA	1,478.72	1,478.47	YPAPMPPyPA	1,153.57	1,153.43	MP	229.10	229.08
PPY	358.18	358.23	YPAPMPPyPAP	1,250.63	1,250.73	MPPy	557.28	557.92

y- b- cleavages (+1)	Expected	Observed	y- b- cleavages (+1)	Expected	Observed
MPPyP	654.33	654.38	PAP	266.15	266.23
MPPyPAP	822.42	822.39	PAPM	397.19	397.14
MPPyPAPMP	1,050.51	1,050.47	PAPMP	494.24	494.28
MPPyPAPMPP	1,147.57	1,147.46	PAPMPPY	754.36	754.43
MPPyPAPMPPY	1,310.63	1,310.66	PAPMPPYP	851.41	851.42
MPPyPAPMPPYP	1,407.68	1,407.52	PAPMPPYPA	922.45	922.38
MPPyPAPMPPYPA	1,478.72	1,478.47	PAPMPPYPAPM	1,150.54	1,150.45
PPy	426.24	426.23	PAPMPPYPAPMP	1,247.60	1,247.60
PPyPA	594.33	594.29			
PPyPAPM	822.42	822.39	y- b- cleavages (+2)		
PPyPAPMPPY	1,179.59	1,179.49	MPPYPAPMPPyPA	739.86	739.82
PPyPAPMPPYPA	1,347.68	1,347.60	PMPPYPAPMPPy	704.34	704.32
PyP	426.24	426.23	MPPYPAPMPP	540.26	540.27
PyPAP	594.33	594.29			
PyPAPMP	822.42	822.39	Miscellaneous Ions		
PyPAPMPPY	1,082.54	1,082.45	PyPAPMPPYPAP +2	674.34	674.51
PyPAPMPPYP	1,179.59	1,179.49	PPyPAPMPPYPAP -prenyl +2	688.84	688.96
PyPAPMPPYPA	1,250.63	1,250.73	PPyPAPMPPYPAP +2	722.87	723.08
PyPAPMPPYPAP	1,347.68	1,347.60	PYPAPMPPYP -prenyl +1	1,111.53	1,111.48
PyPAPMPPYPAPM	1,478.72	1,478.47	APMPPYPAPMP -prenyl +1	1,150.54	1,150.45
yPAPMPP	822.42	822.39	PPYPAPMPPyPAP -prenyl +1	1,279.62	1,279.54
yPAPMPPYP	1,082.54	1,082.45			
yPAPMPPYPAP	1,250.63	1,250.73			
yPAPMPPYPAPM	1,381.67	1,381.54			
yPAPMPPYPAPMP	1,478.72	1,478.47			

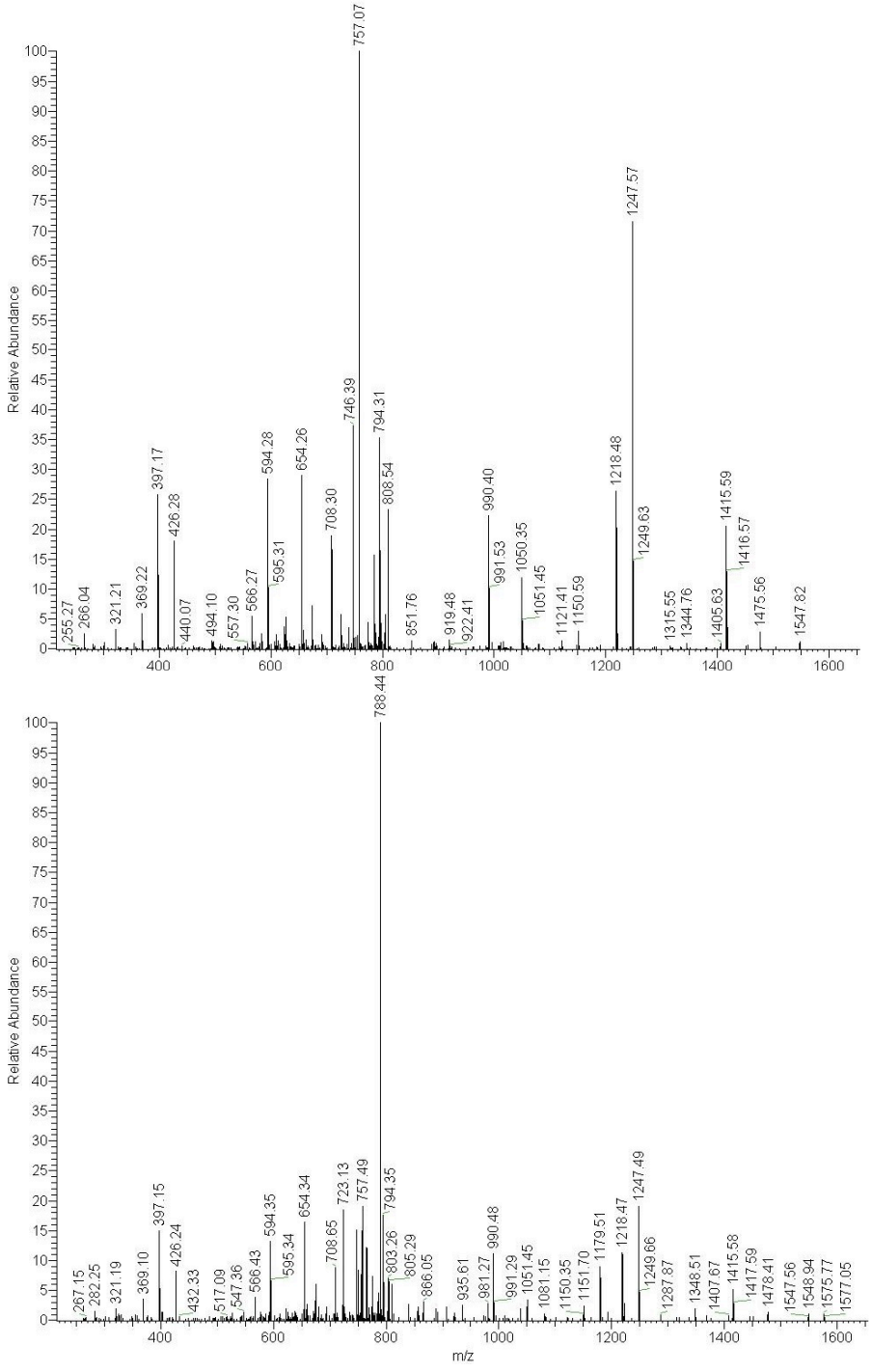
(q) FT-ICR data showing ion whose mass (+/- 2ppm) corresponds to doubly prenylated cyclo[APMPPYPAPMPPYP]



(r) LC-FT-ICR chromatogram selected for mass of doubly prenylated cyclo[APMPPYPAPMPPYP]



(s) MS-MS spectra for early (top) and late (bottom) eluting doubly prenylated products of the reaction of LynF with cyclo[APMPPYPAPMPPYP] (see chromatogram on previous page)

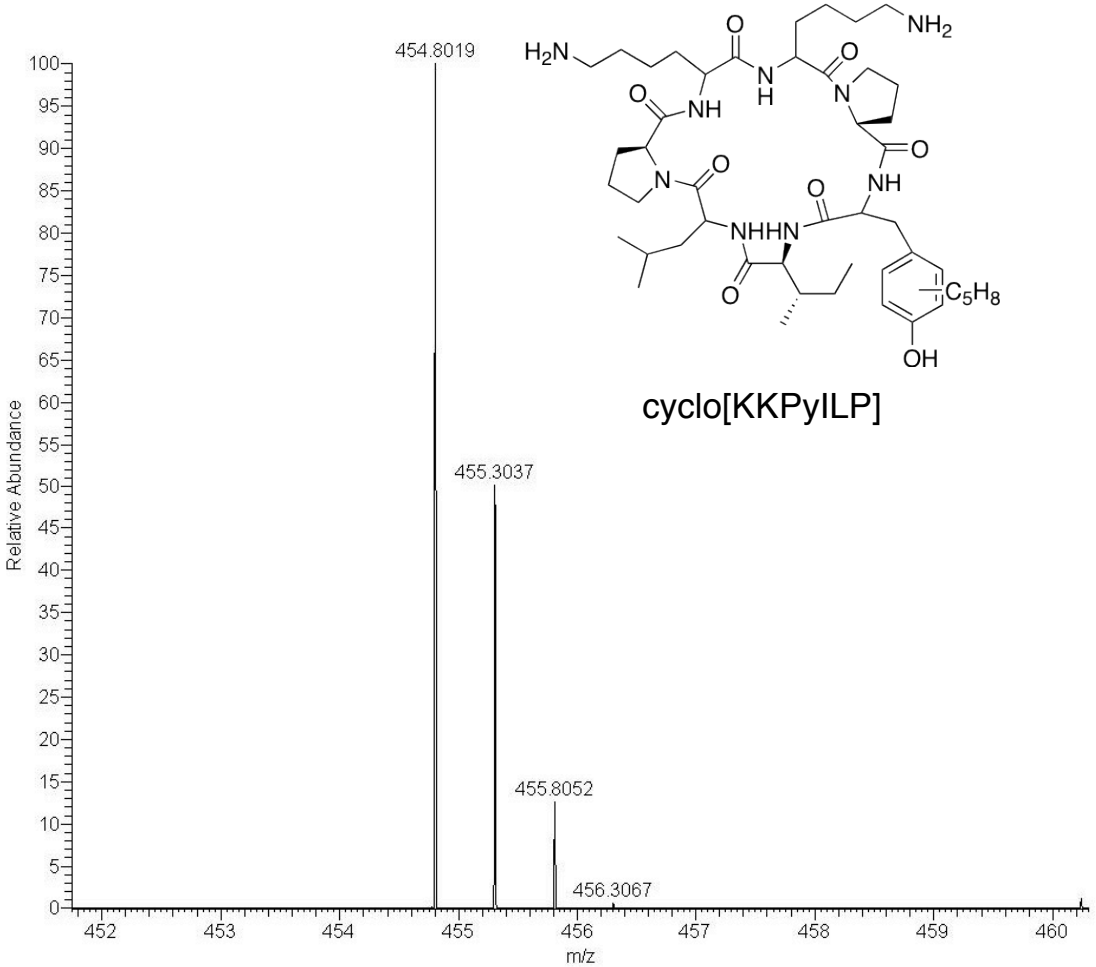


(t) MS-MS assignments for doubly prenylated cyclo[APMPPYPAPMPPYP] early (top) and late (bottom) eluting isomers

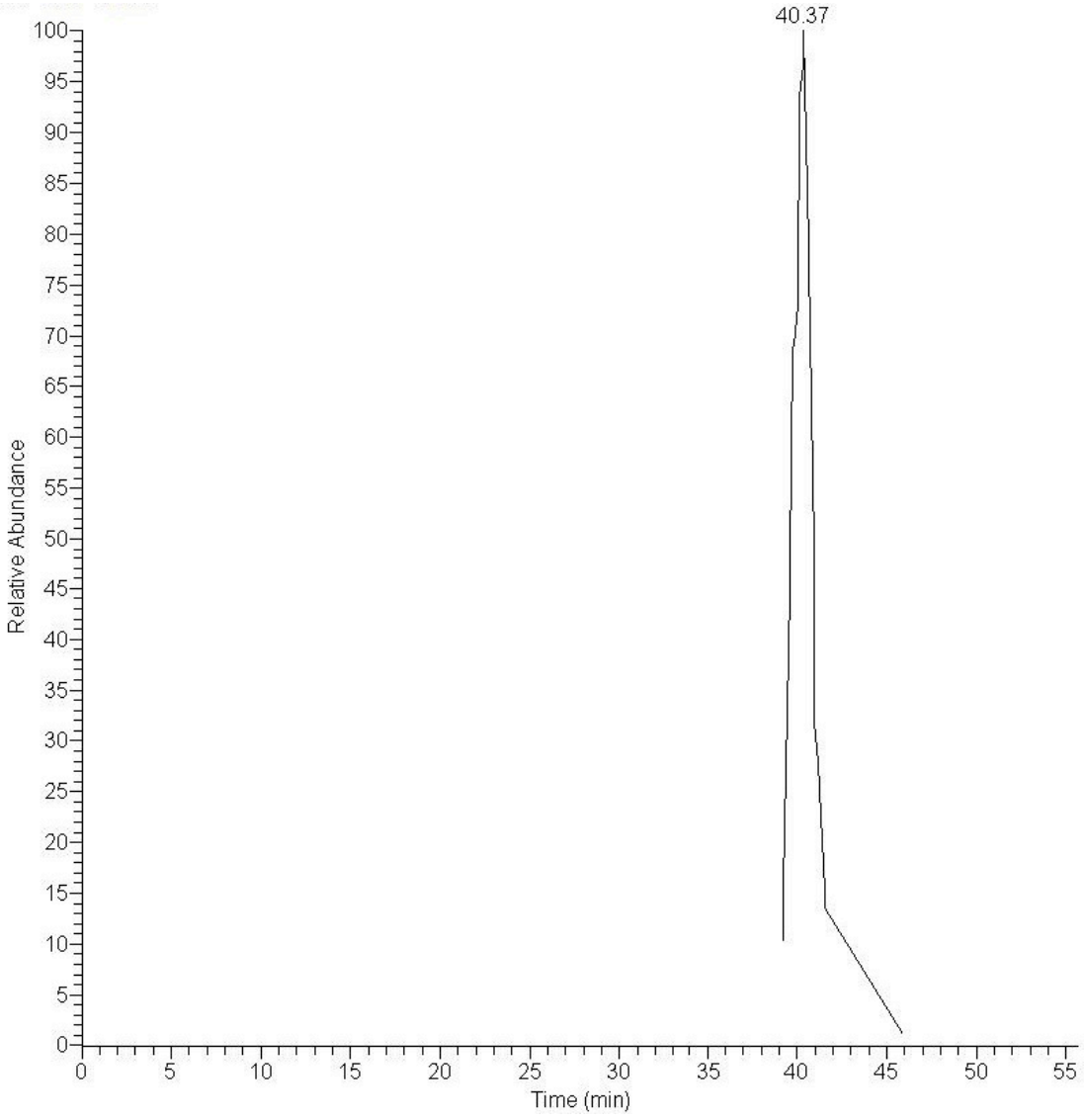
cyclo[APMPPYPAPMPPYP] 1st peak								
y- b- cleavages (+1)	Expected	Observed	y- b- cleavages (+1)	Expected	Observed	y- b- cleavages (+1)	Expected	Observed
APMP	397.19	397.17	PPyPAPMPP	1016.53	1016.45	PAPMPPyPAPMP	1315.66	1315.55
APMPPy	725.37	725.40	PPyPAPMPPy	1247.66	1247.57			
APMPPyPAP	990.51	990.40	PPyPAPMPPyP	1344.71	1344.76	y- b- cleavage (+2)		
APMPPyPAPM	1121.55	1121.41	PPyPAPMPPyPA	1415.75	1415.59	PPyPAPMPPyPAP	756.90	757.07
APMPPyPAPMP	1218.60	1218.48	PyP	426.24	426.28	PyPAPMPPyPAP	708.38	708.30
APMPPyPAPMPP	1315.66	1315.55	PyPA	497.27	497.35	PPyPAPMPPyP	672.35	672.83
APMPPyPAPMPPy	1546.79	1546.40	PyPAP	594.33	594.28	APMPPyPAPMPP	657.83	658.47
PMP	326.15	326.18	PyPAPM	725.37	725.40	PyPAPMPPyP	623.83	624.14
PMPP	423.21	423.13	PyPAPMPP	919.47	919.48	APMPPyPAPMP	609.30	609.98
PMPPy	654.33	654.26	PyPAPMPPy	1150.61	1150.59			
PMPPyPAP	919.47	919.48	PyPAPMPPyP	1247.66	1247.57	Miscellaneous Ions		
PMPPyPAPM	1050.51	1050.35	PyPAPMPPyPAP	1415.75	1415.59	PAPM -Met sidechain	321.16	321.21
PMPPyPAPMP	1147.57	1147.56	PyPAPMPPyPAPM	1546.79	1546.40	PAPM y- a- cleavage	369.20	369.22
PMPPyPAPMPPy	1475.75	1475.56	yPAP	497.27	497.35	PyPAP y- a- cleavage	566.33	566.27
MPP	326.15	326.18	yPAPMP	725.37	725.40	PMPPyPA -Met sidechain	746.39	746.39
MPPy	557.28	557.30	yPAPMPPyP	1150.61	1150.59	APMPPyP y- a- cleavage	794.43	794.31
MPPyP	654.33	654.26	yPAPMPPyPA	1221.64	1221.67			
MPPyPA	725.37	725.40	yPAPMPPyPAPMP	1546.79	1546.40			
MPPyPAPMP	1050.51	1050.35	PAP	266.15	266.04			
MPPyPAPMPP	1147.57	1147.56	PAPM	397.19	397.17			
MPPyPAPMPPyP	1475.75	1475.56	PAPMP	494.24	494.10			
PPy	426.24	426.28	PAPMPPyP	919.47	919.48			
PPyPA	594.33	594.28	PAPMPPyPA	990.51	990.40			
PPyPAP	691.38	691.16	PAPMPPyPAP	1087.56	1087.74			
PPyPAPMP	919.47	919.48	PAPMPPyPAPM	1218.60	1218.48			

cyclo[APMPPYPAPMPPYP] 2nd peak								
y- b- cleavages (+1)	Expected	Observed	y- b- cleavages (+1)	Expected	Observed	Miscellaneous Ions	Expected	Observed
APMP	397.19	397.17	PyPAP	594.33	594.34	PPyPAPMPPyPAP -prenyl, +2	722.87	723.13
APMPPy	725.37	725.38	PyPAPM	725.37	725.38	All, -prenyl, +2	788.39	788.45
APMPPyPAP	990.51	990.52	PyPAPMPP	919.47	919.49	PPyPAPMPPy, -prenyl +1	1179.59	1179.53
APMPPyPAPM	1121.55	1121.56	PyPAPMPPy	1150.61	1150.46	PPyPAPMPPyPA, -prenyl +1	1347.68	1347.65
APMPPyPAPMP	1218.60	1218.43	PyPAPMPPyP	1247.66	1247.55	PyPAPMPPyPAP -prenyl +2	674.34	674.55
APMPPyPAPMPP	1315.66	1315.68	PyPAPMPPyPAP	1415.75	1415.75			
PMPP	423.21	423.26	yPAPMP	725.37	725.38			
PMPPy	654.33	654.38	yPAPMPPyP	1150.61	1150.46			
PMPPyPAP	919.47	919.49	PAP	266.15	266.10			
PMPPyPAPM	1050.51	1050.51	PAPM	397.19	397.17			
PMPPyPAPMPPy	1475.75	1475.66	PAPMPPyP	919.47	919.49			
MPPy	557.28	557.38	PAPMPPyPA	990.51	990.52			
MPPyP	654.33	654.38	PAPMPPyPAPM	1218.60	1218.43			
MPPyPA	725.37	725.38	PAPMPPyPAPMP	1315.66	1315.68			
MPPyPAPMP	1050.51	1050.51						
MPPyPAPMPPyP	1475.75	1475.66	y- b- cleavages (+2)					
PPy	426.24	426.28	PyPAPMPPyPAP	708.38	708.66			
PPyP	523.29	523.43	PPyPAPMPPyPAP	756.90	756.79			
PPyPA	594.33	594.34						
PPyPAP	691.38	691.37	Miscellaneous Ions					
PPyPAPMP	919.47	919.49	PAPM -Met sidechain	321.16	321.20			
PPyPAPMPP	1016.53	1016.41	PAPM y- a- cleavage	369.20	369.11			
PPyPAPMPPy	1247.66	1247.55	PyPAP y- a- cleavage	566.33	566.41			
PPyPAPMPPyPA	1415.75	1415.75	PMPPyPA -Met sidechain	746.39	746.40			
PyP	426.24	426.28	APMPPyP y- a- cleavage	794.43	794.55			

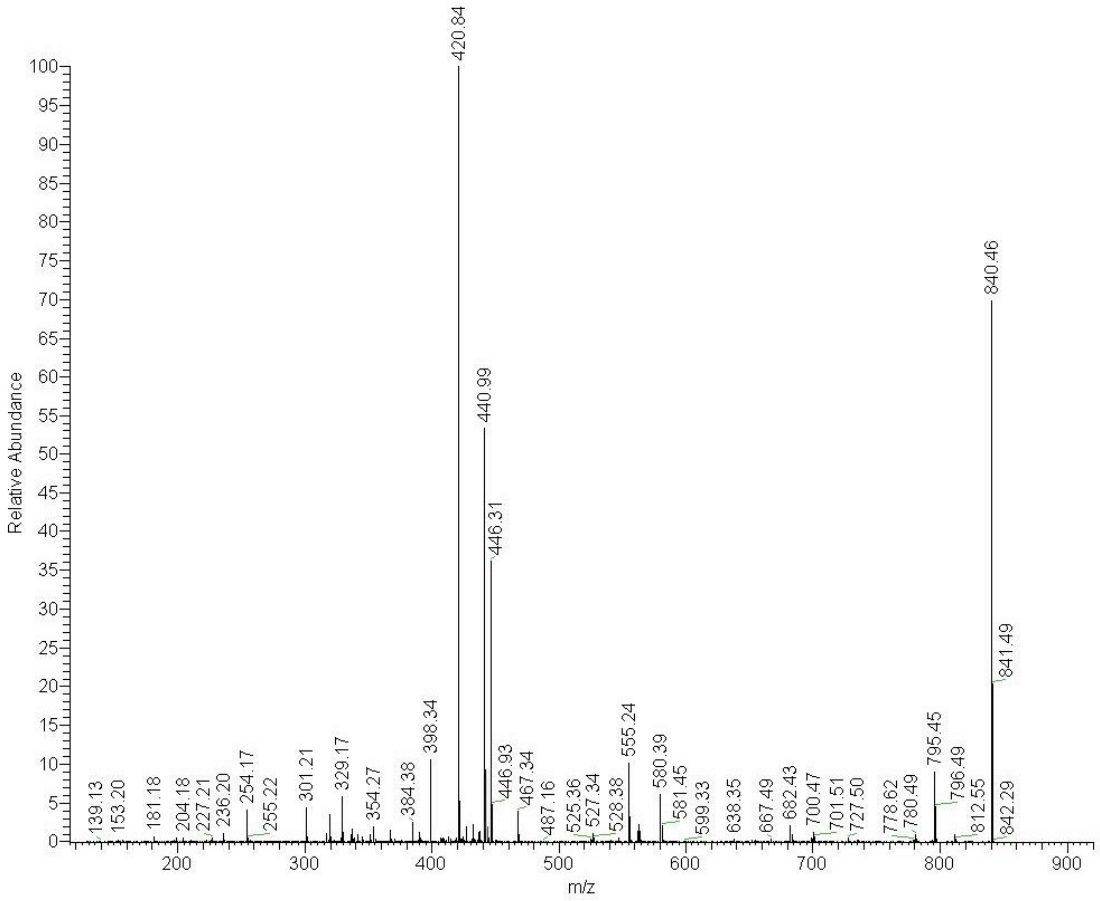
(u) FT-ICR data showing ion whose mass (+/- 2 ppm) corresponds to prenylated cyclo[KKPYILP]



(v) LC-FT-ICR chromatogram selected for mass of prenylated cyclo[KKPYILP]; in this case alone was only a single broad chromatographic peak observed, most likely indicating that C- and O-prenylated isomers were not well separated



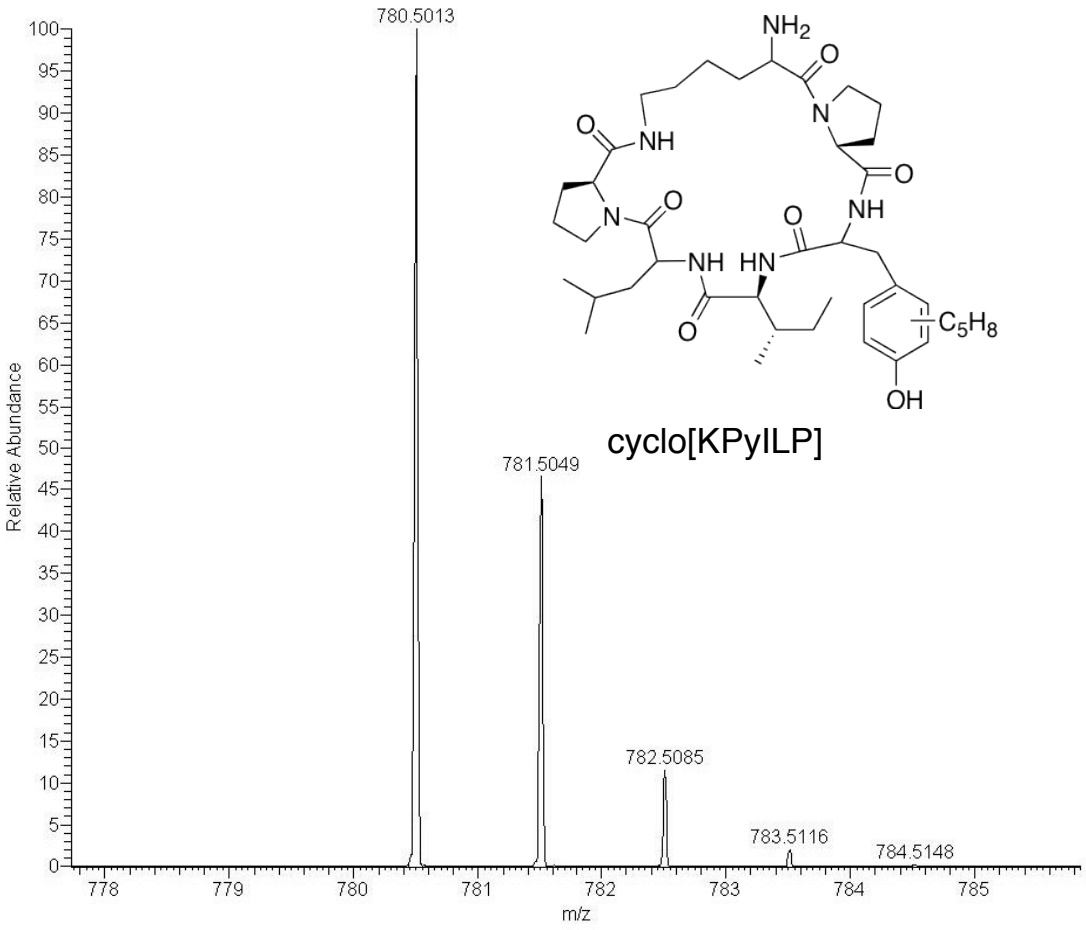
(w) MS-MS spectrum for product of reaction of LynF with cyclo[KKPYILP]



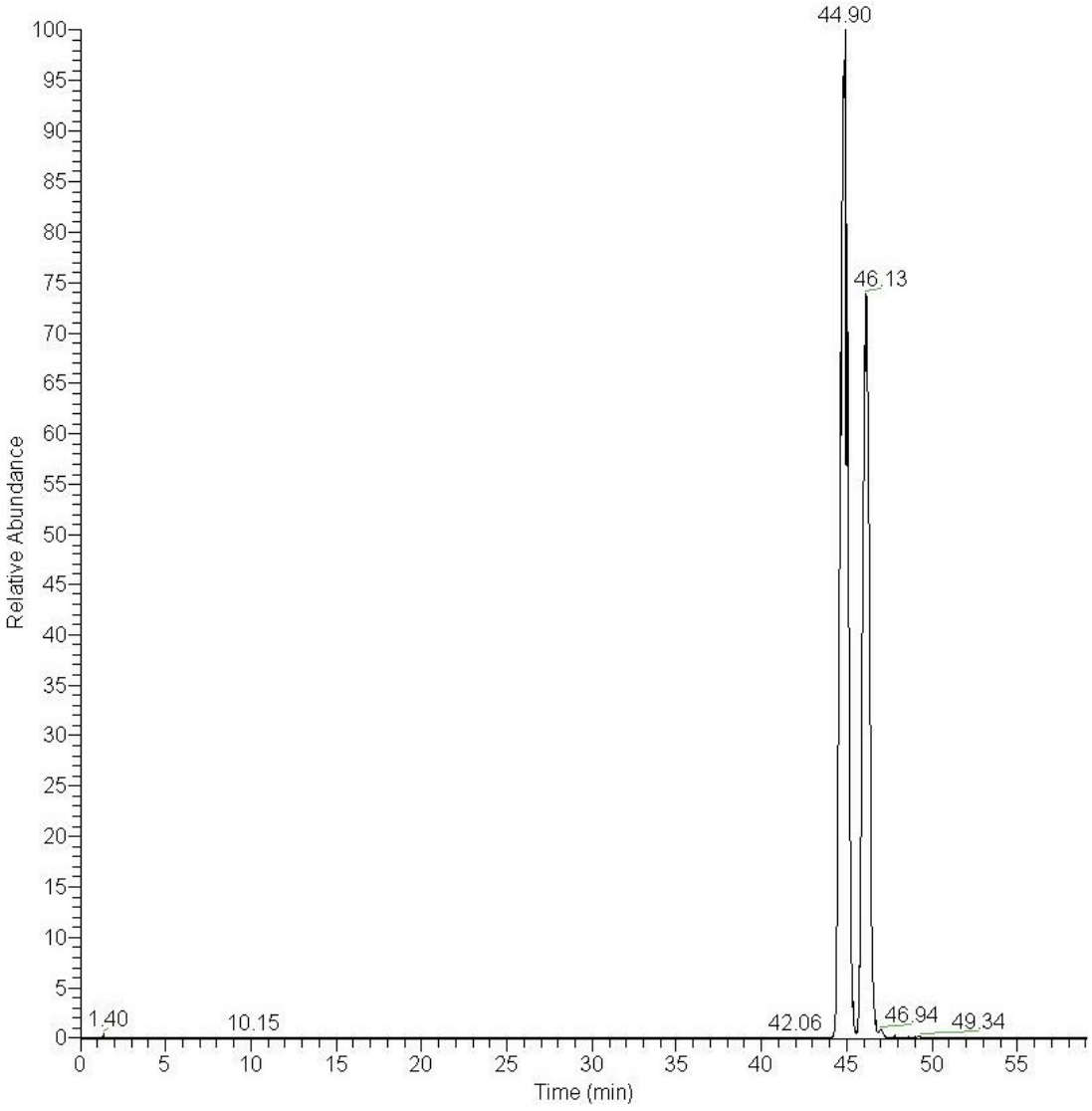
(x) MS-MS assignments for prenylated cyclo[KKPYILP] single peak; assignments consistent with presence of O-prenylated compound

cyclo[KKPYILP] single peak					
y- b- cleavages (+1)	**Expected**	**Observed**	**y- b- cleavages (+1)**	**Expected**	**Observed**
KK	257.20	-	**ILPKKP**	677.47	-
KKP	354.25	354.27	**LP**	211.14	-
KKPy	585.37	-	**LPK**	339.24	339.28
KKPyI	698.46	698.44	**LPKK**	467.33	467.34
KKPyIL	811.54	811.53	**LPKKP**	564.39	-
KP	226.16	-	**LPKKPy**	795.51	795.45
KPy	457.28	-	**PK**	226.16	-
KPyI	570.36	-	**PKK**	354.25	354.27
KPyIL	683.45	-	**PKKP**	451.30	-
KPyILP	780.50	780.49	**PKKPy**	682.43	682.43
Py	329.18	329.17	**PKKPyI**	795.51	795.45
PyI	442.27	442.16			
PyIL	555.35	555.24	**y- b- cleavages (+2)**		
PyILP	652.40	-	**KKPyI**	349.74	349.83
PyILPK	780.50	780.49	**PKKPyI**	398.26	398.35
yI	345.22	345.23	**PKKPy**	341.72	341.87
yIL	458.30	-			
yILP	555.35	555.24	**Miscellaneous Ions**		
yILPK	683.45	-	**All -prenyl +1**	840.54	840.46
yILPKK	811.54	811.53	**All -prenyl +2**	420.77	420.85
IL	227.18	227.21	**All -NH3, +2**	446.29	446.31
ILP	324.23	-	**All -acylium, +2**	440.81	440.98
ILPK	452.32	-	**Py y- a- cleavage +1**	301.19	301.21
ILPKK	580.42	580.39			

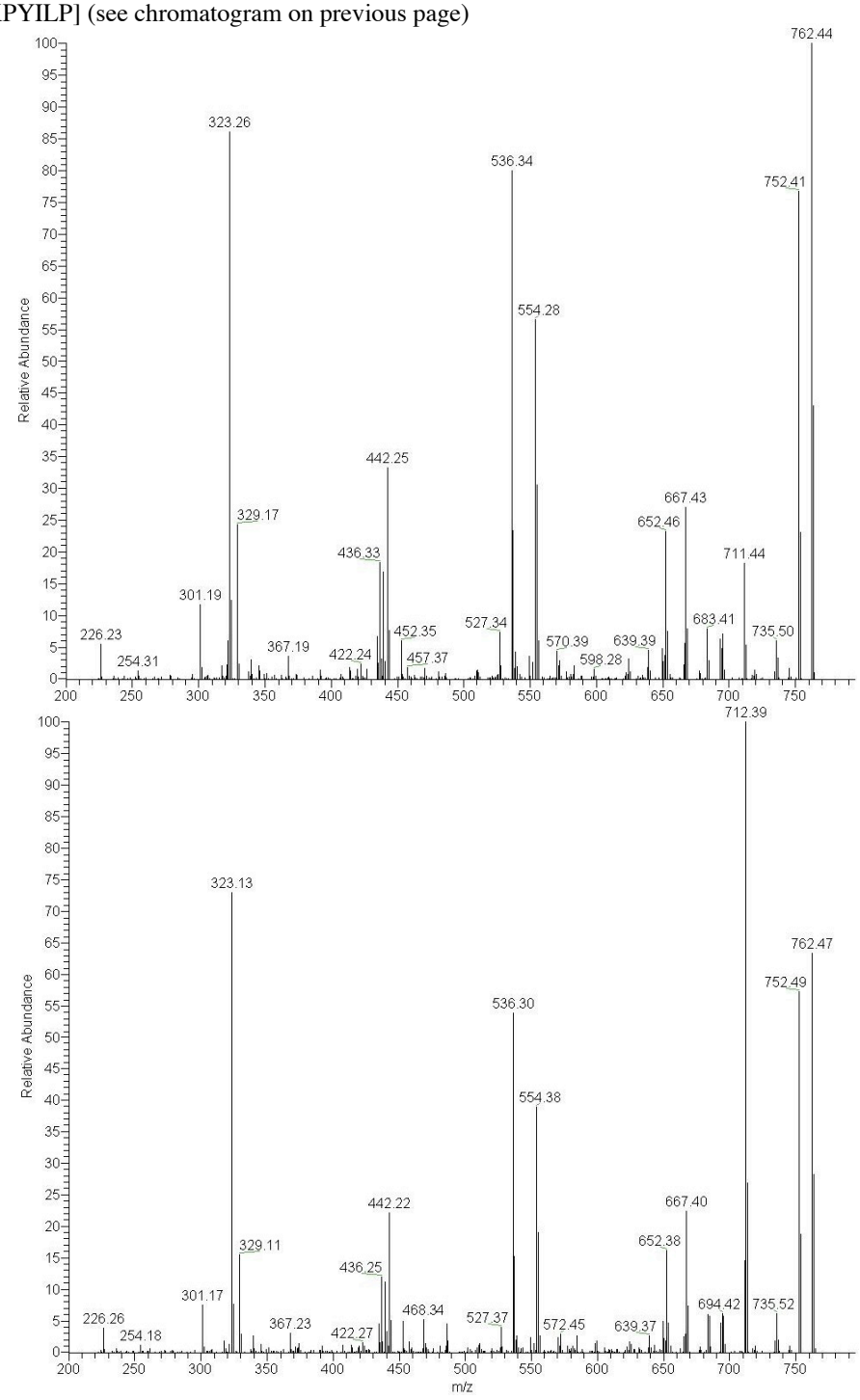
(y) FT-ICR data showing ion whose mass (+/- 2 ppm) corresponds to prenylated cyclo[KPYILP]



(z) LC-FT-ICR chromatogram selected for mass of prenylated cyclo[KPYILP]



(aa) MS-MS spectra for early (top) and late (bottom) eluting products of reaction of LynF with cyclo[KPYILP] (see chromatogram on previous page)

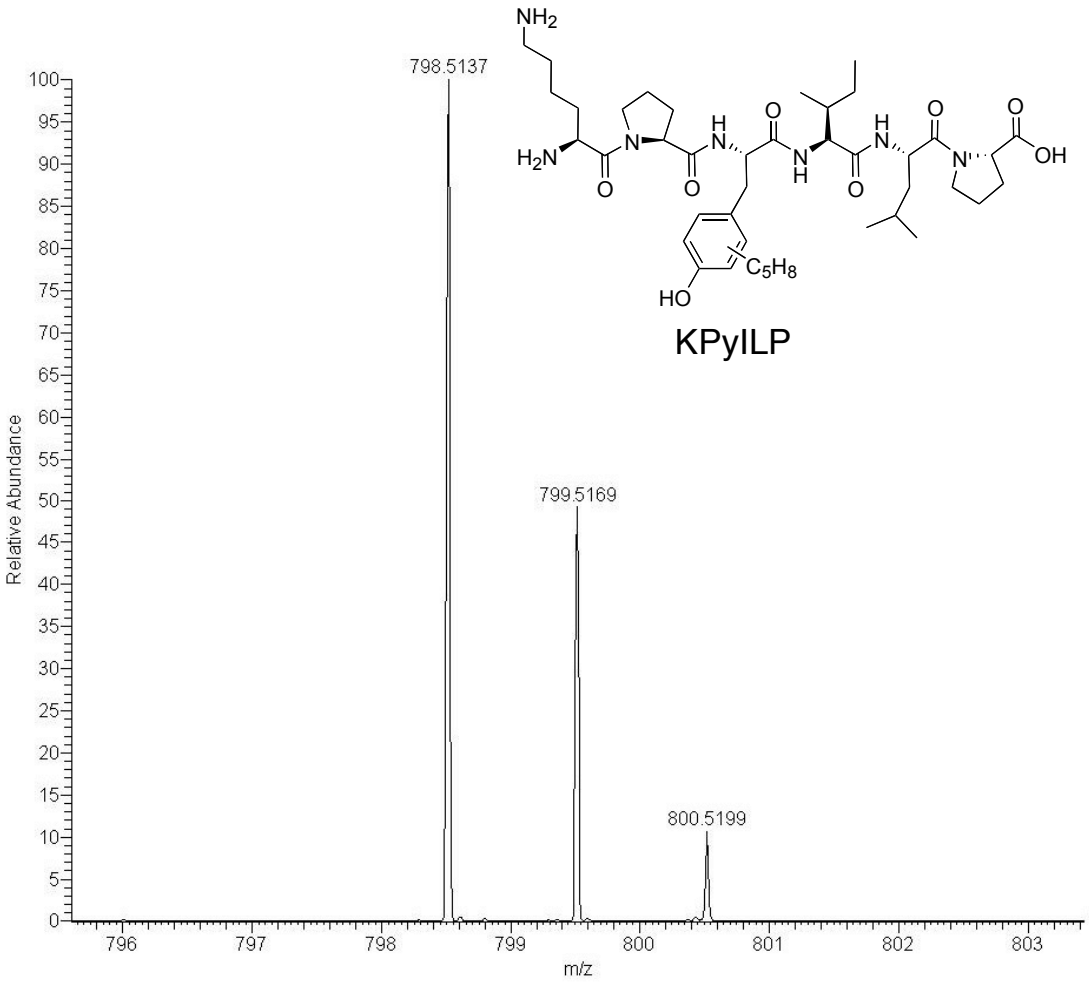


(bb) MS-MS assignments for prenylated cyclo[KPYILP] early (top) and late (bottom) eluting isomers

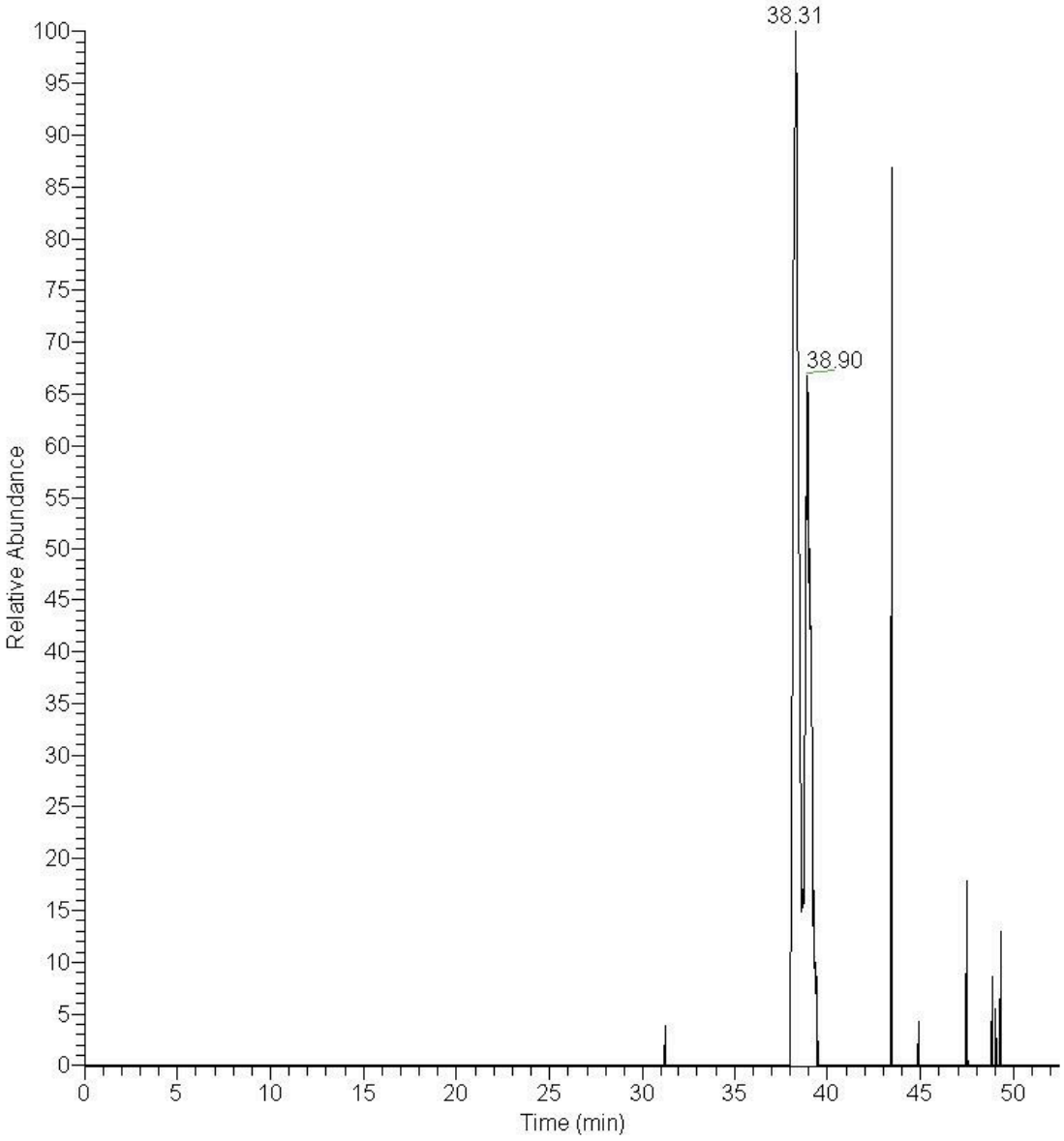
cyclo[KPYILP] 1st peak (C-prenyl)					
y- b- cleavages (+1)	**Expected**	**Observed**	**y- b- cleavages (+2)**	**Expected**	**Observed**
KP	226.16	226.23	**KPyIL**	342.23	342.25
KPy	457.28	457.37	**PKPyI**	334.22	334.27
KPyI	570.36	570.39			
KPyIL	683.45	683.41	**Miscellaneous Ions**		
Py	329.18	329.17	**All -H2O**	762.49	762.44
PyI	442.27	442.25	**All -acylium**	752.51	752.41
PyIL	555.35	555.32	**PKPy -H2O**	536.32	536.34
PyILP	652.40	652.46	**Py y- a- cleavage +1**	301.19	301.19
yI	345.22	345.19			
yIL	458.30	-			
yILP	555.35	555.32			
yILPK	683.45	683.41			
IL	227.18	-			
ILP	324.23	-			
ILPK	452.32	452.35			
ILPKP	549.38	549.36			
LP	211.14	-			
LPK	339.24	339.31			
LPKP	436.29	436.33			
LPKPy	667.42	667.43			
PK	226.16	226.23			
PKP	323.21	323.26			
PKPy	554.33	554.28			
PKPyI	667.42	667.43			

cyclo[KPYILP] 2nd peak (O-prenyl)					
y- b- cleavages (+1)	**Expected**	**Observed**	**y- b- cleavages (+2)**	**Expected**	**Observed**
KP	226.16	226.26	**KPyIL**	342.23	-
KPy	457.28	-	**PKPyI**	334.22	-
KPyI	570.36	-			
KPyIL	683.45	-	**Miscellaneous Ions**		
Py	329.18	329.11	**All -H2O**	762.49	762.47
PyI	442.27	442.22	**All -acylium**	752.51	752.49
PyIL	555.35	555.25	**All -prenyl**	712.44	712.39
PyILP	652.40	652.38	**All -H2O, -prenyl**	694.43	694.42
yI	345.22	345.11	**PKPy -H2O**	536.32	536.30
yIL	458.30	-	**Py y- a- cleavage +1**	301.19	301.17
yILP	555.35	555.32			
yILPK	683.45	-			
IL	227.18	-			
ILP	324.23	-			
ILPK	452.32	-			
ILPKP	549.38	-			
LP	211.14	-			
LPK	339.24	-			
LPKP	436.29	436.25			
LPKPy	667.42	667.40			
PK	226.16	226.26			
PKP	323.21	323.13			
PKPy	554.33	554.38			
PKPyI	667.42	667.40			

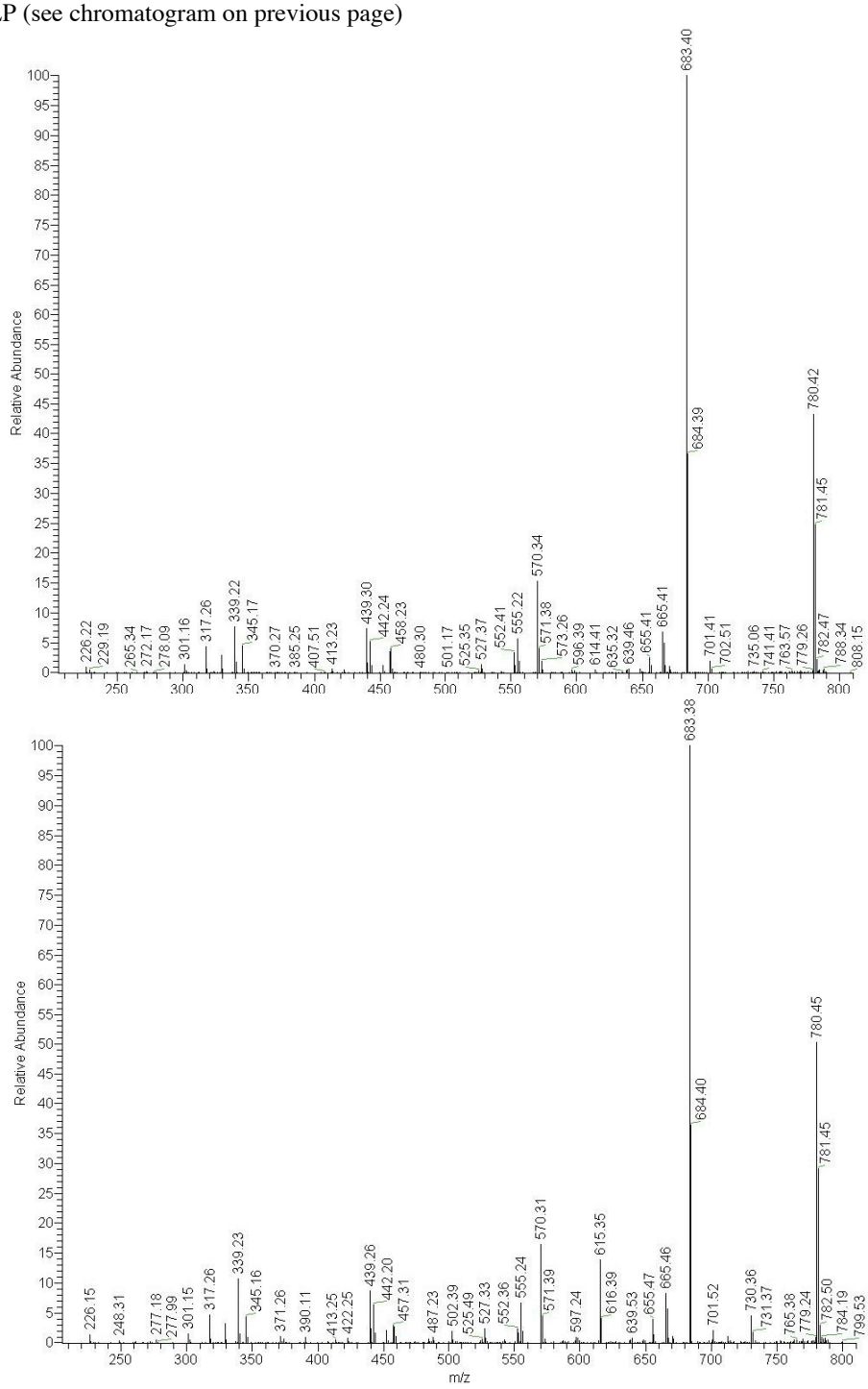
(cc) FT-ICR data showing ion whose mass (+/- 2 ppm) corresponds to prenylated KPYILP



(dd) LC-FT-ICR chromatogram selected for mass of prenylated KPYILP



(ee) MS-MS spectra for early (top) and late (bottom) eluting products of reaction of LynF with KPYILP (see chromatogram on previous page)

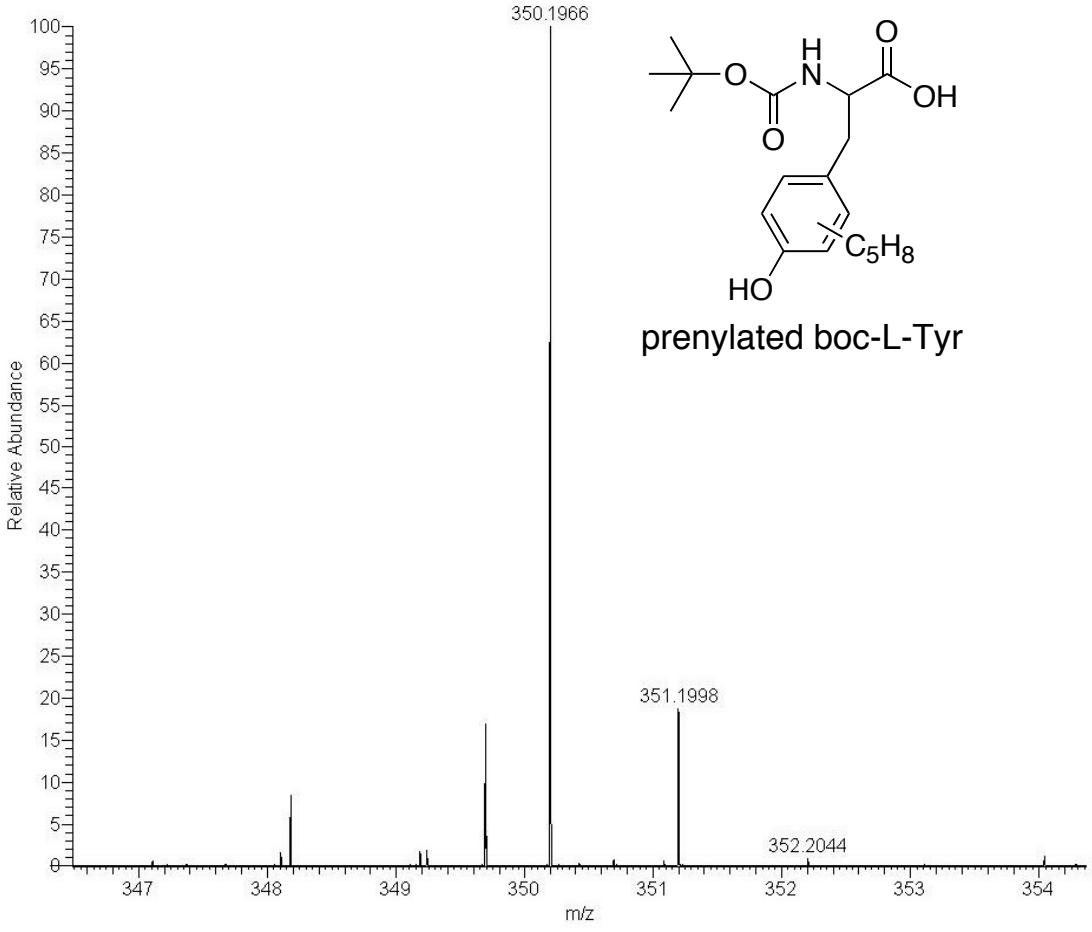


(ff) MS-MS assignments for prenylated KPYILP early (left) and late (right) eluting isomers

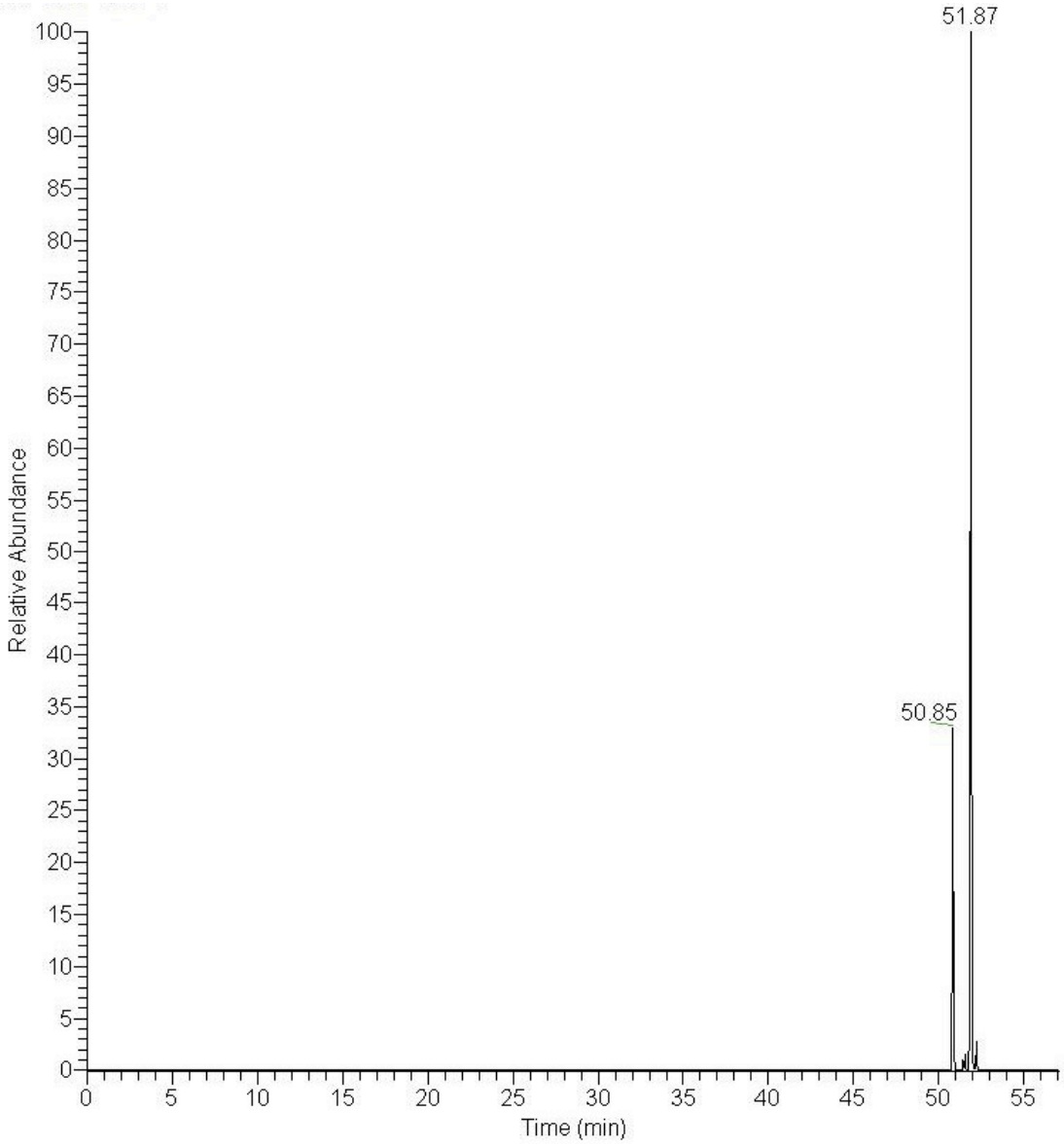
KPYILP 1st peak (C-prenyl)		
b-ions (+1)	**Expected**	**Observed**
KP	226.16	226.22
KPy	457.28	457.25
KPy°	439.27	439.30
KPyI	570.36	570.34
KPyI°	552.35	552.41
KPyIL	683.45	683.41
KPYIL°	665.44	665.41
KPyILP°	780.50	780.42
y-ions (+1)		
PyILP	670.41	-
yILP	573.36	573.26
ILP	342.24	-
LP	229.15	-
y- b- cleavages (+1)		
LPK	339.24	339.22
yI	345.22	345.17
yIL	458.30	458.23
PyI	442.27	442.24
PyIL	555.35	555.22
y- a- cleavages (+1)		
yI	317.22	317.26
yILP	527.36	527.37
Miscellaneous Ions		
KPyIL a-cleavage +1	655.45	655.47

KPYILP 2nd peak (O-prenyl)		
b-ions (+1)	**Expected**	**Observed**
KP	226.16	226.13
KPy	457.28	457.31
KPy°	439.27	439.26
KPyI	570.36	570.31
KPyI°	552.35	552.35
KPyIL	683.45	683.38
KPYIL°	665.44	665.46
KPyILP°	780.50	780.44
y-ions (+1)		
PyILP	670.41	670.35
yILP	573.36	-
ILP	342.24	-
LP	229.15	-
y- b- cleavages (+1)		
Py	329.19	329.17
LPK	339.24	339.24
PyI	442.27	442.20
PyIL	555.35	555.24
yI	345.22	345.15
yIL	458.30	458.31
y- a- cleavages (+1)		
yI	317.22	317.24
yILP	527.36	527.31
Miscellaneous Ions		
KPYILP -prenyl +1	730.45	730.40
KPYILP a-cleavage +1	655.45	655.46
KPyIL -prenyl +1	615.39	615.36
KPyI -prenyl +1	502.30	502.39

(gg) FT-ICR data showing ion whose mass (+/- 2 ppm) corresponds to prenylated boc-L-tyrosine



(hh) LC-FT-ICR chromatogram selected for the mass of prenylated boc-L-tyrosine



(jj) ESI-MS spectrum showing ion corresponding to the mass of prenylated N-acetyl-L-Tyr

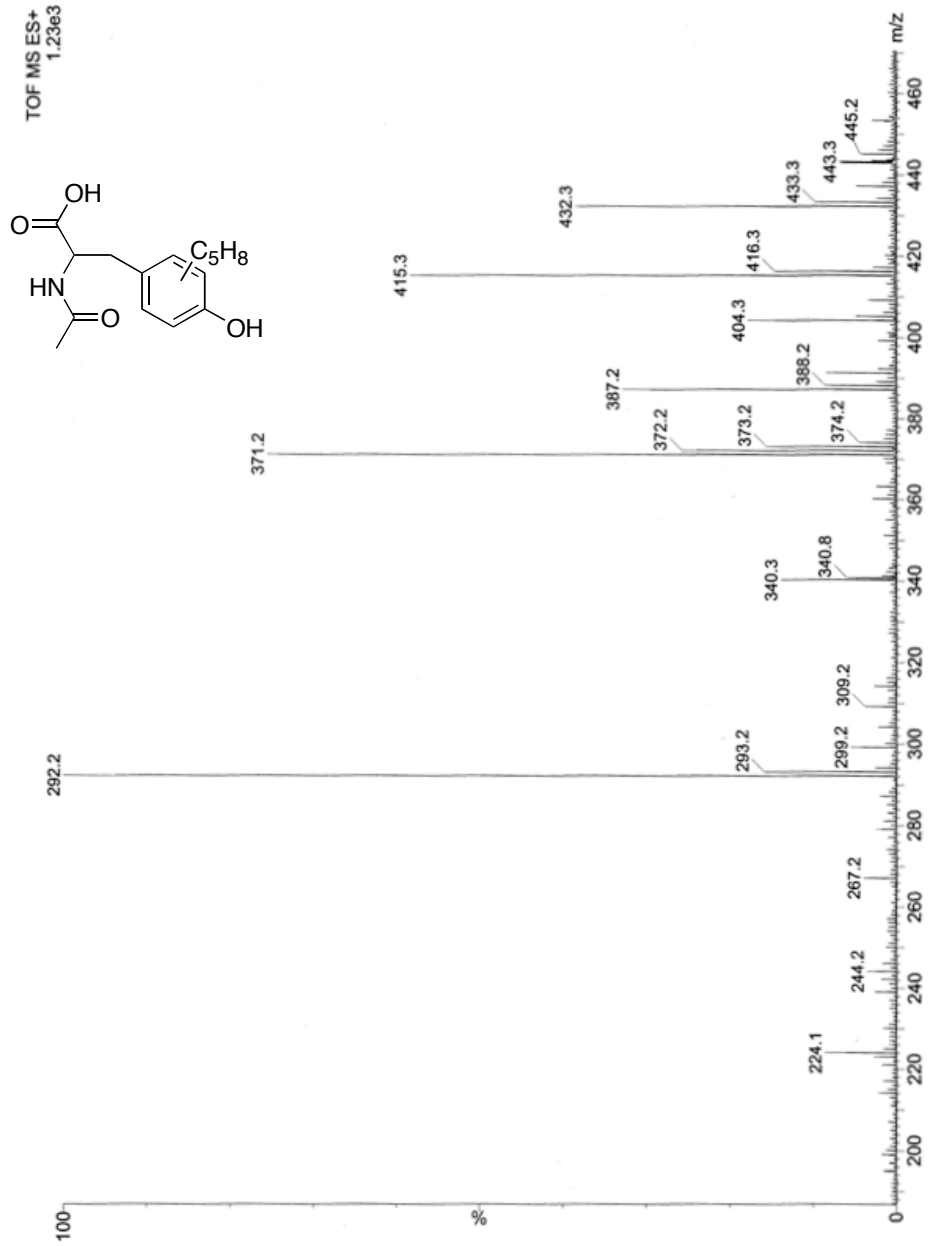
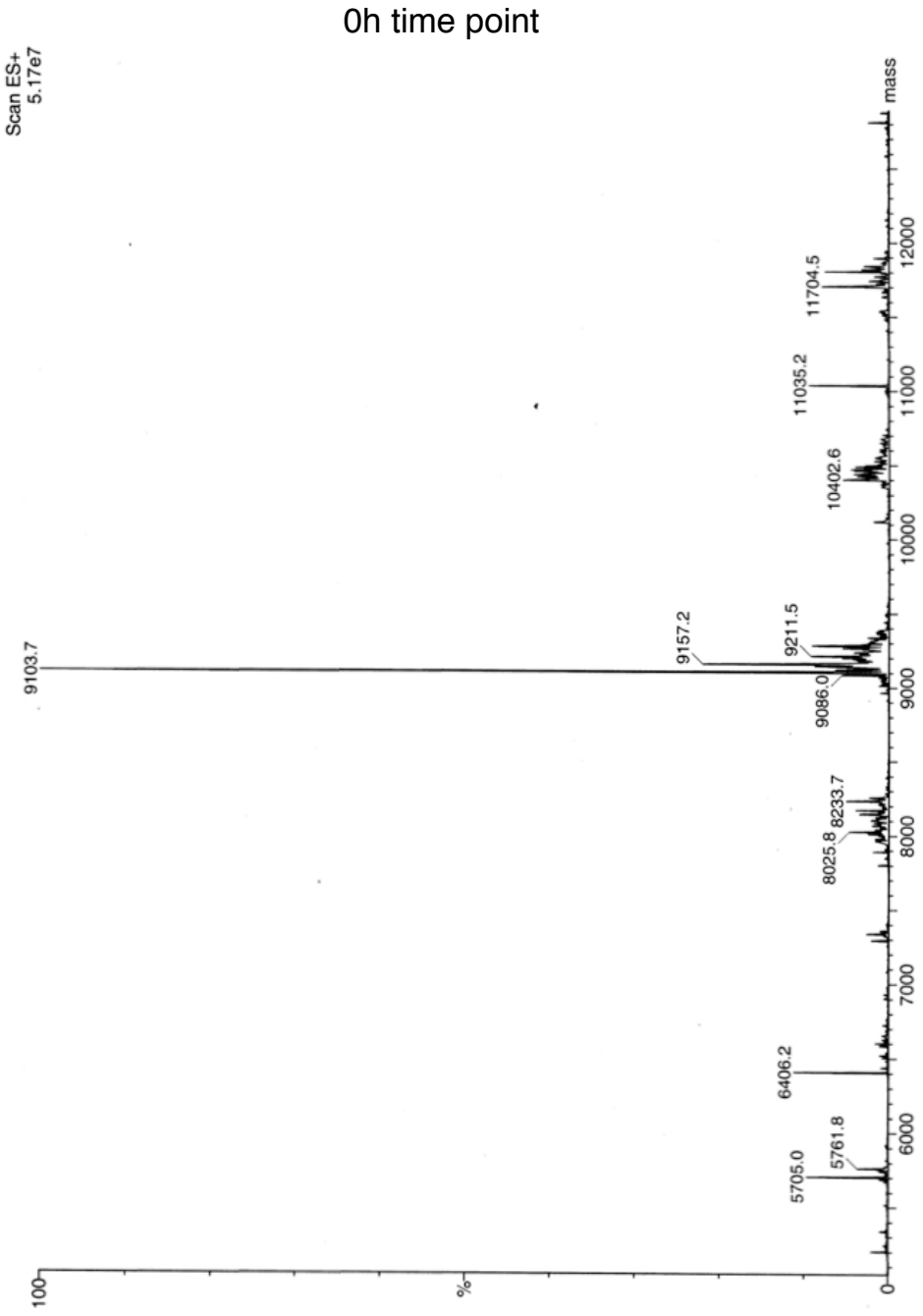
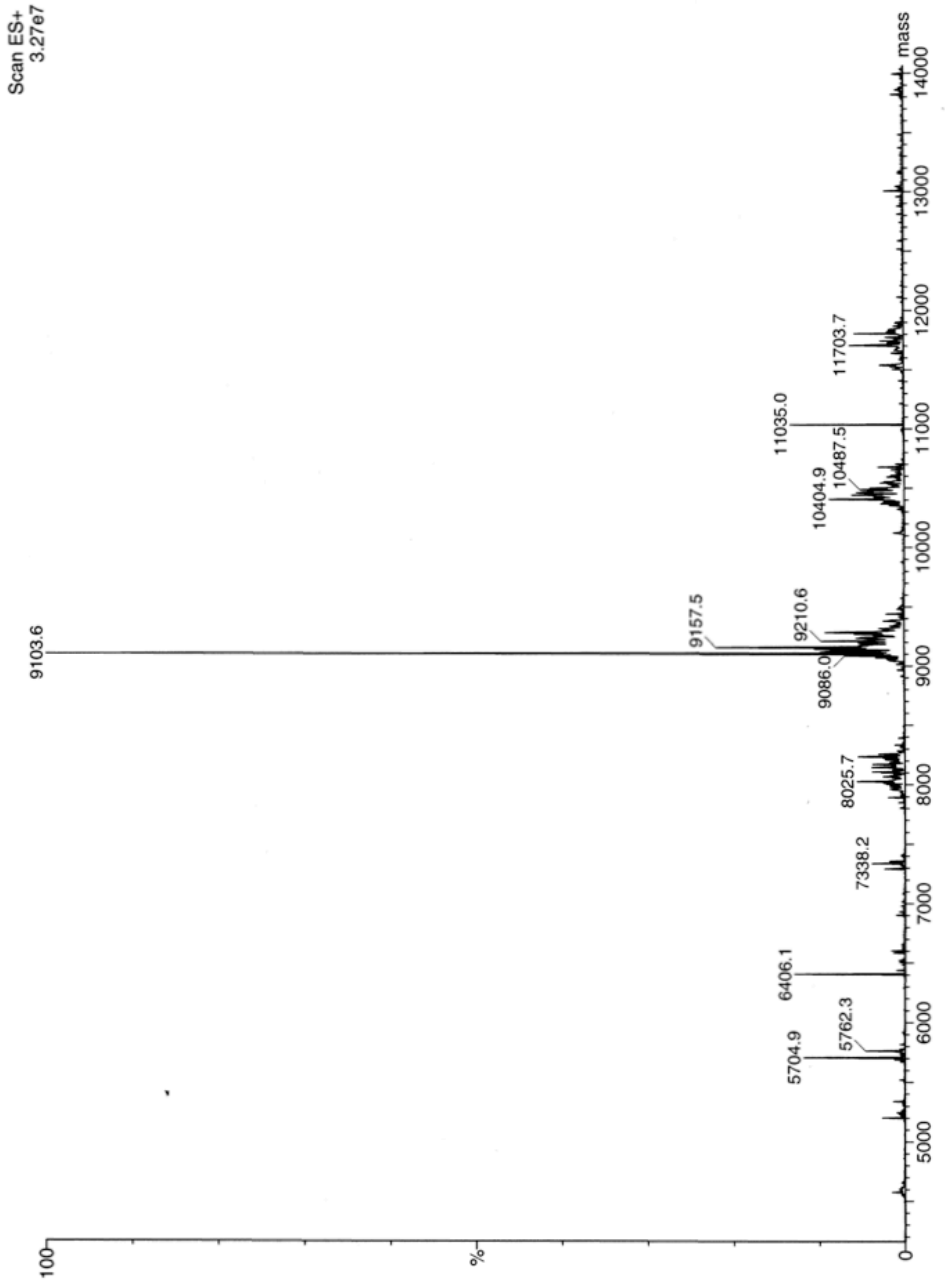


Figure S2. Precursor peptides are not substrates for prenyltransfer (a) ESI-MS intact analysis of time points taken from a reaction containing TruLy1, LynF and standard additives at 0, 6, 12, 24 h (deconvoluted spectra shown); expected mass of unmodified TruLy1 is 9104 Da. No reaction is observed even after 24 h incubation at 37 °C (b) Incorporation of tritiated DMAPP into various substrates; background subtracted counts-per-minute (CPM) shown

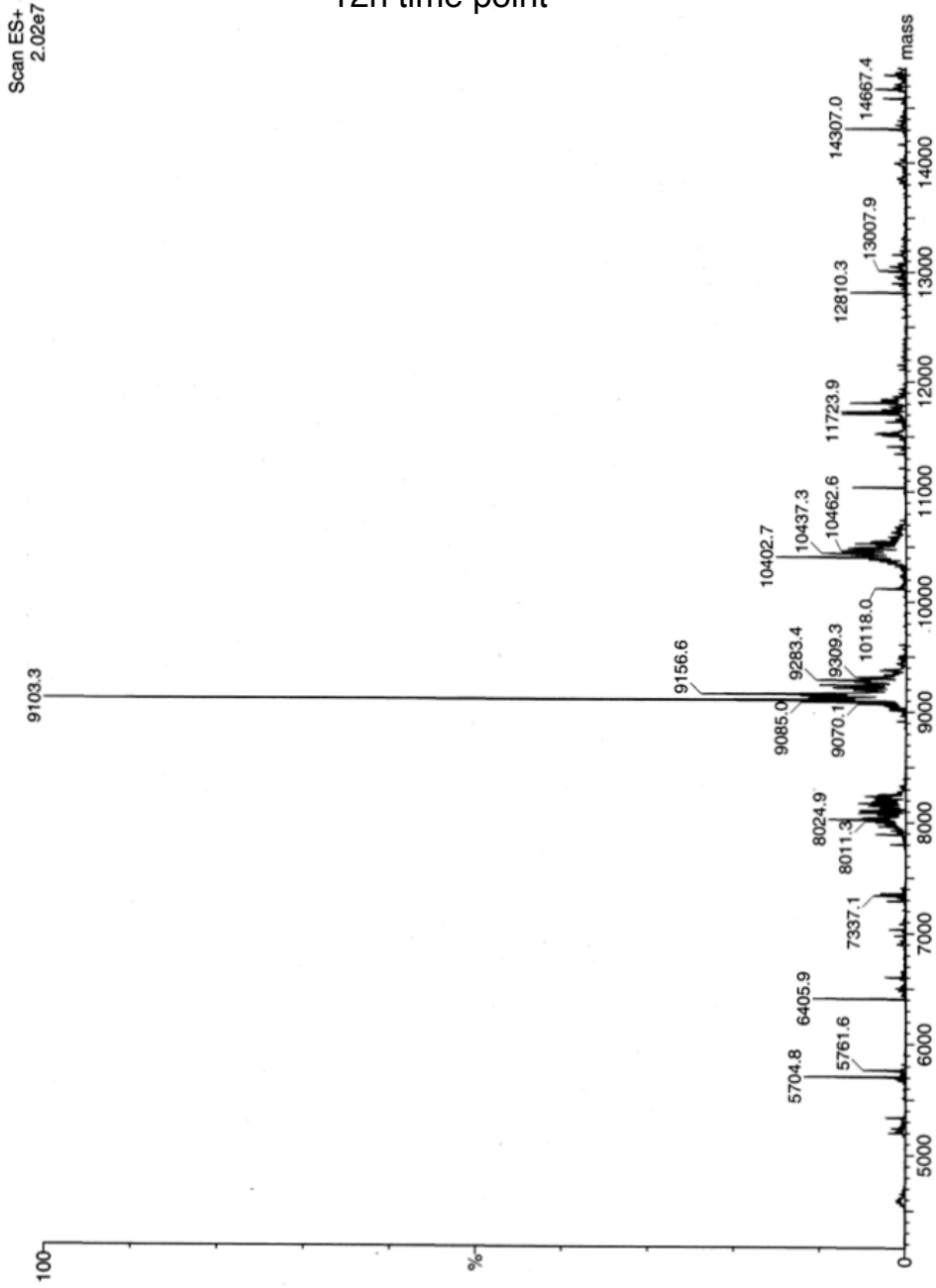
(a)


0h time point
Scan ES+
5.17e7
mass
12000
11000
10000
9000
8000
7000
6000
11704.5
11035.2
10402.6
9211.5
9157.2
9103.7
9086.0
8233.7
8025.8
6406.2
5761.8
5705.0
100
%
0

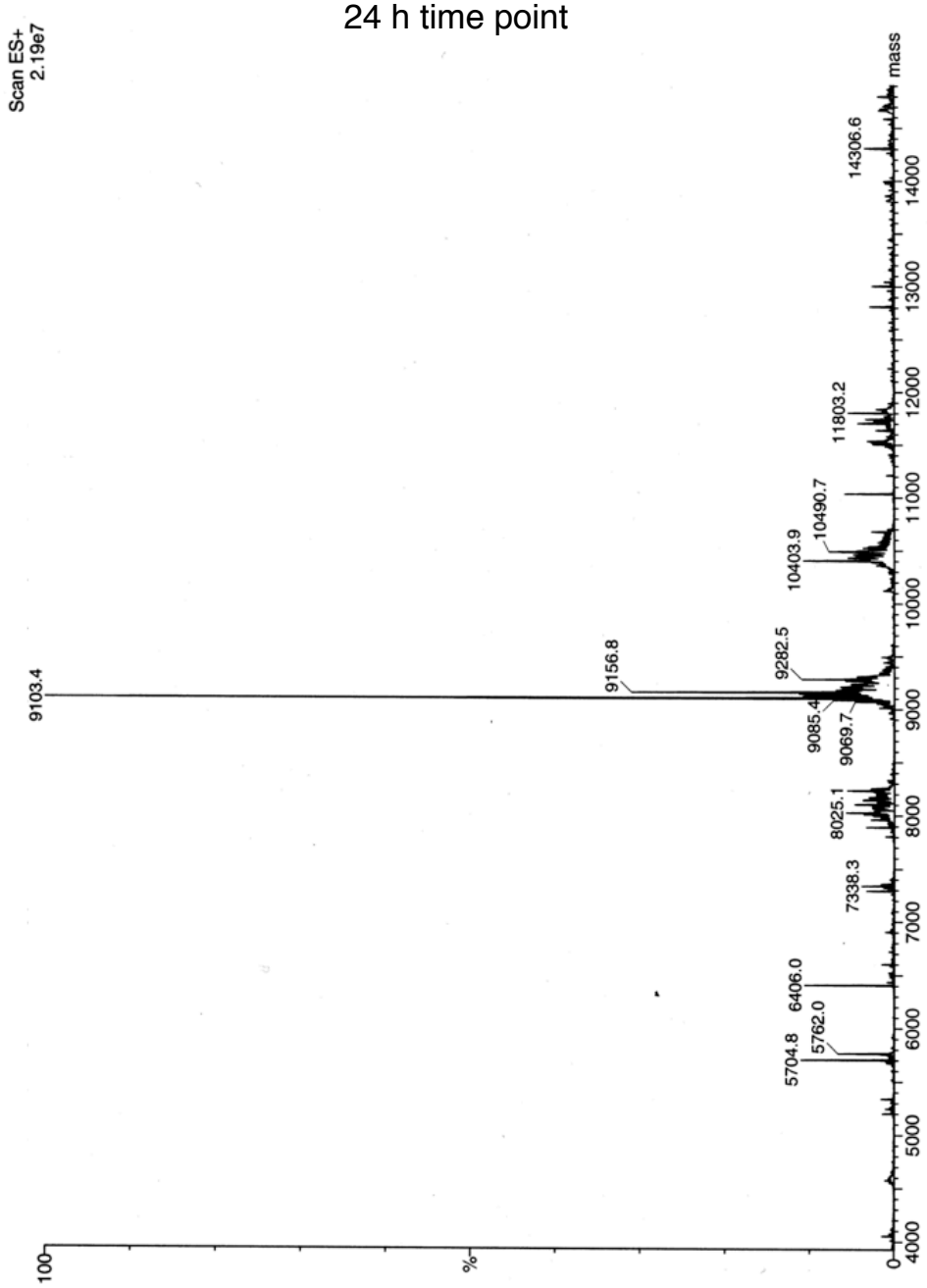
6h time point



12h time point



24 h time point



(b)

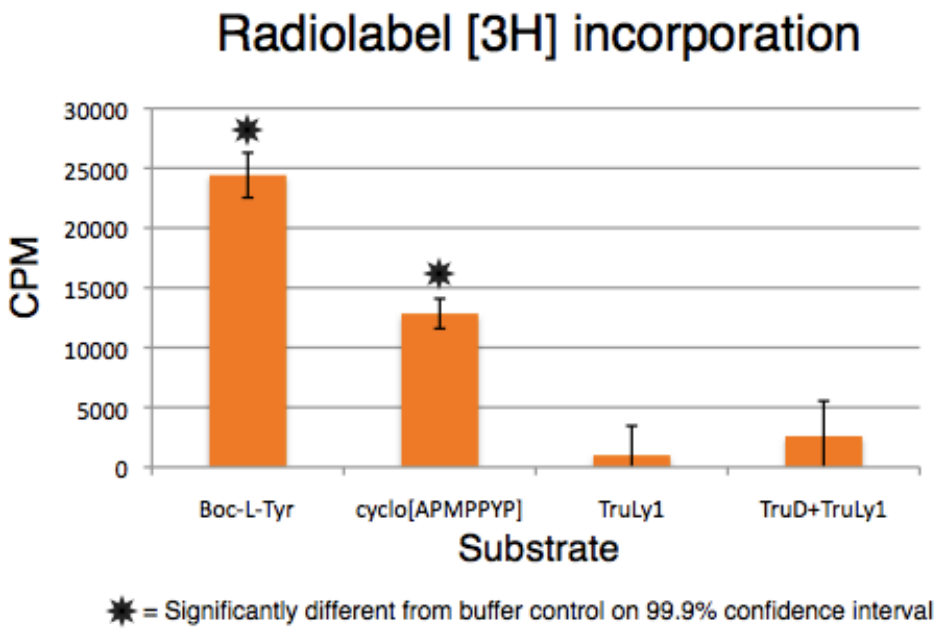
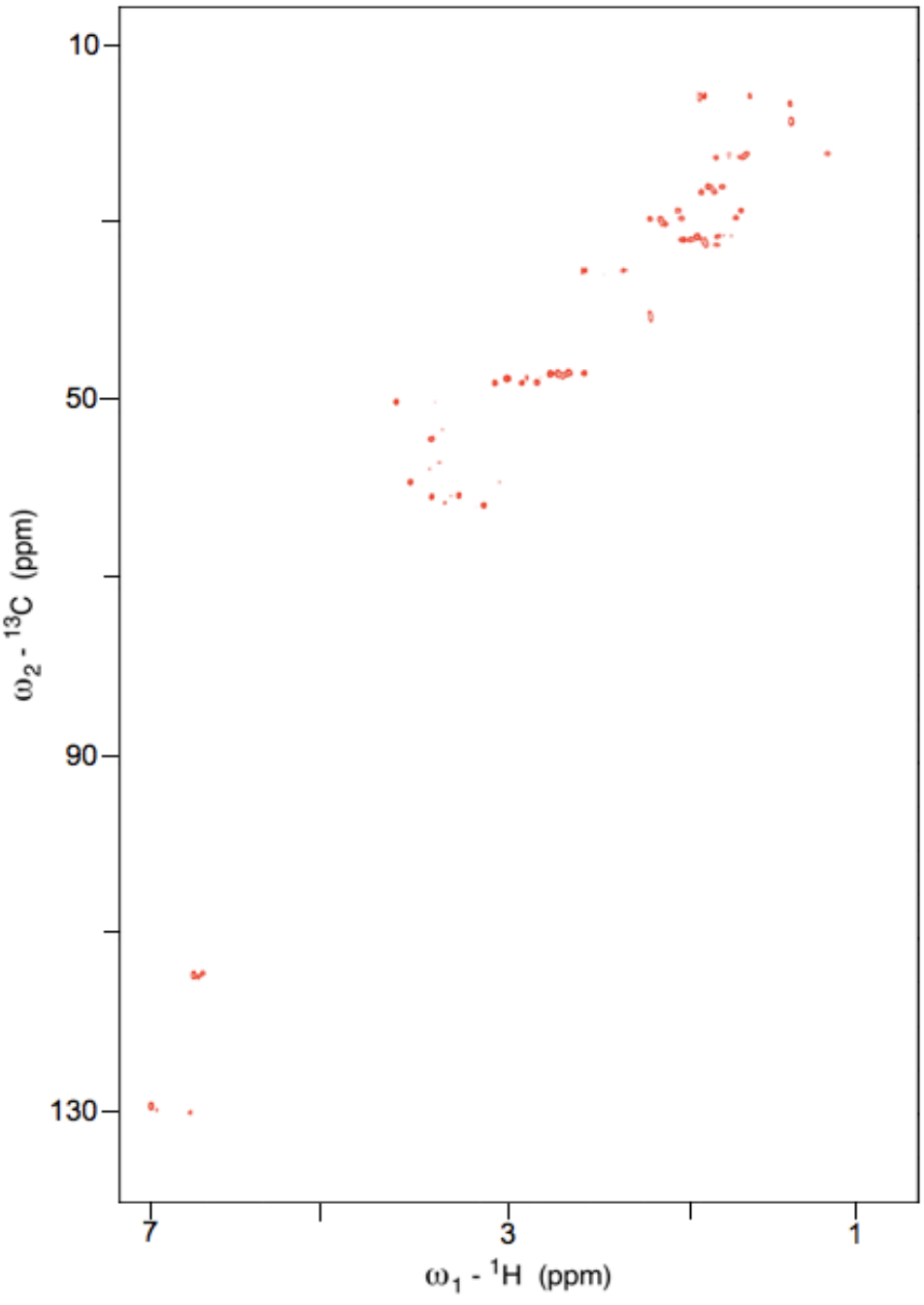
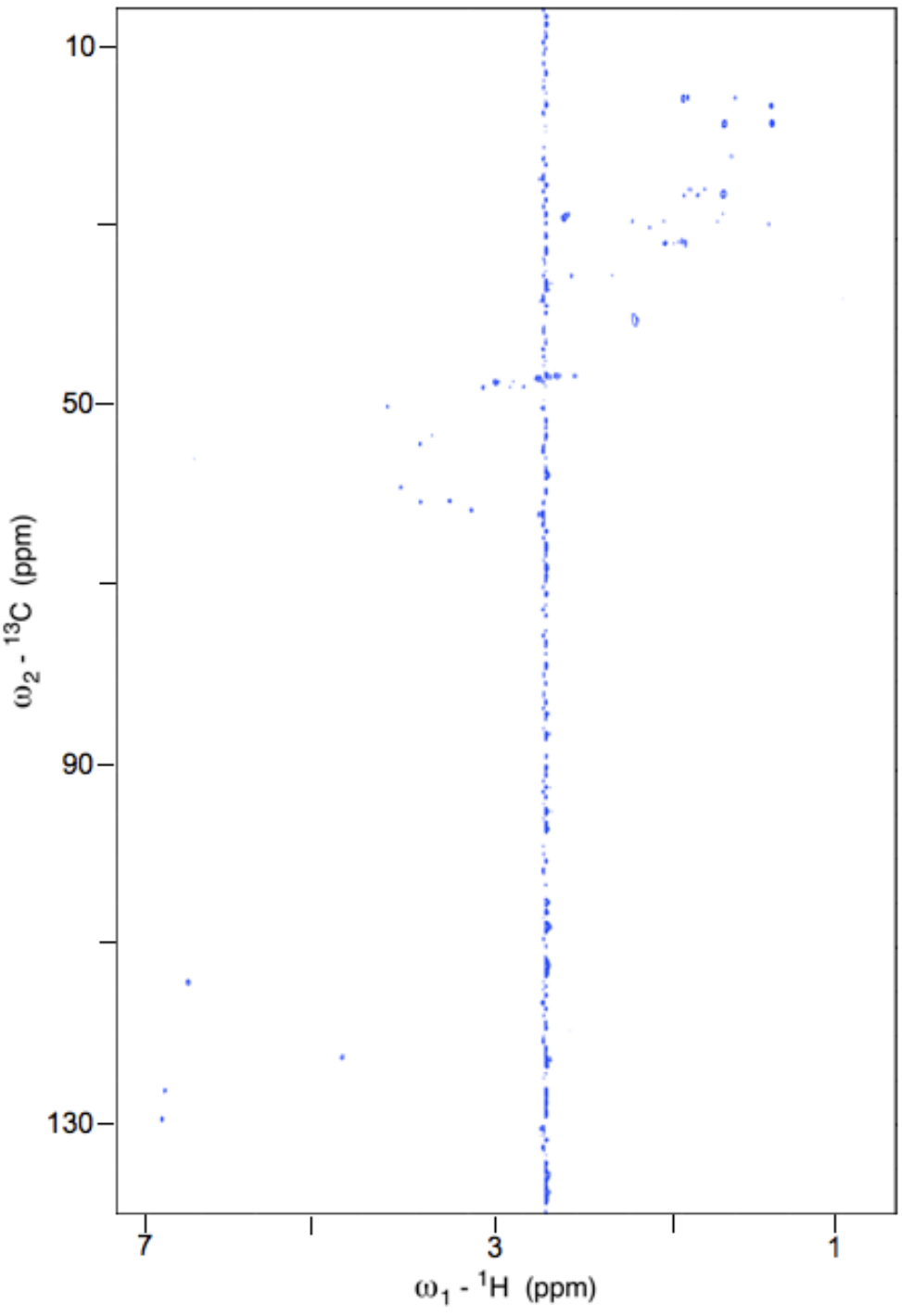

Radiolabel [3H] incorporation
30000
25000
20000
15000
10000
5000
0
CPM
Boc-L-Tyr
cyclo[APMPPYP]
TruLy1
TruD+TruLy1
Substrate
= Significantly different from buffer control on 99.9% confidence interval

Figure S3. NMR characterization of substrates cyclo[APMPPYP] and boc-L-tyrosine (a) HSQC of cyclo[APMPPYP] (b) HSQC of prenylated cyclo[APMPPYP] (c) Overlay of (a) and (b) data (d) table of assignments (e) HSQC of C-prenyl boc-L-Tyr (f) Table of assignments and numbered structure

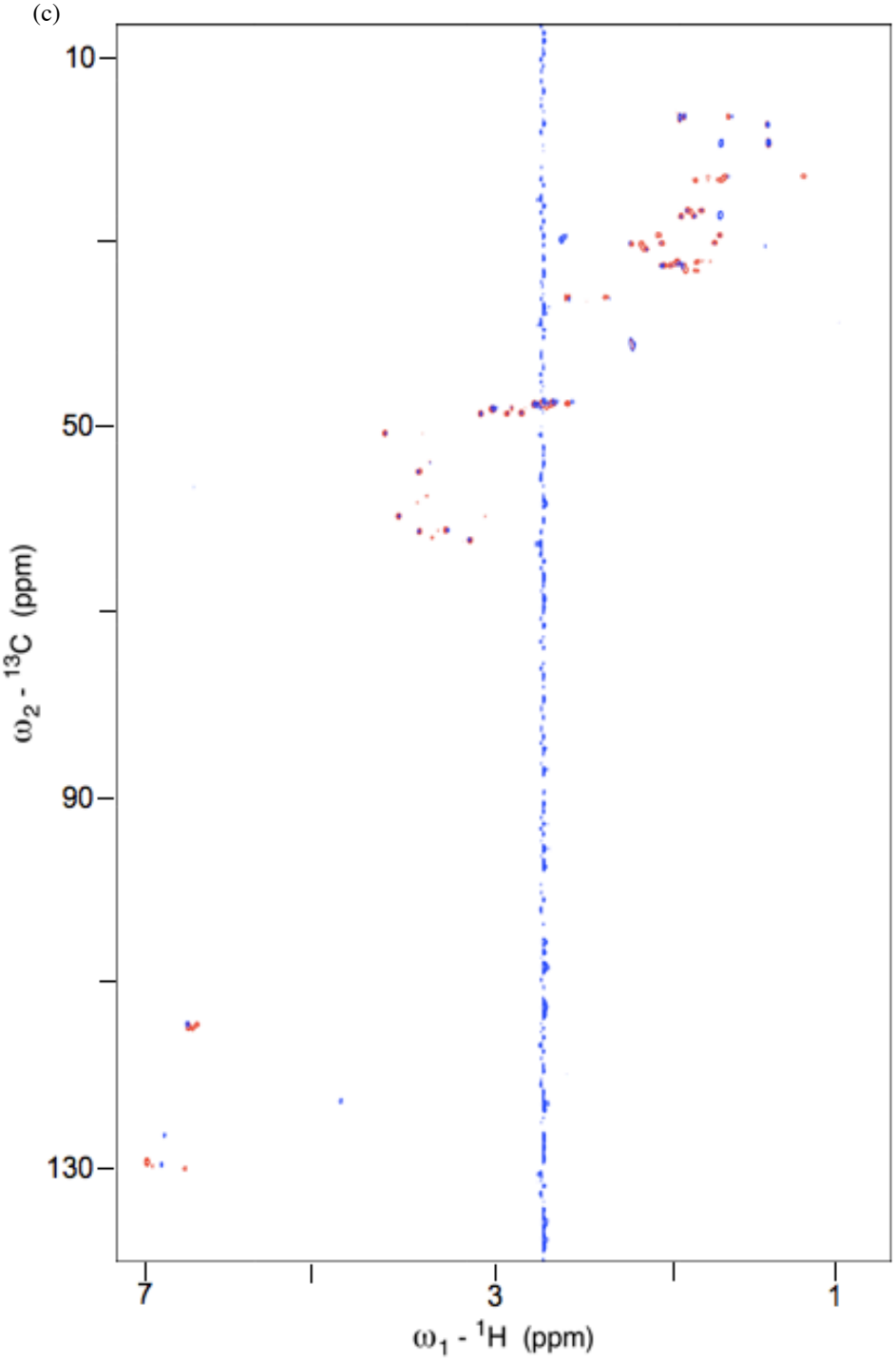
(a)



(b)



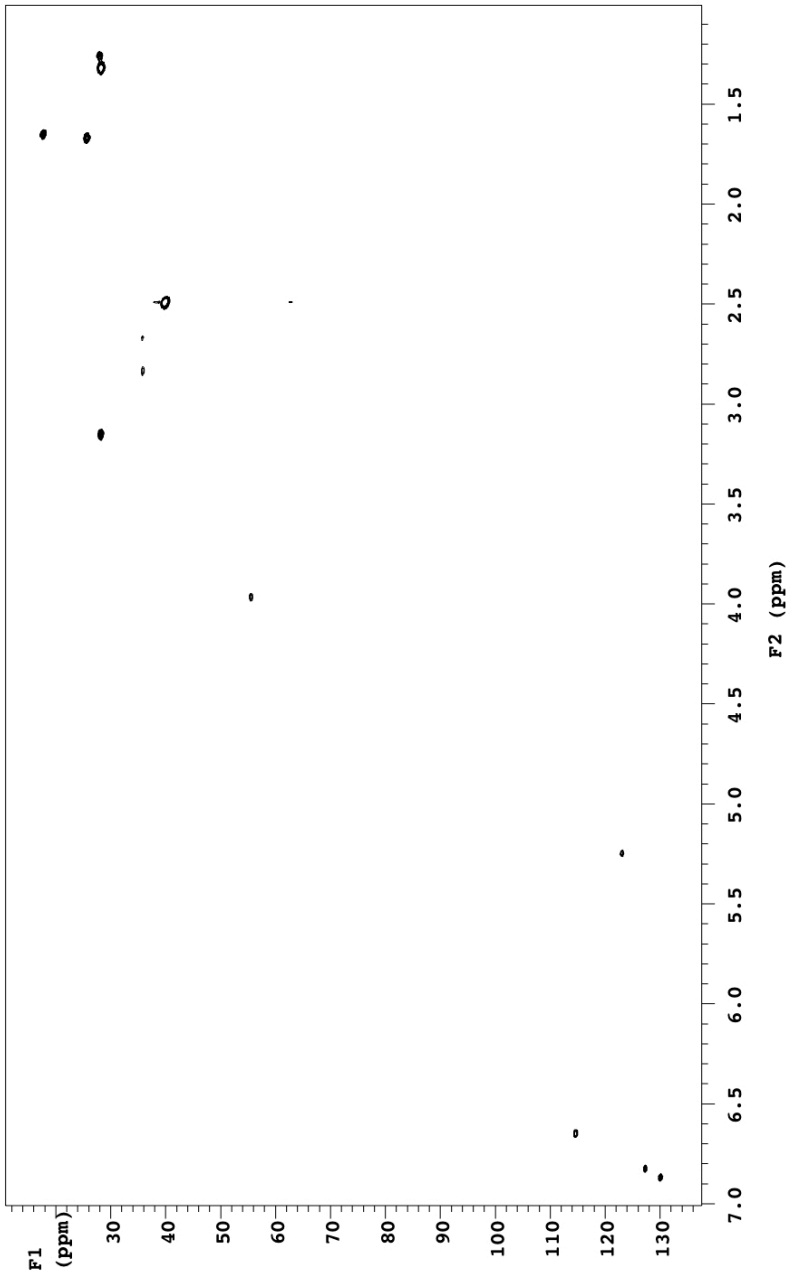
(c)



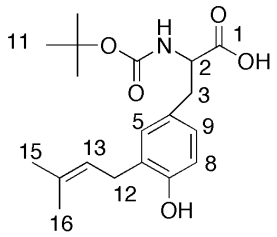
(d)

cyclo[APMPPYP] unmod	13C	1H	cyclo[APMPPY*P] product	13C	1H
Tyr Cδ	129.86	7.06	**Tyr Cδ1**	130.08	6.92
Tyr Cε	114.96	6.67	**Tyr Cδ2**	126.85	6.89
Pro Cα	61.12	4.01	**C2 of 3-methyl-2-buten-1-yl adduct**	122.9	5.23
Pro Cα	60.15	4.25	**Tyr Cε**	114.61	6.68
Pro Cα	60.25	4.50	**Pro Cα**	61.07	4.02
Pro Cα	58.64	4.69	**Pro Cα**	60.13	4.22
Tyr Cα	53.66	4.50	**Pro Cα**	60.13	4.50
Met Cα	49.52	4.82	**Pro Cα**	58.59	4.69
Ala Cα	47.22	3.91	**Tyr Cα**	53.57	4.50
Pro Cδ	47.22	3.67	**Met Cα**	49.48	4.81
		3.53	**Ala Cα**	47.24	3.91
Pro Cδ	46.74	3.81	**Pro Cδ**	47.06	3.66
Pro Cδ	46.12	3.41			3.53
		3.34	**Pro Cδ**	46.75	3.79
Pro Cδ	46.09	3.24			3.62
		3.09	**Pro Cδ**	46.02	3.40
DMSO	39.51	2.49	**Pro Cδ**	45.79	3.22
Tyr Cβ	34.53	3.11			3.05
		2.74	**DMSO**	39.51	3.49
Pro Cβ	31.4	1.98	**Tyr Cβ**	34.61	3.08
		1.88			2.70
Met Cγ	30.82	2.19	**C1 of 3-methyl-2-buten-1-yl adduct**	28.13	3.15
		2.12	**Pro Cβ**	31.34	1.88
Pro Cβ	30.48	2.07		30.87	2.02
		1.88	**Met Cγ**	30.83	2.20
Met Cβ	28.57	2.50			2.12
		2.40	**Pro Cβ**	30.52	2.05
Pro Cβ	28.36	2.21			1.89
		1.71	**Met Cβ**	29.18	2.39
Pro Cβ	27.68	2.24	**Pro Cβ**	28.45	2.21
		1.67			1.71
Pro Cγ	25.5	2.02	**Pro Cβ**	27.62	2.24
		1.91			1.67
Pro Cγ	24.89	1.96	**Pro Cγ**	25.39	2.02
		1.83			1.90
Pro Cγ	21.5	1.89	**trans methyl of 3-methyl-2-buten-1-yl adduct**	25.27	1.65
		1.65	**Pro Cγ**	24.82	1.97
Pro Cγ	21.14	1.77			1.83
		1.61	**Pro Cγ**	21.43	1.91
Ala Cβ (rotamers)	17.32	1.20			1.62
	15.42	1.21	**Pro Cγ**	21.03	1.77
Met Cε (rotamers)	14.54	2.03			1.58
		1.99	**cis methyl of 3-methyl-2-buten-1-yl adduct**	17.31	1.65
		1.58	**Ala Cβ (rotamers)**	17.31	1.20
				15.45	1.21
			Met Cε (rotamers)	14.06	2.03
					1.99
					1.55

(e)


F1 (ppm)
30
40
50
60
70
80
90
100
110
120
130
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
F2 (ppm)

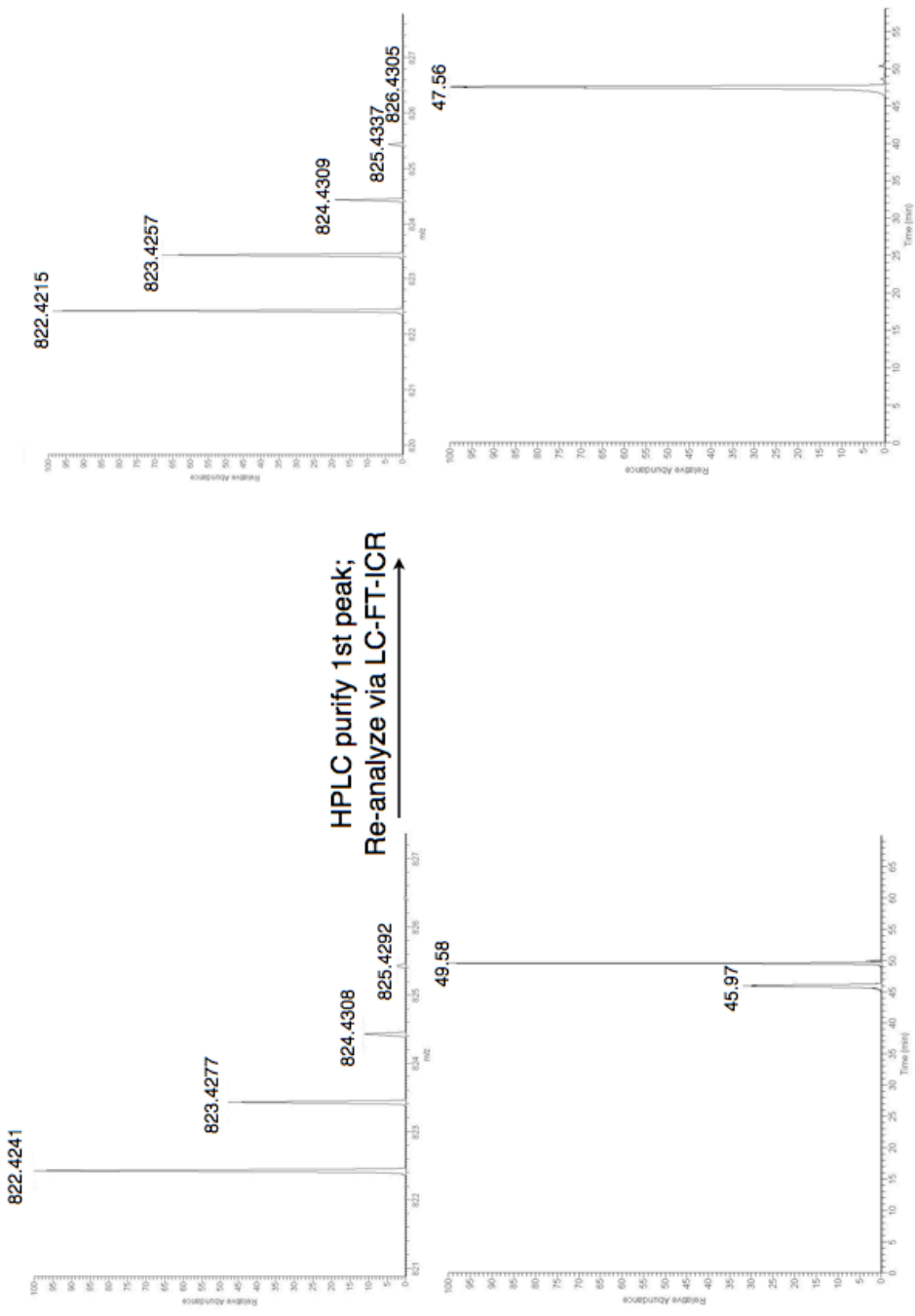
(f)



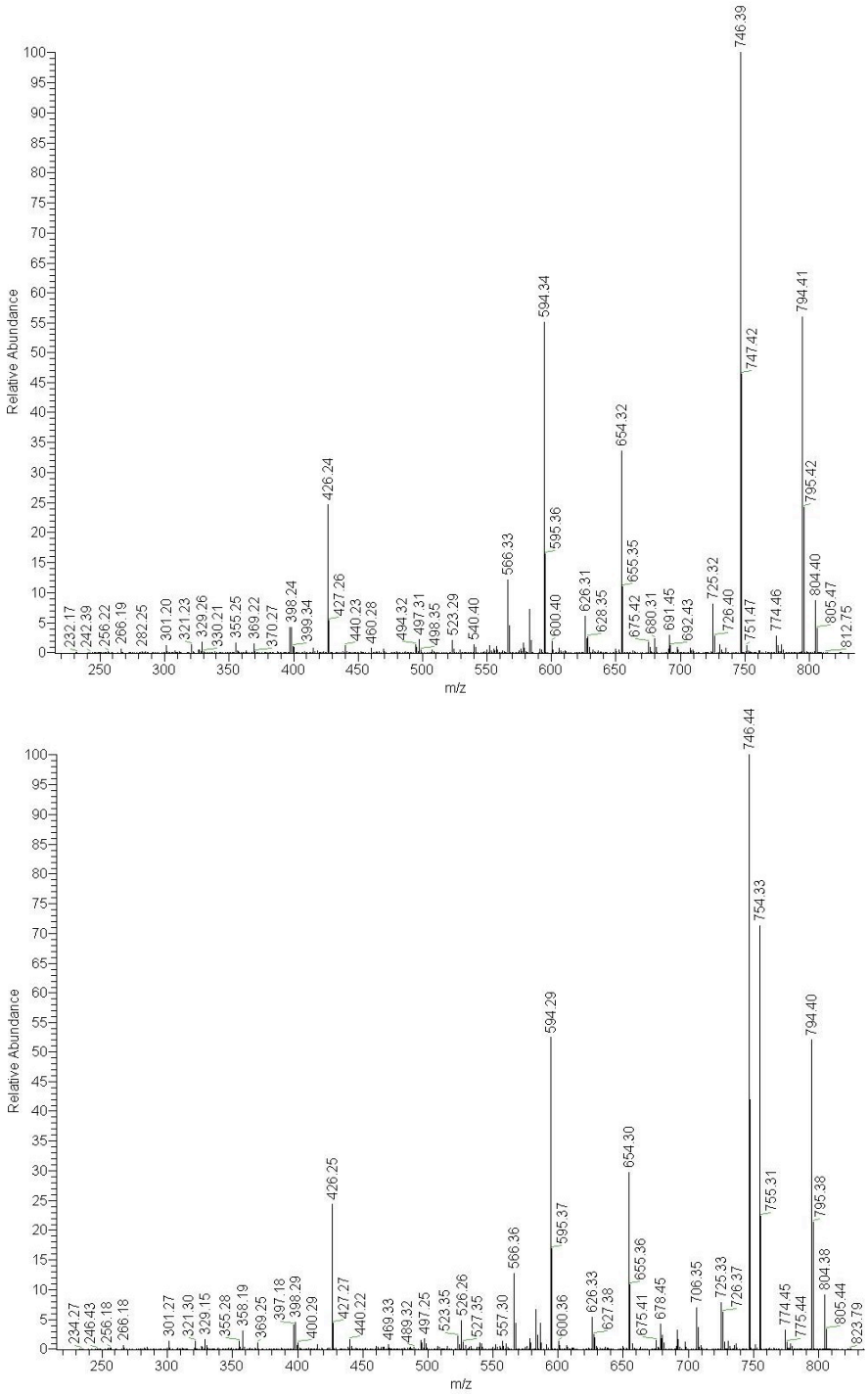
Position	δ 13C	δ 1H
5	129.5	6.87
9	126.8	6.82
13	122.6	5.25
8	114.2	6.65
2	55.0	3.96
DMSO	39.5	2.49
3a	35.5	2.66
3b	35.5	2.83
12	27.9	3.15
11a	27.7	1.32
11b	27.4	1.26
15	25.2	1.67
16	17.2	1.65

Figure S4. Spectral properties of C- and O-prenylated tyrosyl componds (a) Left: LC-FT-ICR peaks and chromatograms selected for the mass of prenylated cyclo[APMPPYP] from unpurified reaction mixtures; note that two separable compounds corresponding to the mass of cyclo[APMPPYP] are present. Right: Analysis by LC-FT-ICR of the NMR-characterized C-prenylated cyclo[APMPPYP]; for this compound only a single peak with a mass corresponding to prenylated cyclo[APMPPYP] is observed (b) MS-MS spectra for early (top) and late (bottom) eluting-species observed in initial cyclo[APMPPYP] reaction mixtures and tables of assignments; assignments for early-eluting species consistent with C-prenylation, while those for the late-eluting species consistent with O-prenylation (c) Table of assignments for O- and C-prenylated compounds (d) MS-MS spectrum for NMR-characterized C-prenylated cyclo[APMPPYP]; MS-MS shows no loss of isoprene (e) Table of assignments for C-prenylated reference compound.

(a)


822.4215
823.4257
824.4309
825.4337
826.4305
47.56
822.4241
823.4277
824.4308
825.4292
49.58
45.97
HPLC purify 1st peak;
Re-analyze via LC-FT-ICR

(b)

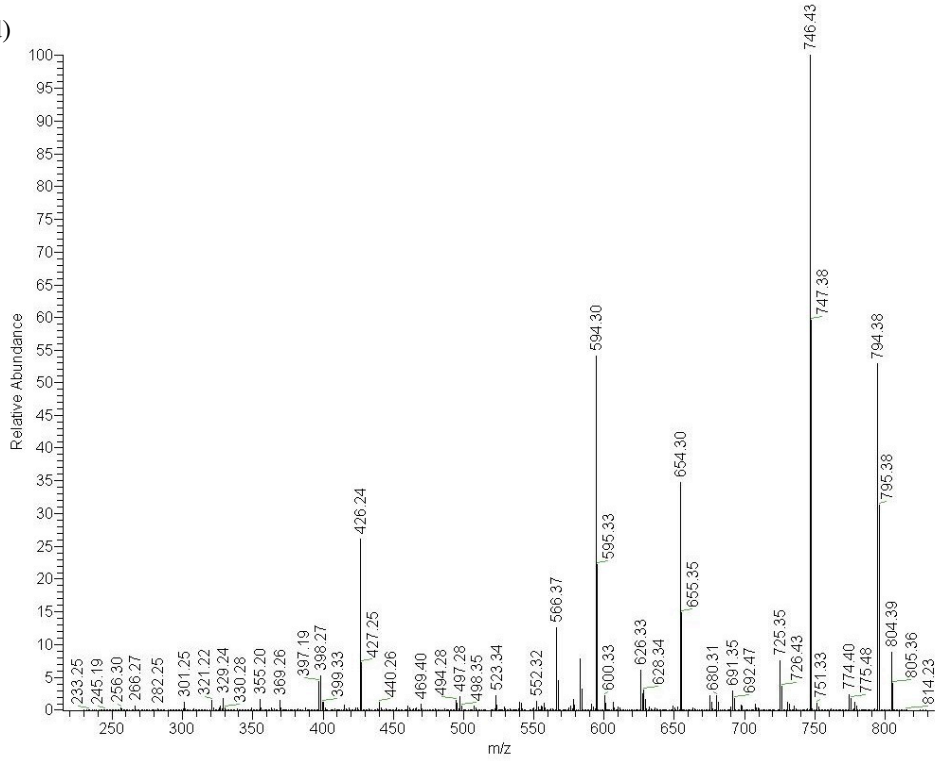

Relative Abundance
m/z
746.39
594.34
794.41
747.42
654.32
426.24
795.42
595.36
566.33
655.35
804.40
626.31
725.32
805.47
812.75
746.44
754.33
594.29
794.40
654.30
426.25
595.37
755.31
795.38
566.36
804.38
805.44
823.79

(c)

APMPPYP 1st peak (C-prenyl)		
b-ions (+1)	**Expected**	**Observed**
AP	169.10	-
APM	300.14	-
APMP	397.19	397.22
APMPP	494.24	494.33
APMPPy	725.37	725.31
y-ions (+1)		
PMPPyP	769.39	769.36
MPPyP	672.34	672.27
PPyP	541.30	541.25
PyP	444.25	444.26
Py	347.19	347.23
y- b- cleavages (+1)		
Py	329.19	329.17
PPy	426.24	426.24
MPPy	557.28	557.25
y- a- cleavages (+1)		
Py	301.19	301.22
Miscellaneous Ions		
APMPPy a-cleavage +1	697.37	697.31

APMPPYP 2nd peak (O-prenyl)		
b-ions (+1)	**Expected**	**Observed**
AP	169.10	-
APM	300.14	-
APMP	397.19	397.22
APMPP	494.24	-
APMPPy	725.37	725.31
y-ions (+1)		
PMPPyP	769.39	769.36
MPPyP	672.34	672.27
PPyP	541.30	541.25
PyP	444.25	444.26
Py	347.19	347.23
y- b- cleavages (+1)		
Py	329.19	329.17
PPy	426.24	426.24
MPPy	557.28	557.23
y- a- cleavages (+1)		
Py	301.19	301.22
Miscellaneous Ions		
All -prenyl	772.37	772.28
APMPPy a-cleavage +1	697.37	697.29
APMPPy -prenyl	657.31	657.26
PPyP -prenyl	473.24	473.23

(d)

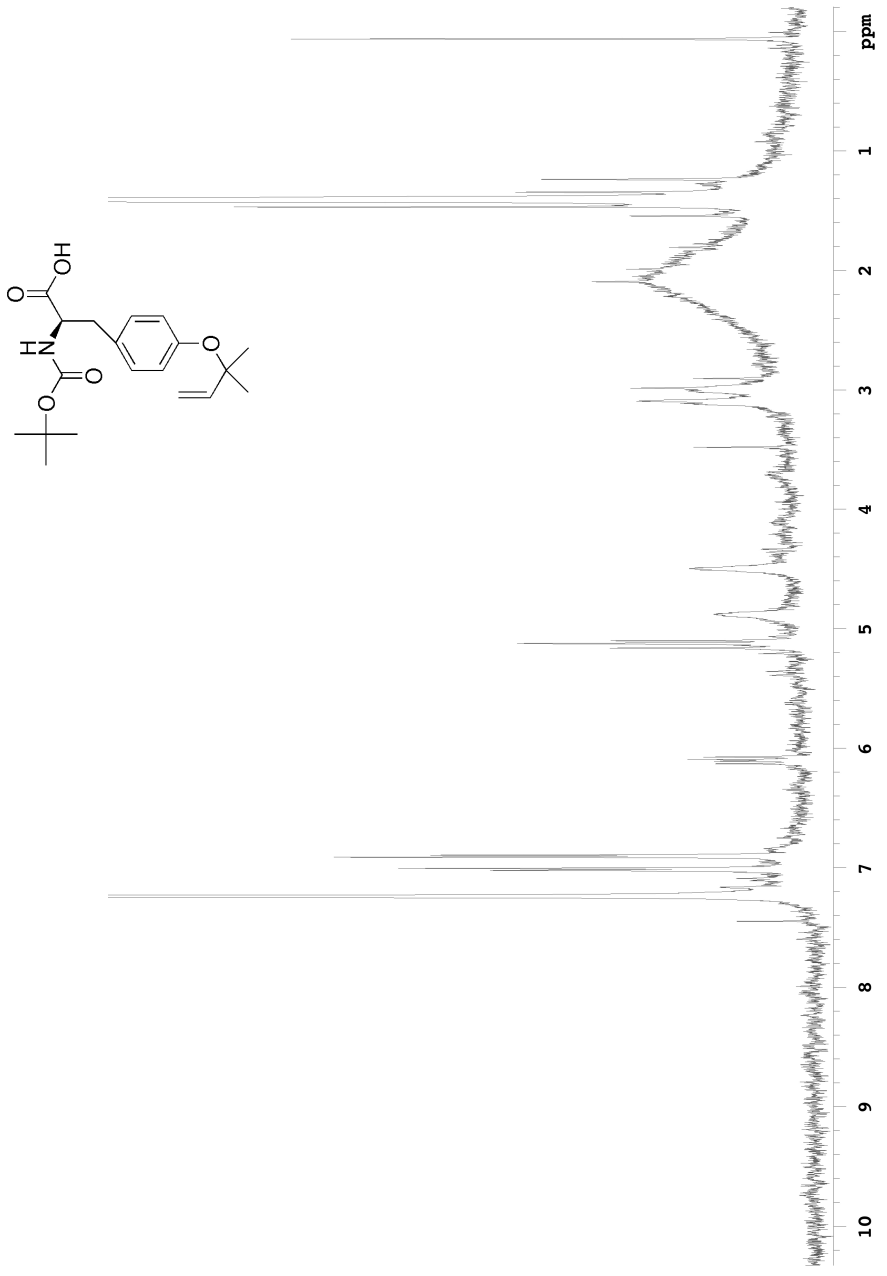

Relative Abundance
m/z
746.43
747.38
594.30
794.38
654.30
795.38
426.24
595.33
655.35
566.37
804.39
725.35
805.36
626.33
814.23
397.19
398.27
427.25
691.35
233.25
245.19
256.30
266.27
282.25
301.25
321.22
329.24
330.28
355.20
369.26
399.33
440.26
469.40
494.28
497.28
498.35
523.34
552.32
600.33
628.34
680.31
692.47
726.43
751.33
774.40
775.48

(e)

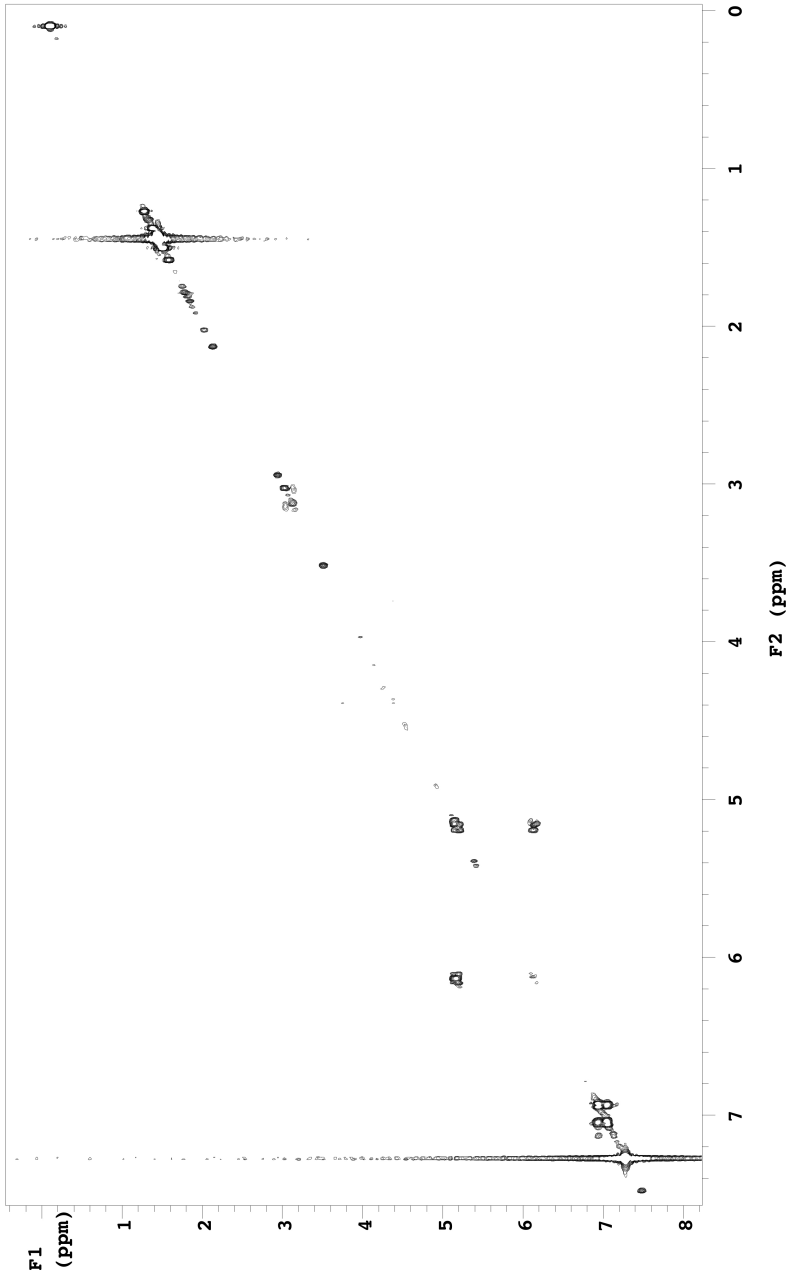
cyclo[APMPPYP] NMR-characterized C-prenyl reference compound					
y- and b- cleavages (+1)	**Expected**	**Observed**	**y- and b- cleavages (+1)**	**Expected**	**Observed**
AP	169.10	-	**yP**	329.19	329.24
APM	300.14	-	**yPA**	400.22	-
APMP	397.19	397.19	**yPAP**	497.28	497.24
APMPP	494.24	494.28	**yPAPM**	628.32	628.34
APMPPy	725.37	725.35	**yPAPMP**	725.37	725.35
PM	229.10	-	**PA**	169.10	-
PMP	326.15	-	**PAP**	266.15	266.28
PMPP	423.21	423.35	**PAPM**	397.19	397.19
PMPPy	654.33	654.30	**PAPMP**	494.24	494.28
PMPPyP	751.39	751.32	**PAPMPP**	591.30	-
MP	229.10	-			
MPP	326.15	-	**Miscellaneous Ions**		
MPPy	557.28	557.29	**-Acylium (C=O)**	794.42	794.38
MPPyP	654.33	654.30	**-Methyl sulfide**	774.42	774.40
MPPyPA	725.37	725.35	**-Met side-chain**	746.39	746.43
PP	195.11	-	**-H2O**	804.41	804.39
PPy	426.24	426.24	**PyPAP y- a- cleavage**	566.33	566.38
PPyP	523.29	523.34	**PAPM y- a- cleavage**	369.20	369.21
PPyPA	594.33	594.30	**PAPM -Met sidechain**	321.16	321.22
PPyPAP	691.38	691.35	**PAPM y- a- cleavage**	301.19	301.25
Py	329.19	329.24			
PyP	426.24	426.24			
PyPA	497.28	497.28			
PyPAP	594.33	594.30			
PyPAPM	725.37	725.35			

Figure S5. Analysis of reverse O-prenylated intermediate and C-prenylated product (a) 1H NMR of boc-D-Tyr reverse O-prenylated intermediate (b) 2D COSY spectrum of boc-D-Tyr reverse O-prenylated intermediate (c) 1H NMR of boc-D-Tyr forward C-prenylated final product (d) circular dichroism spectral overlays of (top) boc-L-Tyr (blue) and boc-D-Tyr (red) and (bottom) C-prenyl boc-L-Tyr (blue) and C-prenyl boc-D-Tyr (red)

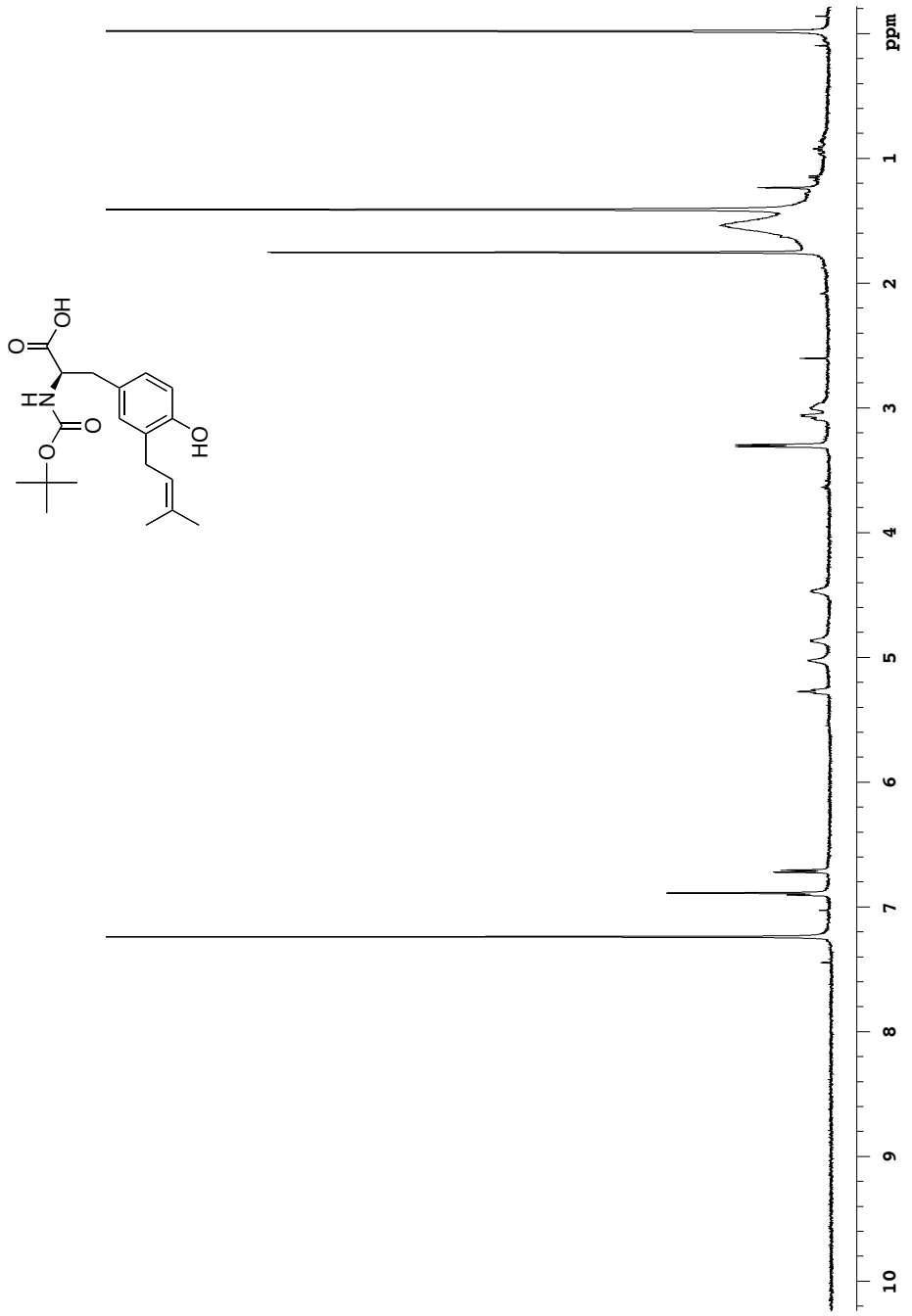
(a)



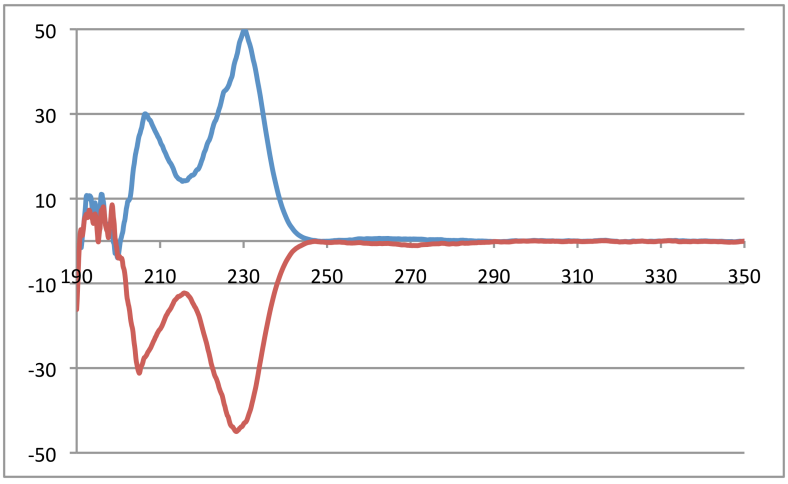
(b)



(c)



(d)


50
30
10
-10
-30
-50
190
210
230
250
270
290
310
330
350

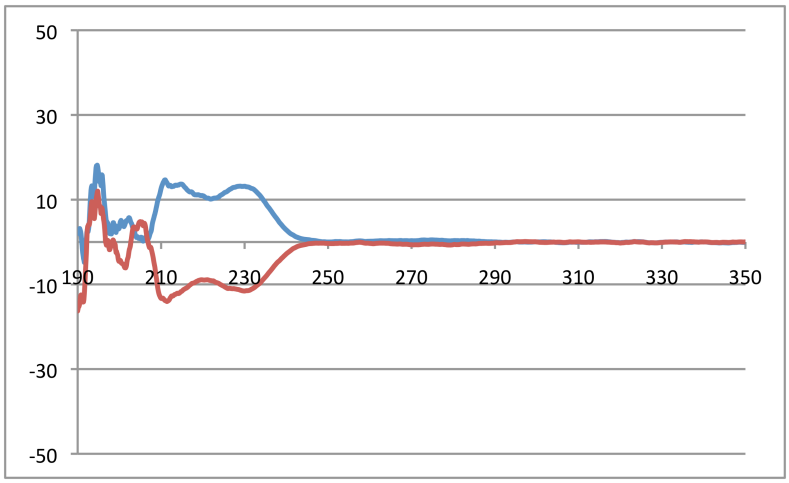

50
30
10
-10
-30
-50
190
210
230
250
270
290
310
330
350

Figure S6. Non-enzymatic rearrangement of reverse O-prenylated tyrosine. Bar graph showing proportions of reverse O-prenylated intermediate and C-prenylated final product after incubation at 37 °C for 0 and 8 h under different conditions. The O-prenylated intermediate is sufficiently unstable that after islolation it was partially rearranged. Proportions of products determined by integration of A220 diode array trace of the relevant peaks, which are well-separated by HPLC.

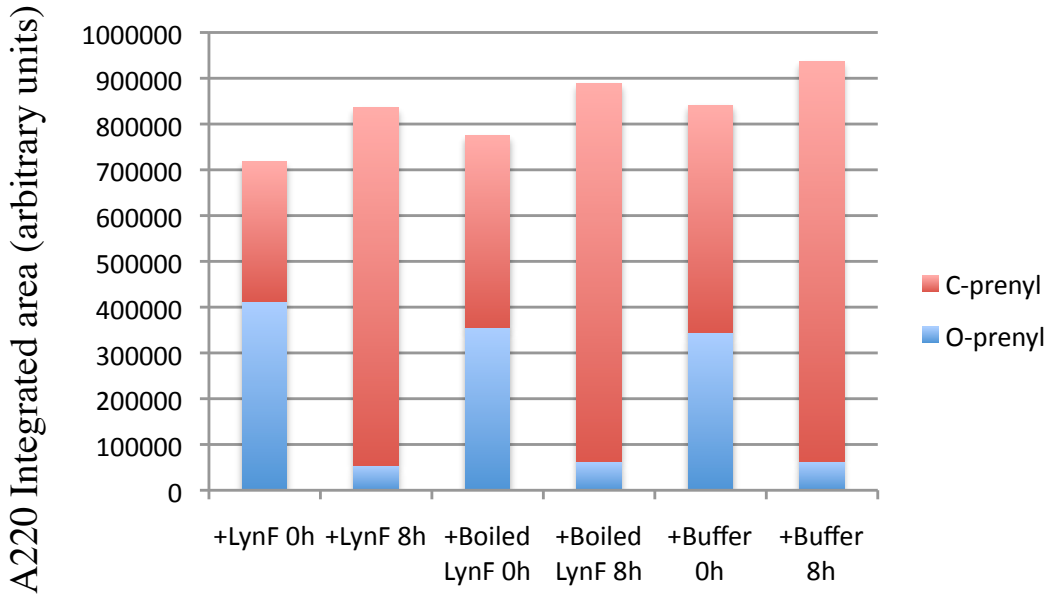
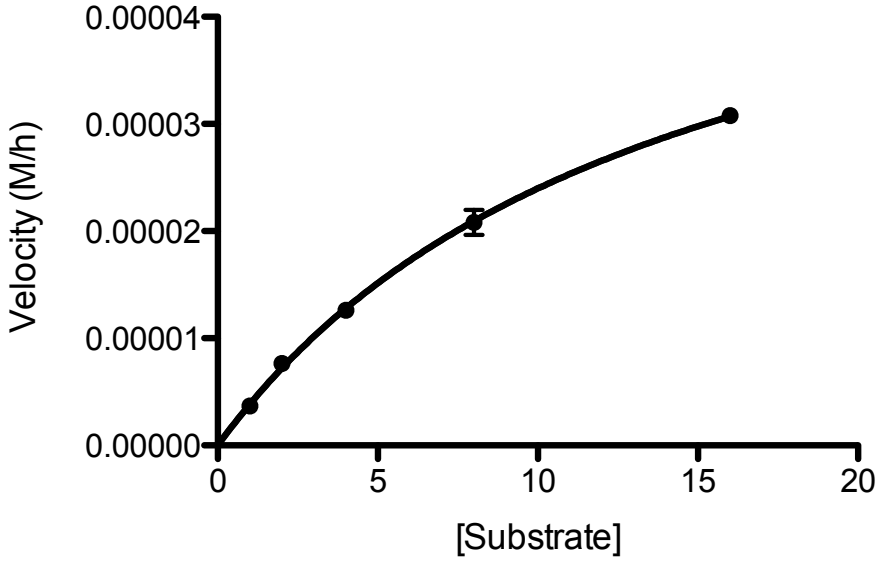


Figure S7. Kinetics of prenylation of cyclo[APMPPYP] and boc-L-tyrosine. Rates were measured by integration of product peaks observed in the monitoring of the HPLC elution 220 nm. Both the C- and O-prenylated products were summed to obtain the total amount of prenylated products at each point. Areas were then plotted on a calibration curve created via injections of known quantities of boc-L-Tyr to derive concentrations (a) Kinetics for reaction of Boc-L-Tyr with LynF (b) Kinetics for reaction of cyclo[APMPPYP] with LynF (c) Controls showing linearity of reaction rate through 6 h.

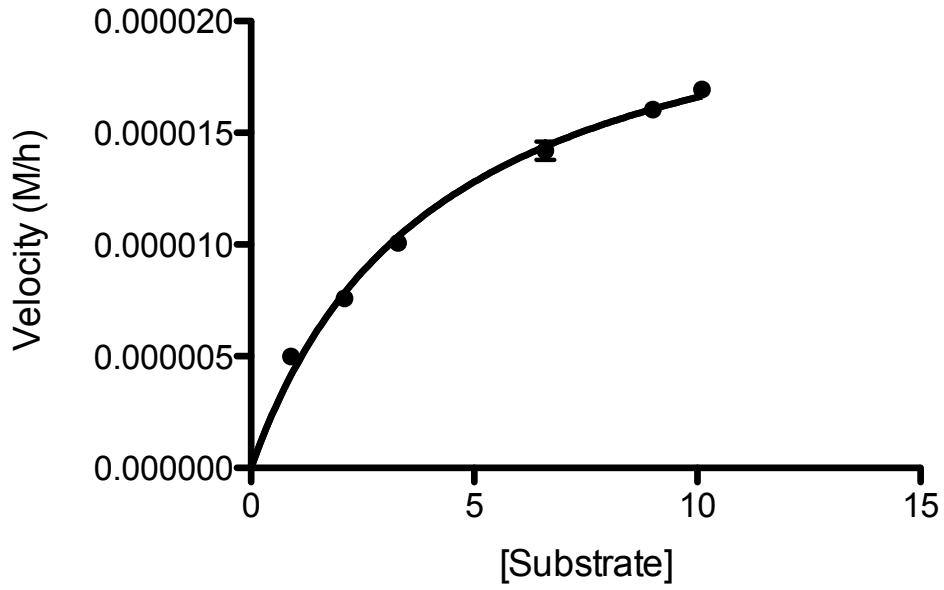
(a)

Best-fit values for Boc-L-Tyr kinetics	
Vmax	5.7E-05 M/h
Km	14.0 mM
kcat	63.3 h-1
Std. Error	
Vmax	3.9E-06 M/h
Km	1.6 mM
95% Confidence Intervals	
Vmax	4.9e-005 to 6.6e-005 M/h
Km	10.49 to 17.59 mM

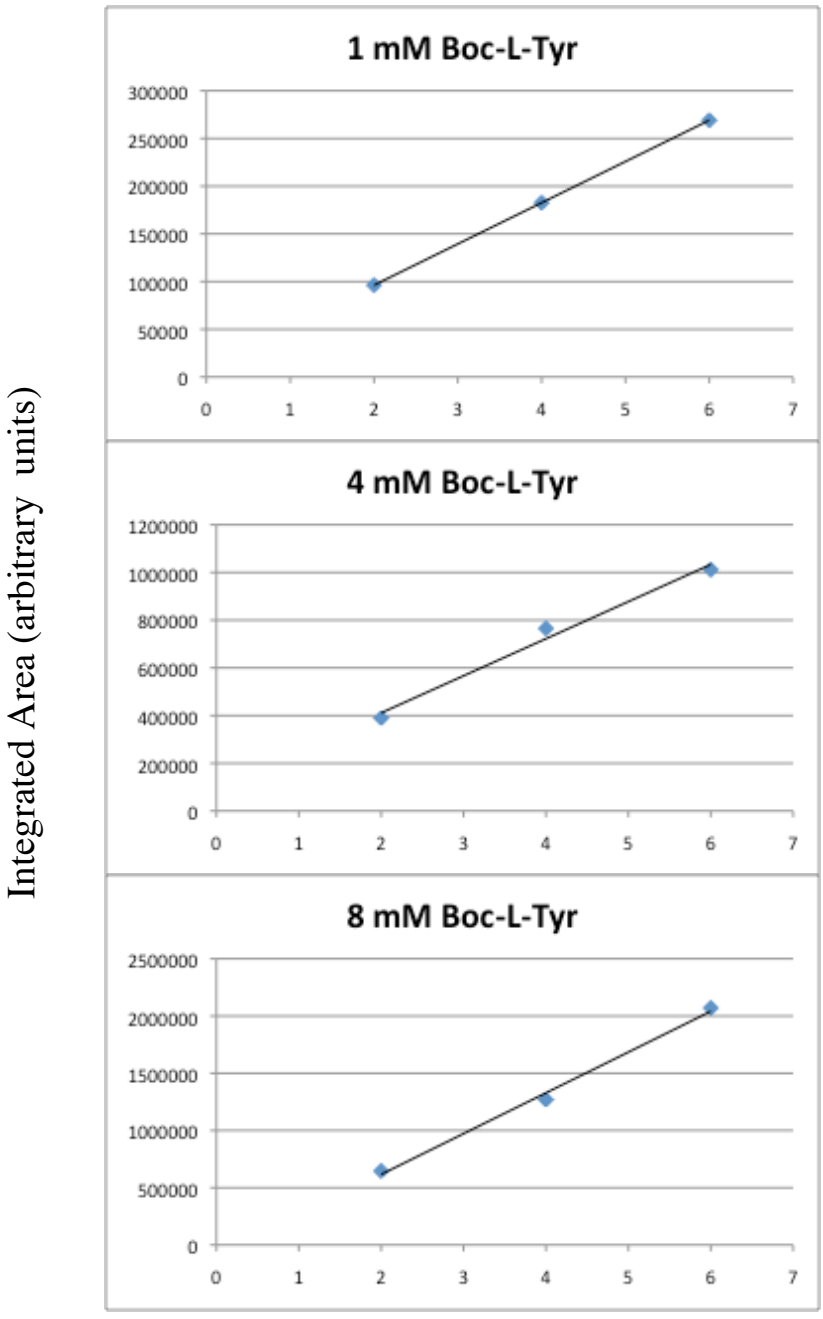


(b)

Best-fit values for cyclo[APMPPYP] kinetics	
Vmax	2.4E-05 M/h
Km	4.2 mM
kcat	13 h-1
Std. Error	
Vmax	8.3E-07 M/h
Km	0.4 mM
95% Confidence Intervals	
Vmax	2.2e-005 to 2.5e-005 M
Km	3.4 to 5.0 mM



(c)


1 mM Boc-L-Tyr
300000
250000
200000
150000
100000
50000
0
0 1 2 3 4 5 6 7
4 mM Boc-L-Tyr
1200000
1000000
800000
600000
400000
200000
0
0 1 2 3 4 5 6 7
8 mM Boc-L-Tyr
2500000
2000000
1500000
1000000
500000
0
0 1 2 3 4 5 6 7
Integrated Area (arbitrary units)
Time (h)

S67

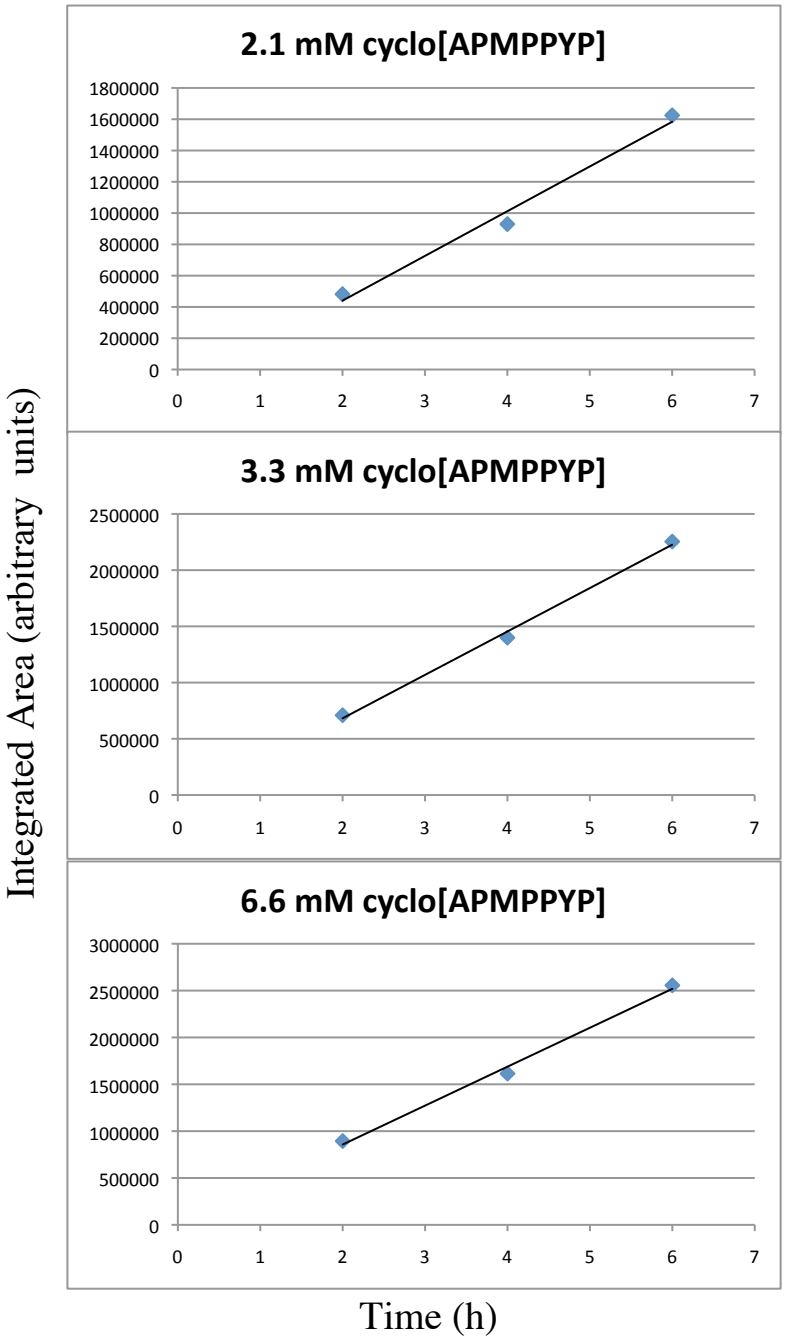
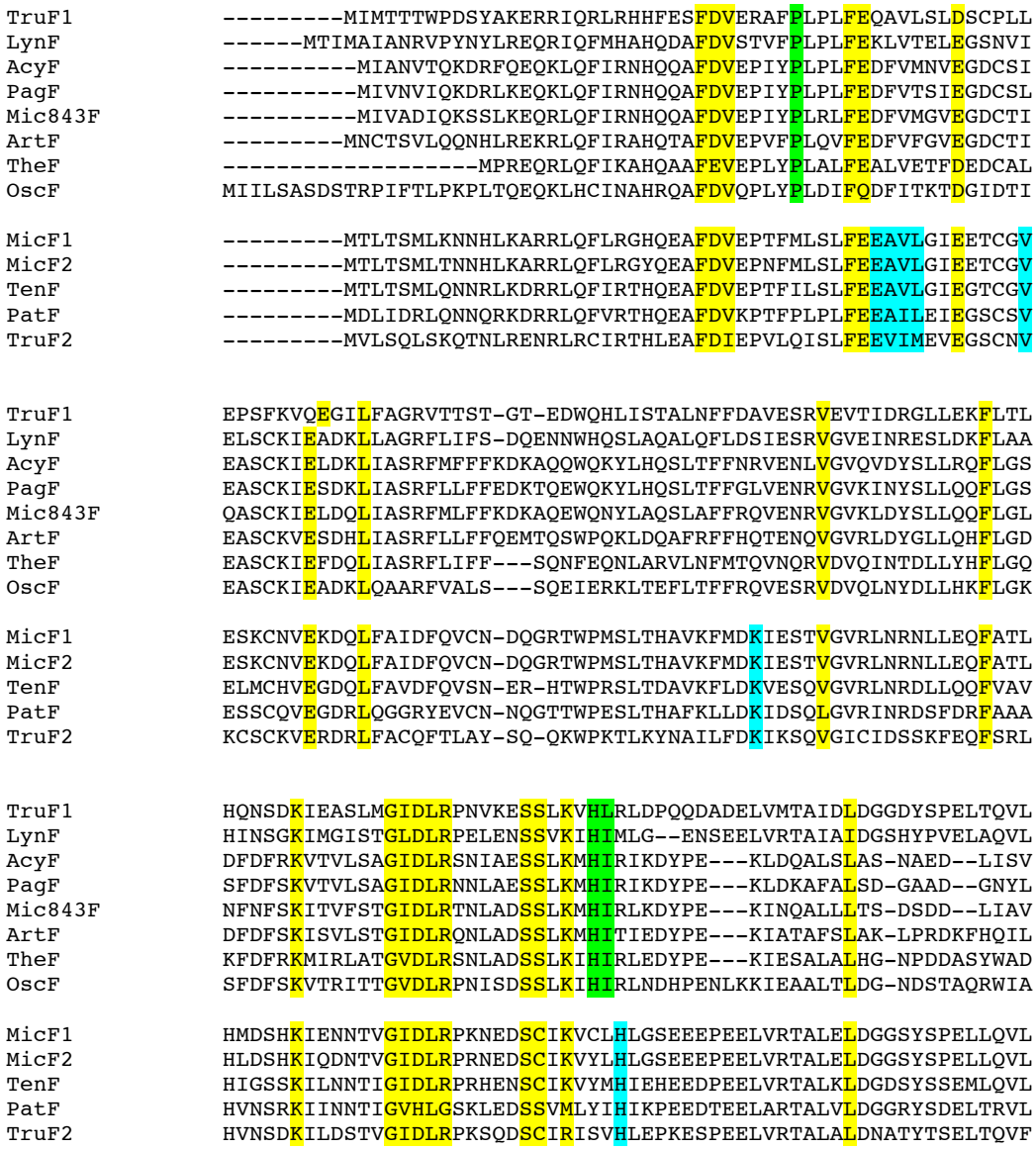

2.1 mM cyclo[APMPPYP]
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
200000
0
0 1 2 3 4 5 6 7
3.3 mM cyclo[APMPPYP]
2500000
2000000
1500000
1000000
500000
0
0 1 2 3 4 5 6 7
6.6 mM cyclo[APMPPYP]
3000000
2500000
2000000
1500000
1000000
500000
0
0 1 2 3 4 5 6 7
Integrated Area (arbitrary units)
Time (h)

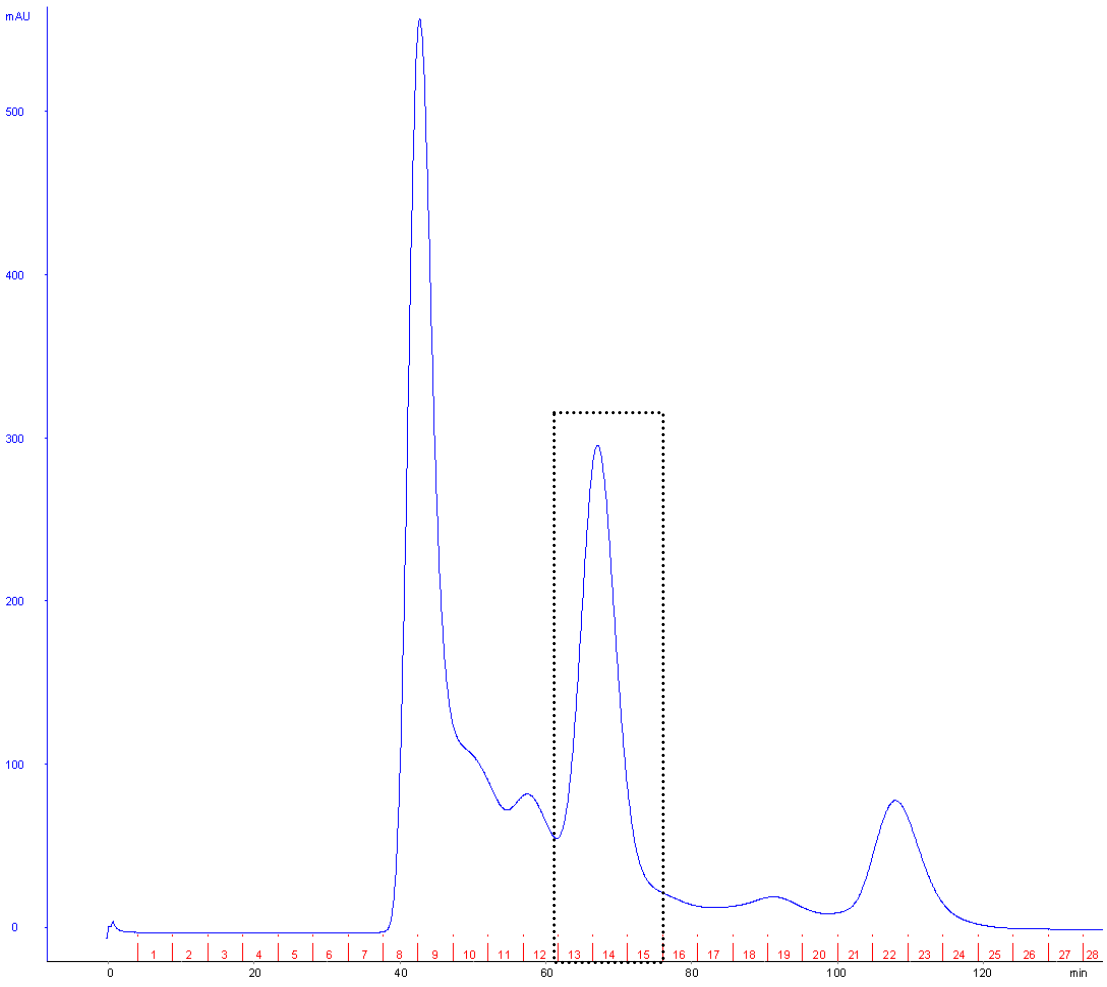
Figure S8. Multiple sequence alignments of LynF relatives. Highlighted in yellow are residues conserved across all known LynF homologues, in green are residues conserved across enzymes predicted to catalyze prenylation, in cyan are residues conserved across predicted non-prenylating enzymes



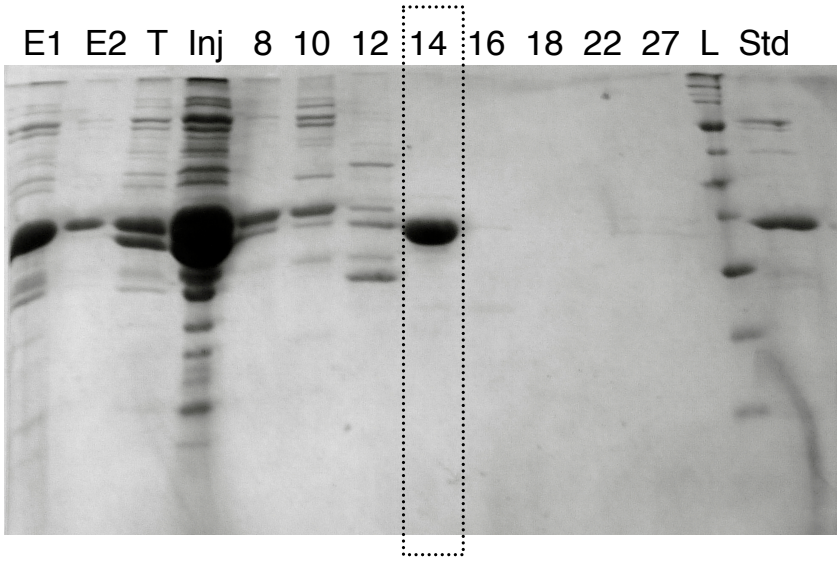
```
TruF1           LKDTFLIGFDFFLDGGSAVEMYTICPGKKPLAMLGKKGAYLKPYVLSNFSHKVTSLLQEV
LynF            LKDTMMIGFDFFLNGHSEVELYISCSRKKDSLPN-NRGESTRYYIRQKFSPKVSSLLDAS
AcyF            RPFLSLVGFDFYFNGRSEIELYPEIQAEDFAKSE------TQNLVWRHFPKFVLDPLEVT
PagF            KDFVNLIGFDFYFNGKSEIEIYAEVQEDDFFKPE------INNLVWQHFPKTALQPLKAS
Mic843F         RDFLSIVGFDFYFDGRSAIKIYPEVAETDFFKPE------TQDKVWRHLPKFVLEPLKAT
ArtF            LSSVSLIGFDFYLDGRSEIELYASLKEEEFNSPH------VQSFLTSNFCASALKPLAAS
TheF            LNAIANIGLDFYLDGRSEIEFYPELSEERFQQPE------MQVLLQQMFPPFVLAPLKAS
OscF            LQTVHLIGFDFYLNGRSEIELYCELTEKQFQQPD------IQSFLQQTFPPFVLEPLKVS

MicF1           LKSTIVIGFNLFLNGYSDIELWALAPGEQYEITNSDRGKYLKHYIQRNFSPKVNSLLKEC
MicF2           LKSTIVIGFNLFLNGYSDLELWATCPGEQYEVPNSDRGKYLKQYVQNNFSQKINDLLRES
TenF            LKSTIIIGFNIFFNGYSDVELLVATVGDKYESHKFNRGKYLKHYIQKNFSLKANYMLRES
PatF            LRDTMVIGFELFFDGRSRVDLGPCAPGK--SGTLKMKGKHLEQYTQKNLSRKVNSIFREG
TruF2           LQDCTAIIFECFFDGRSRIELGAVAPGKKHGFSG-NHGRALTAYAQKYFSPKAVSLSEVS

TruF1           AALTVGFSKEN-PRPVLYFEFETLREVKYNLLFNSLGDKIYDFCLHNQIENFVSIGVTEP
LynF            DFFVGGFSKAN-VEPVLYYAFENIKDIPKYFVFNDLGNRVYDFCRSQDSITMTWIGINER
AcyF            GSFLVGFSKAN-PNPVLYYNLKNKQDLANYFKLNDAAQRVHSFYQNQDILPHMWVGTVQK
PagF            SLFFTGLSKAN-NNPVLYYHLKNRQDLTNYFKLNDTAQRVHSFYQHQDILPYMWVGTAQK
Mic843F         SLFGFGFSKTN-NNPVLYYRLKNRQDLTNYFKLNDTAQRVHSFYQHQDILSSVWLGTAQQ
ArtF            SAFYMGLSIAN-ENPVLYYLLKNKQELQNYFRLNDTGNRVHSLL----------------
TheF            EIFGFGLSKAN-PSAVLYYQLKNKQDLPSYFAINDKAHQVHGYYLHQDTRPYMCVGVAQS
OscF            SVFFTGLSKDN-TEPVLYYCLKDKKDLLSYFPINDTAQRVHAFYQNQPVYSSMWVGVAQG

MicF1           TFFCVSFFHKK--EPVIIFHYEDTKEIPKNFFFNSLGDRIYSFCQGQDCMTYAGVAVTER
MicF2           TFLVVSFSNQK--EPALIFHYEDIKEIPKNFLLNSLGDRIYSFCQGQDCMTYAGVAVTER
TenF            NILLVSFSQAKKVNPLLIFHYEDIKDIRKYFSFNSLGDRSYSFFQSQDCITYAGVSVREL
PatF            YLFGAFFSKTR-VEPILFFYHSIIKDLPKYFTFNSLGDKIYNFCQSQGCITDVAIAVTET
TruF2           DLFGMTISKYK-AEPVLHFGFNNIKDISNYFLFNTLGNRIYSFCQNQDCILLAIIGVNEK

TruF1           DLEKRRLENFRFYYRKAV---------------
LynF            DLDRERLNNFRLYYRRSFG--------------
AcyF            ELEKTRIENVRLYYYKFFN--------------
PagF            ELEKTRIENIRLYYYKSFKMESN----------
Mic843F         ELEKTRIENVRLYYYKLFGKL------------
ArtF            -----------------------------------
TheF            ELAKTRIDQIRLYYHQFFKVQNP----------
OscF            ELQKTRIDNIRLYYSKKNYSK------------

MicF1           ELEKDRLENFSILYNQRDECKPLLHIKKREDFS
MicF2           ELEKDRLENFSILYNQRDECKPLLHIKKREDFS
TenF            ELQKDRLEKFSLFYNKRDKCQHLPLFSTLNE--
PatF            ELEKSRLENFCFYYDQWDECKPSSDYDTERHLH
TruF2           ELYSNRLENFLFDYAKNDESRMMRV--------
```

Figure S9. Sizing column chromatogram for LynF and resulting SDS-PAGE gel (a) Size-exclusion chromatogram showing elution of LynF. Protein elution was detected at 280 nm, and is shown in mAu; along the x-axis is time in minutes and fraction number (the fraction size was 4.5 mL). Purest fractions ultimately used for experiments are highlighted with box (b) SDS-PAGE gel showing analysis of size-exclusion chromatography fractions; fractions 13-15 of size purification used for most experiments described in paper. Key to abbreviations: E1=Ni-NTA elution 1, E2=Ni-NTA elution 2, T=after treatment with thrombin to cleave his-tag, Inj=sample of material injected on sizing column, 8-27=sizing column fraction number, L=ladder, Std=previously prepared LynF sample purified by Ni-NTA chromatography only.

(a)



(b)



Materials and Methods

Synthesis of substrates.

Dimethylallyl pyrophosphate (DMAPP) was synthesized following previously established procedures.[1-3] Briefly, disodium dihydrogen pyrophosphate (15 mmol, 3.3g) was dissolved in 15 mL of a 10% (v/v) solution of ammonium hydroxide. That solution was put over a column containing Dowex AG 50W-X8 column (58 mequiv, acidic form), and eluted with 110 mL of deionized water to give the free acid. The eluent was titrated to pH 7.3 (as measured by a pH meter) with a 40% solution of tetra-*n*-butylammonium hydroxide. The solution was then dried by lyophilization to give tris(tetra-*n*-butylammonium) hydrogen pyrophosphate (**32**). Next, **32** (2.5 mmol, 2.3g) was dissolved in 2.5 mL of acetonitrile. The pyrophosphate solution was placed on ice, with stirring under a blanket of argon. Dimethylallyl bromide (1.2 mmol, 1 M in acetonitrile) was then added dropwise to the pyrophsophate solution. The reaction proceeded for 2 h, during which time it was allowed to come to room temperature. After solvent removal, the residue was dissolved in ion-exchange buffer (1:49 v/v Isopropanol : 25 mM aqueous ammonium bicarbonate) and run on a column containing Dowex AG 50W-X8 (18.8 mequiv, ammonium form). Finally, chromatography on cellulose (2, 3) afforded dimethylallyl pyrophosphate, ammonium salt. ^{1}H NMR (400 MHz, D_2O) δ 5.39 (1H, t, J=7.03 Hz), 4.39 (2H, dd, $J_{H,H}$=6.64 Hz, $J_{H,P}$=6.64 Hz), 1.70 (3H, s), 1.66 (3H, s); ^{31}P NMR (162 MHz, D_2O) δ -5.96 (1P, d, $J_{P,P}$=21.7 Hz), -9.23 (1P, d, $J_{P,P}$=20.6).

Unless otherwise noted, peptide substrates were synthesized at the University of Utah DNA/peptide synthesis core facility using standard Fmoc chemistry. Cyclic peptide **6** was cyclized in solution, leading to small amounts of **9** and **10**. HPLC purification effected purification of **6** from **9** and **10**, which co-eluted and were used as a mixture in prenyltransferase experiments. Substrates **11** and **12** were was synthesized as a linear precursor (KKPYILPAYDGE) and then enzymatically cyclized as reported elsewhere.[4] Synthesis and characterization of **26** and **27** has been previously reported.[5,6]

Boc-protection of 4-iodo-L-phenylalnine and 4-methoxy-L-phenylalanine was performed according to previously established procedures.[7] Briefly, to a stirred solution of amino acid (0.3 M in 50/50 THF/H_2O) was added di-tert-butyl dicarbonate (1.1 equiv), and triethylamine (1 equiv). The reactions were stirred at room temperature until the starting material was consumed, as determined by TLC. Reaction mixtures were then dried by rotary evaporation to remove THF, and then lyophilized. The residue was then dissolved in aqueous sodium hydroxide (1 M, 2 equiv) and extracted twice with CH_2Cl_2; the solutions were acidified and extracted three times with CH_2Cl_2. The acid extracts were then combined and dried by rotary evaporation. Prior to use in enzyme reactions, small amounts of each boc-protected amino acid were subjected to additional purification by reversed-phase HPLC. To wit, the amino acid derivatives were injected on a C18 onyx monolithic semiprep column (Phenomenex) on a linear gradient from 50% buffer A (H_2O, 0.05% TFA) to 100% buffer B (acetonitrile) over 60 minutes.

Boc-4-iodo-L-Phe: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.62 (2H, br), 6.93 (2H, br), 4.90 (1H, s), 4.56 (1H, br), 3.13 (1H, br), 3.03 (1H, br), 1.42 (9H, s); ^{13}C NMR (100 MHz, DMSO d_6) δ 173.4, 155.4, 137.9, 136.8, 131.6, 92.2, 78.1, 54.9, 35.9, 28.1. **Boc-4-methoxy-L-Phe**: ^{1}H NMR (400 MHz, DMSO d_6) δ 7.14 (2H, br), 6.83 (2H, br), 4.00 (1H, br), 3.70 (3H, br), 2.91 (1H, br), 2.74 (1H, ap t, J=10.8 Hz), 1.31 (9H, s). ^{13}C NMR (100 MHz, DMSO d_6) δ 173.7, 157.9, 155.5, 130.1, 129.9, 113.6, 78.0, 55.5, 55.0, 35.6, 28.2.

Boc-L-tyrosine, sodium hydrogen pyrophosphate, dimethylallyl bromide, tetrabutylammonium hydroxide, and dopamine HCl were purchased from Sigma. N-acetyl-L-tyrosine and phenol were purchased from Fisher Scientific. All other Tyr and Phe derivatives were purchased from ChemImpex.

Genes and Cloning

A codon-optimized version of *lynF* was synthesized and cloned into pET28 in frame with the N-terminal his-tag sequence using NdeI and EcoRI (Genscript). The vector pRSF-truE1-DlacI was described previously.[8] Oligonucleotides were designed to swap the region of truE1 encoding patellin 2 (TVPTLC) for a region encoding a cassette found in the *lyn* pathway (VCMPCYP). The resulting construct (pRSF-LynE-Dlac) was confirmed using DNA sequencing. This construct was then cloned into pET28b using NdeI and BamHI.

Protein Expression and Purification

The plasmid encoding *lynF* was transformed into BL21(DE3) cells via electroporation and grown at 30 °C. Colonies were picked the following day and grown overnight in 2 x 45 mL starter cultures (LB media, 50 μg/mL kanamycin) at room temperature. The starter cultures were then used to inoculate 8 x 1 L cultures (LB, 50 μg/mL kanamycin) in 2.8 L Fernbach flasks and grown at 30 °C until an A600 of 0.4 was reached. Upon reaching the desired optical density, the temperature was lowered to 15 °C and IPTG (0.1 mM final concentration) was added, and expression was allowed to proceed overnight. Cells were then pelleted by centrifugation, and frozen at -80 °C until purification. Cell lysis was performed according to previously established procedures. Briefly, cells were resuspended in cold lysis buffer (200 mM NaCl, 50 mM Tris pH 7.5, 100 μM EDTA, 5 mL per gram of cell paste) to which lysozyme (600 μg/mL), PMSF (1 mM), and imidazole (10 mM) were added. The suspension was then incubated on ice with stirring for 1 h, after which time $MgCl_2$ (10 mM) and DNase I (20 μg/mL) were added and incubation on ice was continued for a further 0.5 h. After lysis was complete, the suspension was centrifuged at 14,000 RPM in a JA-20 rotor for 1 h at 4 °C. Cleared lysates were then passed through a 0.45 μm syringe filter, and then ATP (2 mM) was added. Lysate was then applied to a gravity column containing a 5 mL bed of nickel-NTA resin (Qiagen). The lysate was allowed to equilibrate on resin for 15 minutes and then allowed to flow through. The column was then washed with 20 column volumes of wash buffer (500 mM NaCl, 25 mM imidazole, 2 mM ATP pH 8.0), and then eluted with 2 x 15 mL of elution buffer (750 mM NaCl, 250 mM imidazole pH 8.0). In order to cleave the his-tag, thrombin was added to the eluents (33 nM), which were then placed in 8,000 MWCO dialysis tubing and stirred in dialysis buffer (0.5 M NaCl, 25 mM HEPES pH 8.0) overnight at 4 °C. Following dialysis, LynF-containing fractions were concentrated and run on an S-75 sizing column, for which an isocratic buffer (0.5 M NaCl, 25 mM HEPES pH 8.0) was employed. Fractions were tested for activity, the portion eluting between 65 and 77 mL was significantly more active as determined by MALDI-MS analysis of reactions containing **6**. The active fractions were then dialyzed into storage buffer (0.5 M NaCl, 25 mM HEPES pH 8.0, 10% glycerol), flash frozen in liquid nitrogen, and frozen at -80 °C until use.

Expression of *truLy1* was performed in a similar fashion, with the main difference being that rather than attempting to isolate soluble protein, *truLy1* was strongly overexpressed with the intent of driving the protein into inclusion bodies, after which time purification under denaturing conditions was performed. To wit, 8 x 1L cultures (sLB: 10 g/L soytone, 10 g/L NaCl, 5 g/L yeast extract, 50 μg/mL kanamycin) containing the *truLy1* plasmid were grown at 37 °C and upon reaching an A600 of 0.4, induced with 1 mM IPTG and allowed to grow overnight. Cell lysis and denaturing Ni-NTA purification was effected according to the manufacturer's protocol (Qiagen). The concentration of TruLy1 was assessed by amino acid analysis.

Enzyme Assays

Enzyme reactions typically contained enzyme (3.8 μM), variable substrate concentration (100 μM for most substrates; higher concentrations, i.e. 1 mM, were occasionally employed with boc-protected amino acid derivatives. Exceptions include substrates **11** and **12**, which were used at

20 μM final concentration as well as substrates **9**, and **10**, which were used at 70 and 30 μM, respectively), and several additives (1 M NaCl, 40 mM glycylglycine pH 9.0, 12 mM $MgCl_2$, 3 mM tris(2-carboxyethyl) phosphine (TCEP), and 1 mM DMAPP). Reactions were incubated at 37 °C for 24 h in a DNA Engine Peltier thermocycler (Bio-Rad). Enzyme reactions with full-length precursor peptide contained TruLy1 (28 μM), ATP (0.8 mM), with or without heterocyclase enzyme TruD (90 nM), and additives as above. Controls were run to ensure that LynF was active in the presence of TruD and TruLy1, and vice versa.

Reactions prepared for kinetic analysis of boc-L-Tyr were prepared in triplicate, and contained variable substrate concentrations (1, 2, 4, 8, and 16 mM), constant enzyme (0.9 μM), DMSO (4% v/v) for solubility, and additives as described above. The reactions were incubated for 3 h at 37 °C as above, halted by addition of guanidinium HCl (1 M), and then frozen at -80 °C until HPLC analysis. A sample of each reaction (15 μL) was injected onto a 214MS C4 5μ column (Grace-Vydac) and run on a linear gradient from 99% buffer A (H_2O, 0.5% TFA) / 1% buffer B (AcN) to 0% A / 100% B over 45 minutes. Peaks corresponding to O-prenyl and C-prenyl products were integrated, and summed to arrive at the total integrated area corresponding to prenylated product. The integrated areas were plotted on a calibration curve to determine what molar amounts of product they represented. The data were then analyzed using GraphPad Prism to derive kinetic curves and parameters.

Kinetic analysis of cyclo[APMPPYP] was performed in a similar fashion, with variable substrate concentrations (0.9, 2.1, 3.3, 6.6, 9, and 10.1 mM) and constant enzyme (1.8 μM). Reactions were incubated as above and quenched in an identical manner. HPLC conditions differed only in that the run began by holding the solvent mixture at 99% buffer A / 1% buffer B for the first 5 minutes, then stepped to 80% buffer A / 20% buffer B at which time a linear gradient to 60% buffer A / 40% buffer B over 34 minutes was begun. Additionally, controls were run with both boc-L-Tyr and cyclo[APMPPYP] to ensure that the rate of product formation was linear over the first several hours. To wit, reactions containing boc-L-Tyr (1, 4, and 8 mM), constant enzyme (0.9 μM), DMSO (4% v/v) for solubility, and additives as described above were incubated at 37 °C. Aliquots were removed at three time points (2, 4, and 6 h), and analyzed as above. A similar set of controls was performed containing cyclo[APMPPYP] (2.1, 3.3, and 6.6 mM), constant enzyme (1.8 μM), DMSO (4% v/v), and additives as described above. Data presented in figure 5 were generated via analysis of a time-course of the LynF (5.4 μM) reaction with boc-L-Tyr (1 mM) containing standard additives. Reaction aliquots (17.5 μL) were removed at 1, 2, 4, 6, and 24 h and immediately mixed with guanidinium HCl (3.5 μL, 6 M) and then placed at -80 °C until HPLC analysis.

Experiments assaying prenylation of TruLy1 were performed as follows: LynF (1.5 μM) or buffer control was incubated with either TruLy1 (14 μM), boc-L-Tyr (14 μM), cyclo[APMPPYP] (14 μM), without substrate, additives as above with the exception of DMAPP, and ^{3}H-labeled DMAPP (0.1 μM) (American Radiolabeled Chemicals), which had previously been dried under a stream of argon, and resuspended in ultrapure water. Reactions were covered with a mineral oil overlay (30 μL) and placed in a heat-block at 37 °C for 24h. Once complete, reaction aliquots (20 μL) were then quenched by addition of acid (2 M HCl in 80% EtOH, 4 μL), incubated a further 30 minutes to hydrolyze any remaining DMAPP. The resulting mixture was then added to a 7 mL scintillation vial along with water (150 μL), and placed in a sand-bath at 110 °C for 1.5 h. Resolubilization was then effected via addition of Soluene 350 (1 mL) (Perkin-Elmer), and heating in a sand bath at 50 °C for 1 h. Scintillation fluid (up to 7 mL) (Hionic-Fluor, Perkin-Elmer) was then added. Samples were placed in the dark for 30 minutes and then read by scintillation counting (5 minutes per sample). Radioassay samples were also analyzed by autoradiography, in which case a reaction aliquot (15 μL) was mixed with 6X SDS-PAGE buffer (0.6 M DTT, 0.35 M tris pH 6.8, 30% (v/v) glycerol, 10% (w/v) SDS, 0.012% (w/v) bromophenol blue, 3 μL), boiled for 5 minutes, and run on an 18%

polyacrylamide gel. After staining with Coomassie R250, gels were destained, and enhanced for autoradiography using En3hance (Perkin-Elmer) according to the manufacturer's instructions. Film (BioMax XAR, Kodak) was then exposed to the enhanced gel at -80 °C for at least 24 h and then developed. Lastly, possible prenylation of TruLy1 was assessed via ESI-MS intact analysis, as well as by proteolytic digest followed by LC-FT-ICR. In those cases, LynF (1.5 μM) was incubated with TruLy1 (14 μM) or boc-L-Tyr (640 μM), additives as above, and ATP (800 μM). Reactions were run at 37 °C for 24h; samples were taken at 0, 6, 12, and 24h time points, and were then analyzed by ESI-MS. Control reactions were run containing TruD (90 nM), or lacking ATP and TruD, and also analyzed by ESI-MS after 24h incubation. Boc-L-Tyr containing reactions were analyzed by HPLC as positive controls. Samples analyzed by LC-FT-ICR were digested either with previously characterized cyanobactin protease PatA (2 μM) including added calcium (10 mM),[6] or with trypsin and chymotrypsin.

Isolation of O-prenylated intermediate was performed as follows: a crude fraction of LynF (6.3 μM) was incubated with boc-D-Tyr (3.5 mM, 8.8 μmol), DMSO (1% v/v) for solubility, and additives as above. The reaction was allowed to proceed for 48 h at 37 °C. The reaction was then acidified (pH 4) with 1 M HCl and then extracted 5X with methylene chloride. The combined organic phases were then dried, and purified by HPLC. Purification was performed using a Onyx monolithic C18 semi-preparative column on a linear gradient from 99% buffer A (H_2O, 0.05% TFA) / 1% buffer B (AcN) to 100% buffer B over 60 minutes. The dominant C-prenylated product was purified along with the O-prenyl product.

To assess whether the conversion of reverse O-prenylated Tyr to forward C-prenylated Tyr was accelerated by LynF, the O-prenylated intermediate was purified as above, and then incubated at 37 °C with additives as above, excepting DMAPP, and either LynF (2.3 μM), boiled LynF (2.3 μM), or buffer only. Time points were taken at 0 and 8 h, and then frozen at -80 °C until analysis.

Phylogenetic Tree Construction

The amino acid sequences of LynF homologues from the functionally characterized cyanobactin pathways were aligned using CLUSTALX. Maximum likelihood analysis with molecular clock PROMLK (PHYLIP) using the bootstrap test method (1000 replicates) was performed to assess the phylogenetic relationship between the different homologues. The same tree branches were also supported using other phylogenetic experiments such as Maximum Parsimony (MEGA 4.0) using 1000 bootstrap replicates.

Preparative Enzymatic Synthesis

Preparative synthesis of prenylated boc-L-Tyr was performed as follows: a crude fraction of LynF (6.3 μM) was incubated with boc-L-Tyr (3.5 mM, 17.8 μmol) and additives as above. The reaction was incubated at 37 °C for 48 h and extracted as described above for boc-D-Tyr. Purification was accomplished using a Vydac C4 214TP1010 semi preparative column on a linear gradient from 99% buffer A (H_2O, 0.05% TFA) / 1% buffer B (AcN) to 100% buffer B over 45 minutes.

Preparative synthesis of prenylated boc-D-Tyr was performed as described in the previous section. The purified O- and C-prenylated products were subjected to NMR analysis: **Boc-4-O-(2-methylbut-3-en-2-yl)-D-tyrosine**: ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.01 (2H, d, *J*=8.5 Hz), 6.90 (2H, d, *J*=8.5 Hz), 6.10 (1H, dd, *J*=17.3, 10.5), 5.14 (1H, ap d, *J*=17.8), 5.11 (1H, ap d, *J*=11.5), 4.88 (1H, s), 4.49 (1H, br), 3.10 (1H, br), 2.98 (1H, br), 1.42 (6H, s), 1.40 (9H, s). **Boc-3-(3-methylbut-2-en-1-yl)-D-tyrosine**: ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.90 (1H, d, *J*=8.0 Hz), 6.89 (1H, s), 6.71 (1H, d, *J*=8.0 Hz), 5.27 (1H, t, *J*=6.6 Hz), 4.86 (1H, s), 4.47 (1H, br), 3.30 (2H, d, *J*=7.0 Hz), 3.07 (1H, m), 3.00 (1H, br), 1.75 (6H, ap s), 1.41 (9H, s)

Preparative synthesis of prenylated cyclo[APMPPYP] was performed by incubation of sizing-column purified LynF (5 μM) with cyclo[APMPPYP] (0.4 mM) and additives as above. The reaction was incubated at 37 °C for 88 h, and then frozen at -20 °C. Purification was accomplished using a Vydac C4 214TP1010 semi preparative column on a linear gradient from 99% buffer A (H_2O, 0.05% TFA) / 1% buffer B (AcN) to 100% buffer B over 45 minutes.

Preparative synthesis of prenylated boc-D-Tyr was performed as described in the previous section. The purified O- and C-prenylated products were subjected to NMR analysis: **Boc-4-O-(2-methylbut-3-en-2-yl)-D-tyrosine**: ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.01 (2H, d, *J*=8.5 Hz), 6.90 (2H, d, *J*=8.5 Hz), 6.10 (1H, dd, *J*=17.3, 10.5), 5.14 (1H, ap d, *J*=17.8), 5.11 (1H, ap d, *J*=11.5), 4.88 (1H, s), 4.49 (1H, br), 3.10 (1H, br), 2.98 (1H, br), 1.42 (6H, s), 1.40 (9H, s). **Boc-3-(3-methylbut-2-en-1-yl)-D-tyrosine**: ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.90 (1H, d, *J*=8.0 Hz), 6.89 (1H, s), 6.71 (1H, d, *J*=8.0 Hz), 5.27 (1H, t, *J*=6.6 Hz), 4.86 (1H, s), 4.47 (1H, br), 3.30 (2H, d, *J*=7.0 Hz), 3.07 (1H, m), 3.00 (1H, br), 1.75 (6H, ap s), 1.41 (9H, s)

Preparative synthesis of prenylated cyclo[APMPPYP] was performed by incubation of sizing-column purified LynF (5 μM) with cyclo[APMPPYP] (0.4 mM) and additives as above. The reaction was incubated at 37 °C for 88 h, and then frozen at -20 °C. Purification was accomplished using a Vydac C4 214TP1010 semi preparative column; the run was begun by holding the solvent mixture at 99% buffer A / 1% buffer B for the first 5 minutes, then stepped to 80% buffer A / 20% buffer B at which time a linear gradient to 60% buffer A / 40% buffer B over 55 minutes was begun.

General Methods

ESI-MS, and FT-ICR analyses were performed at the University of Utah Mass Spectrometry and Proteomics core facility. MALDI-MS analyses were performed on a Micromass MALDI micro MX instrument (Waters). HPLC separations were performed on a LaChrom Elite system (Hitachi). NMR spectra were collected either on 400 or 500 MHz spectrometers (Varian). CD spectra were collected on a Jasco J-815 spectrometer, and data were plotted in Excel.

Table S1. Substrates used in this study.

#	Substrate	Result	Analytical Method(s)
3	TruLy1 precursor peptide	NR	ESI-MS, LC-FT-ICR, radioassay, autoradiography
4	TruD-treated TruLy1	NR	ESI-MS, LC-FT-ICR, radioassay, autoradiography
5	APMPPYPSYDDAE	NR	MALDI-MS, HPLC
6	cyclo[APMPPYP]	48%	MALDI-MS, HPLC, LC-FT-ICR, NMR, radioassay
7	APMPPYP	12%	MALDI-MS, HPLC, LC-FT-ICR
8	N-acetyl APMPPYP	10%	HPLC
9	cyclo[APMPPPAPMPPYP]	47%	MALDI-MS, LC-FT-ICR
10	cyclo[APMPPYPAPMPPYP]	43%	MALDI-MS, LC-FT-ICR
12	cyclo[KKPYILP]	37%	MALDI-MS, LC-FT-ICR
12	cyclo[KPYILP]	94%	MALDI-MS, LC-FT-ICR
13	KPYILP	1%	MALDI-MS, LC-FT-ICR
14	Boc-L-Tyr	71%	HPLC, FT-ICR, NMR, CD, radioassay
15	Boc-D-Tyr	66%	HPLC, NMR, CD
16	Boc-4-cyano-L-Phe	NR	HPLC
17	Boc-O-allyl-L-Tyr	NR	HPLC
18	Boc-4-iodo-L-Phe	NR	HPLC
19	Boc-4-methoxy-L-Phe	NR	HPLC
20	L-Phe	NR	HPLC
21	N-acetyl-L-Tyr	3%	HPLC, ESI-MS
22	L-Tyr	NR	HPLC
23	Dopamine	NR	HPLC
24	Phenol	NR	HPLC
25	L-Trp	NR	HPLC
26	cyclo[QGGRGDWP]	NR	MALDI-MS
27	QGGRGDWPAYDGE	NR	MALDI-MS

```
TruLy1:
MGSSHHHHHHSSGLVPRGSHMNKKNILPQLGQPVIRLTAGQLSSQLAELSE
EALGGVDASTLPVPTLCSYDGVDASVCMPCYPSYDD

TruLy1 with 3 heterocycles:
MGSSHHHHHHSSGLVPRGSHMNKKNILPQLGQPVIRLTAGQLSSQLAELSE
EALGGVDASTLPVPTL(Tz)SYDGVDASV(Tz)MP(Tz)YPSYDD
Tz=thiazoline
```

Table S2. Summary of FT-ICR data

Substrate	Product[a]	Expected *m/z*	Observed *m/z*	Δppm
6	cyclo[APMPP**y**P]	822.4219 ($M+H^+$)	822.4216	-0.36
7	APMPP**y**P	840.4324 ($M+H^+$)	840.4328	-0.47
9	cyclo[APMPPPAPMPP**y**P]	706.8589 ($M+2H^{2+}$)	706.8593	+0.56
10	cyclo[APMPP**y**PAPMPPYP]	788.3906 ($M+2H^{2+}$)	788.3912	+0.76
10	cyclo[APMPP**y**PAPMPP**y**P]	822.4219 ($M+2H^{2+}$)	822.4227	+0.97
11	cyclo[KKP**y**ILP]	454.8020 ($M+2H^{2+}$)	454.8019	-0.22
12	cyclo[KP**y**ILP]	780.5018 ($M+H^+$)	780.5013	-0.64
13	KP**y**ILP	798.5124 ($M+H^+$)	798.5137	+1.6
14	O- and C-prenyl boc-L-Tyr	350.1962 ($M+H^+$)	350.1966	+1.1

[a] '**Y**' denotes tyrosine, 'y' denotes prenylated tyrosine

(1) Davisson, V. J.; Woodside, A. B.; Neal, T. R.; Stremler, K. E.; Muehlbacher, M.; Poulter, C. D. *J. Org. Chem.* **1986**, *51*, 4768-4779.
(2) Woodside, A. B.; Huang, Z.; Poulter, C. D. *Organic Syntheses, Coll.* **1988**, *66*, 211.
(3) Woodside, A. B.; Huang, Z.; Poulter, C. D. *Organic Syntheses, Coll.* **1993**, *8*, 616.
(4) McIntosh, J. A.; Robertson, C. R.; Vinayak, A.; Satish, N. K.; Bulaj, G. W.; Schmidt, E. W. *J. Am. Chem. Soc.* **2010**, *132*, 15499-15501.
(5) Donia, M. S.; Hathaway, B. J.; Sudek, S.; Haygood, M. G.; Rosovitz, M. J.; Ravel, J.; Schmidt, E. W. *Nat. Chem. Biol.* **2006**, *2*, 729-735.
(6) Lee, J.; McIntosh, J. A.; Hathaway, B. J.; Schmidt, E. W. *J. Am. Chem. Soc.* **2009**, *131*, 2122-4.
(7) Stankova, I. G.; Videnov, G. I.; Golovinsky, E. V.; Jung, G. *J. Peptide Sci.* **1999**, *5*, 392-398.
(8) Donia, M. S.; Schmidt, E. W. *Chem. Biol.* **2011**, *18*, 508-19.

CHAPTER 8

CONCLUSIONS

1.1 Conclusions

In the prior chapters, I have shown the extensive characterization of a number of enzymes important in the biosynthesis of a group of natural products, the cyanobactins. A question that naturally suggests itself prior to conducting such an investigation, but which due to convention is most likely to be posed at the conclusion of such an inquiry, is the simplest: why?

Indeed, why study natural products biosynthesis? Further, why study the particular natural products that were the focus of this work? Beyond the more fundamental questions one might ask regarding the benefits to the accruing of human knowledge, here I shall attempt to answer the question about the topics specifically treated in this work. Further, since the theoretical benefits of such an inquiry are, by nature, more difficult to specify, only the practical benefits of such an inquiry will be addressed.

To answer the general question first, natural products make up a large fraction of clinically-used drugs. Further, natural products are still actively researched as a source of new molecular entities of interest for drug development. Still, what benefit can the understanding of their biosynthetic origins provide? First, although some natural products are sufficiently potent and selective for their targets that the raw natural product itself is used as a drug, frequently this is not the case. When a natural product possesses a useful activity, but is either insufficiently potent or excessively nonselective, the most obvious solution is to modify its structure so as to remedy its defects. These aims can be successfully achieved by the application of traditional medicinal chemistry, in particular via total chemical synthesis of the natural product and analogues therefrom, it can also be

difficult, expensive, and poor-yielding. An alternative is the use of a biosynthetically-based medicinal chemistry program, using biosynthetic knowledge to create derivatives of the natural product(s). In this case, genetic alteration (genetic disruptions, substitutions, or additions) can be expected to alter the chemical products of the pathway. However, such an attempt would be entirely quixotic without a strong prior foundation of biochemical knowledge supporting the notion that the pathway will remain functional when modified, and that the desired products will result from whatever modification.

Another important consequence of the study of natural products biosynthesis is the development of novel catalysts. To make a gross generalization: all natural products are synthesized by enzyme catalysts. Interestingly, however, these catalysts frequently carry out transformations that would be difficult for synthetic chemists to perform. Indeed, as noted above, the chemical synthesis of complex natural products can frequently be quite difficult. Thus, another benefit to the study of natural products biosynthesis is the isolation of catalysts specific for a given transformation. Biochemical study of these catalysts can then answer the question of how generalizable the activity of a given catalyst is. If an enzyme catalyst were sufficiently robust and chemically useful, then it could also find application outside of the natural pathway. In particular, enzymes can be employed to synthesize fine chemicals or pharmaceuticals that are entirely unrelated to the enzymes' natural substrates.(*1*)

With regard to the study of cyanobactin biosynthesis both of the above-described benefits could potentially apply. First, several cyanobactins possess quite promising anti-cancer activities.(*2-4*) However, none of these compounds has been sufficiently promising to advance far into clinical trials. Consequently, since the work described in

the preceding chapters (as well as in other published works) shows that the cyanobactin biosynthetic enzymes are broadly tolerant of amino acid substitutions in the natural-product encoding sequences, it should be possible to employ the natural pathway to synthesize novel derivatives, which may someday find use in medical practice.

The preceding chapters have suggested several biosynthetic strategies for achieving the above aim. First, in chapters 3 and 4, we have shown that the circularization catalyst, PatG is broadly accepting of profoundly unnatural subsrates, thus PatG's substrate selectivity does not pose a roadblock to the possibility of incorporating non-natural amino acids into cyanobactins. A more conservative approach involves mutation of existing cyanobactins using the 20 canonical amino acids, and again the plausibility of this approach is supported by chapters 3 and 4, as well as chapters 5 and 7. Looking beyond the contents of this work, future engineering efforts on cyanobactins (and others) could employ addition of tayloring domains—such as cytochrome P450s—to existing pathways to create new chemical diversity. Directed evolution may also be a fruitful strategy to direct foreign peptide modifying enzymes to the cyanobactin leader peptide, or to evolve altered chemical reactivity within cyanobactin modifying enzymes.

As for the biocatalytic use of cyanobactin enzymes, it is likely too early to tell whether any enzymes described in this work will find such use. However, several of the enzymes described herein are certainly promising for use as catalysts outside of their natural contexts. For one, if the heterocyclase enzymes' substrate selectivity and catalytic mechanism were fully understood so that they could be directed to modify wholly different sequences that do not possess the standard cyanobactin leader peptide, they could prove quite useful for the introduction of heterocycles into diverse protein and

peptide substrates. Additionally, the prenyltransferase enzymes are also potentially quite promising since of all enzymes presented it seems that their substrate selectivity is most relaxed—acting on small molecule phenols as well as linear and cyclic peptides.

In closing, then, the study of natural products biosynthesis should not be ignored in the drive towards increased reliance upon biotechnological tools to support human existence.

1.2 References

1. Saville, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W., and Hughes, G. J. (2010) Biocatalytic assymetric synthesis of chiral amines from ketones applied to sitagliptin manufacture, *Science 329*, 305-309.

2. Carroll, A. R., Coll, J. C., Bourne, D. J., MacLeod, J. K., Zabriskie, T., Ireland, C. M., and Bowden, B. F. (1996) Patellins 1-6 and trunkamide A: novel cyclic hexa-, hepta- and octa-peptides from colonial ascidians, Lissoclinum sp. , *Aust. J. Chem. 49*, 659-667.

3. Ireland, C., and Scheuer, P. J. (1980) Ulicyclamide and ulithiacyclamide, two new small peptides from a marine tunicate, *J. Am. Chem. Soc. 102*, 5688-5691.

4. Ireland, C. M., Durso, A. R., Newman, R. A., and Hacker, M. P. (1982) Antineoplastic cyclic peptides from the marine tunicate Lissoclinum patella, *J. Org. Chem. 47*, 1807-1811.