BIOSYNTHETIC INVESTIGATIONS OF THE PEPTIDE NATURAL PRODUCTS K-26 AND ANTHRAMYCIN

By

Vanessa Victoria Phelan

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

May, 2010

Nashville, Tennessee

Approved by:

Professor Brian O. Bachmann

Professor Gary A. Sulikowski

Professor John A. McLean

Professor Richard N. Armstrong

TABLE OF CONTENTS

PAGE

To my parents:

Words can never express how thankful I am for all that you have given me.

ACKNOWLEDGEMENTS

It has been my privilege to study at Vanderbilt University in the Department of Chemistry and I am eternally grateful to both the university and department for allowing me to pursue my education in such a supportive and collaborative environment. Without the supportive, collaborative and scientific environment at Vanderbilt, my research and subsequent dissertation would not have been possible.

I would like to thank my advisor, Dr. Brian Bachmann, for his support and guidance throughout my graduate career. His scientific insight, support and openness allowed me to follow my scientific curiosity, has provided me with vast knowledge and piqued my interest in further scientific pursuits. I would also like to thank my Ph.D. committee members Dr. Gary Sulikowski, Dr. John McLean and Dr. Richard Armstrong for providing the opportunity for collaboration and support in both the academic and research aspects of my journey at Vanderbilt.

In addition to my advisor and committee members, it is necessary for me to acknowledge and thank my many collaborators at Vanderbilt without whom my research would have never progressed. To Dawn Overstreet, Dr. M. Wade Calcutt and Dr. David Hachey in the Mass Spectrometry Research Core, thank you for your constant willingness to help me accomplish my goals with your insightful suggestions, willingness to share and endless patience. To Dr. John McLean and his entire research group, specifically Randi Gant-Branum and Michal Kliman, thank you for your help and patience during my endeavors into

MALDI. To Dr. Tony Forster, thank you for opening your laboratory to me and taking the time to teach me. And to my “unofficial” collaborators in the Sulikowski lab, thank you for sharing both your knowledge and your chemicals.

Of course, this research would not have been possible without funding. I am grateful for funding from the National Institutes of Health, the Petroleum Research Fund and the Vanderbilt Institute for Chemical Biology for research funding. The Vanderbilt Institute for Chemical Biology, National Institutes of Health and Graduate School also provided me with support through fellowships and assistantships.

To my colleagues in the Bachmann laboratory, both past and present, I have to give my sincerest thanks. It has been a joy working with all of you. Thank you for the scientific input, friendship and camaraderie that you have provided over the years. I am eternally grateful for you accepting me as the dinosaur loving, color coordinating person that I am. I have to specifically thank Rob Scism and Glenna Kramer as my desk-mates for putting up with my never-ending external monologue.

To my many friends, thank you. My time at Vanderbilt has been difficult and without you, I would not have made it. I especially have to thank Brian Turner, Easton Selby, Victor Ghidu, Bill Evans and Kevin Perzynski for holding me up when the ground disappeared from underneath me. I am eternally grateful for your friendship. There are some things that can never be repaid. To Steve Townsend, Melissa Carter, Anh Hoang, the entire Cliffel lab and my Thursday morning and Monday evening soccer buddies, it has been a pleasure.

You made graduate school bearable. Thank you for both your scientific insight and your friendship.

To my family, thank you for everything. Thank you for your faith, love, support and everything else that you have ever given and instilled in me. Thank you for letting me be who I am without question. Thank you for shaping me into the person I have become.

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

LIST OF ABBREVIATIONS

ACE	Angiotensin converting enzyme
A-domain	Adenylation domain
ADP	Adenosine diphosphate
AEP	2-aminoethylphosphonic acid
AHEP	(*R*)-1-amino-2-(4-hydroxyphenyl)ethylphosphonic acid
AMP	Adenosine monophosphate
AnCE	ACE homologue in *Drosophila melanogaster*
A-site	Acceptor site
ATP	Adenosine triphosphate
C-domain	Condensation domain
CoA	Coenzyme A
C-P	Phosphonate bond
C-P-C	Phosphinate bond
Cy-domain	Cyclization domain
Da	Dalton
DAP	2,3-diaminopropionate
DEAE	Diethylaminoethyl
dK-26	Des-acetyl-K-26
DNA	Deoxyribonucleic acid
E-domain	Epimerization domain
EF-Tu	Elongation factor Tu

ESI	Electrospray ionization
E-site	Exit site
GEB	General enzyme buffer
GTP	Guanosine triphosphate
HA	3-hydroxyanthranilic acid
IC_{50}	Half maximal inhibitory concentration
LB	Luria broth
LC	Liquid chromatography
L-DOPA	L-3,4-dihydroxyphenylalanine
LOD	Limit of detection
LysRS	Lysine tRNA synthetase
m/z	Mass to charge ratio
MALDI	Matrix-assisted laser desorption/ionization
MHA	4-methyl-3-hydroxyanthranilic acid
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MT-domain	Methyltransferasae domain
NAT	*N*-acetyltransferase
NCBI	National Center for Biotechnology Information
NCE	New chemical entity
NMR	Nuclear magnetic resonance
NRPS	Non-ribosomal peptide synthetase
NRRL	Northern Regional Research Laboratory

ORF	Open reading frame
Ox-domain	Oxidation domain
PBD	Pyrrolobenzodiazepine
PCR	polymerase chain reaction
PEP	Phosphoenolpyruvate
polydG	Poly deoxyguanine
PPi	Pyrophosphate
PPTase	Phosphopantetheinyl transferase
PPyr	Phosphonopyruvate
P-site	Peptidyl site
Re-domain	Reductase domain
RNA	Ribonucleic acid
SAM	*S*-adenosyl methionine
SDS-PAGE	Sodium dodecylsulfide polyacrylamide gel electrophoresis
SRM	Selected reaction monitoring
T-domain	Thiolation domain
Te-domain	Thioesterase domain
TIGR	The Institute for Genomic Research
TLC	Thin layer chromatography
TOF	Time-of-flight
tRNA	Transfer ribonucleic acid
TrpRS	Tryptophan tRNA synthetase

CHAPTER I

INTRODUCTION

Natural products have been an invaluable source of therapeutic agents[1, 2]. Of 1024 small molecule new chemical entities (NCEs) introduced between 1981 and 2008, over half were natural products or natural product inspired synthetic derivatives.[2] During this time period, natural products and their derivatives accounted for more than 68% of all anti-infectives (anti-bacterial, -fungal, -parasitic and -viral) and approximately 63% of all anticancer compounds. An additional 225 natural product derived molecules are currently in clinical trials or preclinical development including 86 for cancer treatment and 40 for the treatment of infectious diseases.[3, 4] While natural products are clearly vital for the treatment of human disease, the study of microbial ecosystems has been severely limited by the ability to cultivate less than 1% of the occupying microorganisms leaving a vast untapped resource for drug discovery.[5-8]

Over the past few years, sequencing technologies have advanced substantially, lowering the cost of genome sequencing.[9] Total genome sequencing of a variety of actinomycetes (soil-dwelling bacteria) revealed large numbers of previously unidentified biosynthetic cassettes involved in the biosynthesis of natural products[10-12]. In *Streptomyces avermitilis* alone, over 30 putative gene clusters were identified.[10, 11] In addition to revealing the untapped potential of actinomycetes for possible therapeutics, these genomic

advancements have allowed researchers to correlate previously identified bioactive natural products with their biosynthetic origins. By correlating known natural products to their biosynthetic cassettes, researchers can tap into nature's chemical repertoire by deconstructing the mechanisms required for natural product biosynthesis and harnessing that knowledge to generate novel therapeutics through synthetic biology.[13] Herein, our investigations into understanding the biosynthesis of two previously uninvestigated peptide natural products from actinomycete species: K-26, a potent naturally produced phosphonate containing hypotensive agent, and anthramycin, an anti-cancer pyrrolobenzodiazepine, are described (Figure I-1).[14, 15]

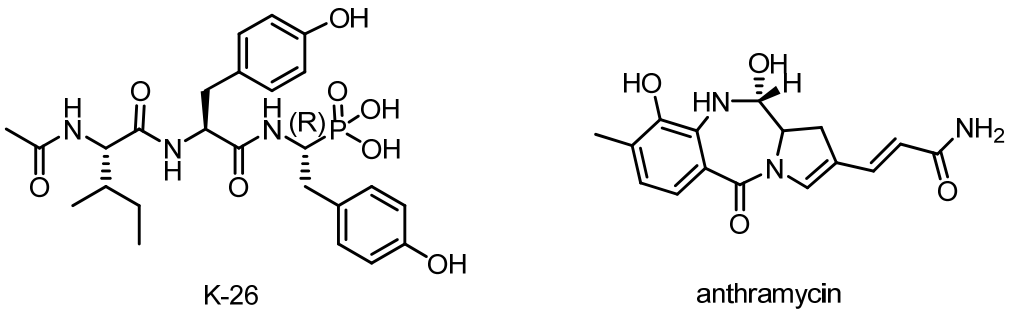


Figure I-1. Chemical Structures of K-26 and Anthramycin K-26 is a unique phosphonate containing tripeptide capable of modulating ACE activity, while anthramycin in an anti-tumor, antibacterial pyrrolobenzodiazepine

Case Study I: K-26

During the last several decades, both natural and synthetic phosphonate and phosphinate containing compounds have been used extensively in medicine and agriculture.[16] (Table I-1) These compounds are similar to phosphate esters and anhydrides, but are distinguished by carbon-phosphorus bonds replacing

one or more oxygen-phosphorus bonds (C-P in phosphonates and C-P-C in phosphinates). The structural similarity of phosphonates and phosphinates to labile phosphate esters and carboxylic acids imparts their ability to act as substrate and transition state analogues of enzyme substrates, inhibiting enzyme catalysis.[17]

Table I-1. Bioactive Phosphonate Containing Compounds

Phosphonate	Class	Application
Cidofovir[1]	antiviral	Cytomegalovirus, HIV
Adefovir[1]	antiviral	hepatitis B, HIV
Fosamax[1]	antiosteoclastic	osteoporosis
Fosfomycin[2]	antibacterial	systemic
Fosmidomycin[2]	antiparasitic	malarial
Bialaphos[2]	herbicidal	weed control

[1]synthetic compounds [2]natural products

In 1959, the first naturally occurring phosphonate, 2-aminoethylphosphonic acid (AEP), was identified.[18] AEP is widely distributed in biological systems as a polar head group of membrane lipids. Subsequently, additional phosphonylated macromolecules including exopolysaccharides and glycoproteins have been described[19, 20]. Furthermore, a variety of phosphonate containing small molecules has been reported. These include bialaphos, an herbicide; fosfomycin, an antibacterial; fosmidomycin, an anti-parasitic compound and K-26, a hypotensive agent.[21] (Figure I-2)

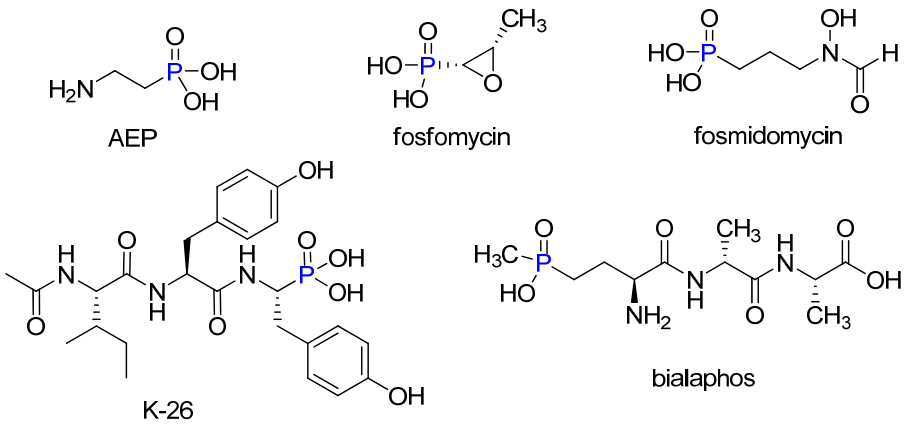


Figure I-2. Phosphonate Natural Products. Representative examples of phosponate containing natural products.

K-26, isolated from *Astrosporangium hypotensionis* (NRRL 12379), is a representative member of an uninvestigated class of phosphonate containing natural products which incorporate a phosphonic acid analogue of tyrosine.[14] NMR, mass spectrometry, degradation and synthetic investigations have demonstrated that K-26 is comprised of *N*-acetylated isoleucine, tyrosine and the non-proteinogenic amino acid (*R*)-1-amino-2-(4-hydroxyphenyl)ethylphosphonic acid (AHEP).[14, 22] AHEP is shared among several related compounds produced by *Streptosporangium* and *Actinomadura* species.[23, 24] (Figure I-3)

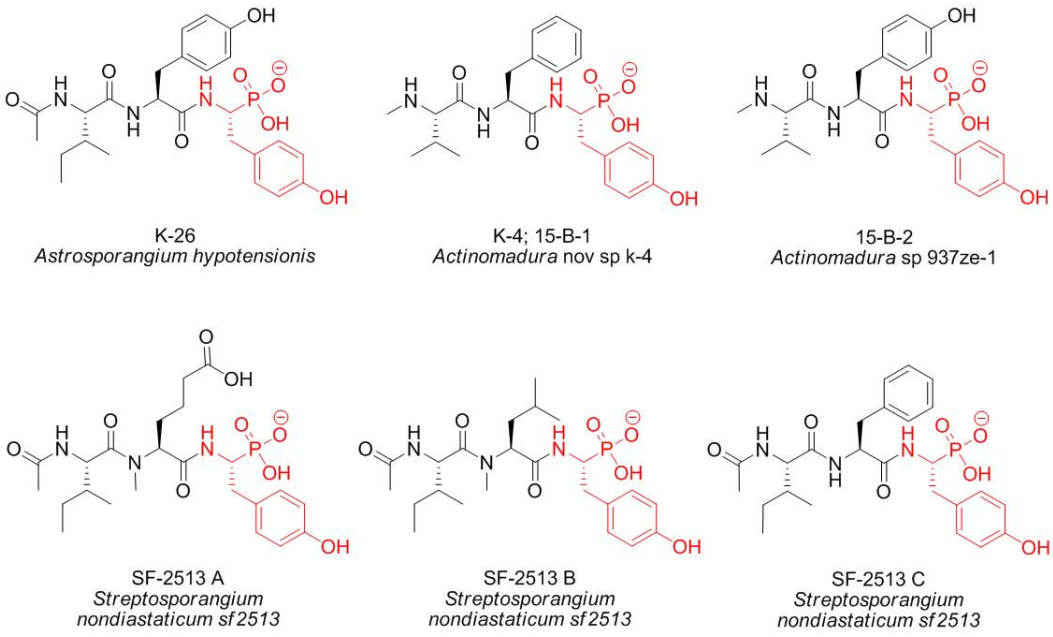


Figure I-3. K-26 and analogues. Highlighted in red is a unique non-proteinogenic amino acid, AHEP. shared among a group of hypotensive tripeptides

K-26 was initially discovered by bioactivity guided fractionation of *A. hypotensionis* culture extracts testing for inhibition of angiotensin converting enzyme (ACE).[14] ACE is a zinc-dependent dipeptidyl carboxypeptidase which catalyzes the cleavage of carboxy-terminal dipeptides from numerous peptide substrates of the Renin-Angiotensin-Aldosterone system.[25] All analogues of K-26 show some inhibition of ACE with K-26 possessing an IC_{50} value of 14.4 nM, comparable to the widely prescribed anti-hypertension drug Captopril (7.7 nM).[26] Initial inhibition studies showed that K-26 is a non-competitive inhibitor of ACE with possible interactions between the phosphonate of AHEP and an active site zinc.[14] Recent structure-activity relationship studies have revealed that the phosphonate moiety of K-26 is required for its inhibitory activity.[26] K-26 was found to be 1500-fold more potent than its carboxylate congener. Additionally,

the absolute configuration of AHEP and *N*-acetylation were found to play an important role in modulating ACE inhibitory activity.

While it may seem unusual to isolate a compound capable of modulating mammalian ACE from a bacterial source, the presence of ACE homologues has been reported in non-vertebrates and unicellular microorganisms.[27-46] Six ACE-like enzymes have been identified in *Drosophila melanogaster*.[28-46] Recent crystallography of one of these enzymes, AnCE, in complex with K-26 revealed that K-26 occupies two of the four available substrate binding pockets.[27, 47] K-26 binds within the S1 and S2 substrate binding sites with two of the phosphonic oxygen atoms coordinating to the active site Zn^{2+} ion, while the third oxygen hydrogen bonds with a histidine and tyrosine within the active site. (Figure I-4) Additional hydrogen bonds mediate and stabilize the binding of K-26 to AnCE. While the inhibition of AnCE by K-26 is lower (160 nM) than the inhibition of ACE (14.4 nM), this crystal structure provides insight into the mechanism of ACE inhibition by K-26 and the potential to develop new investigative tools and therapeutics. It is important to note, however, that the role of K-26 in its environmental interactions, whether endogenous or exogenous, has not been established. Despite the potent hypotensive activity of K-26 and related compounds, the biosynthetic pathway of K-26 has remained cryptic.

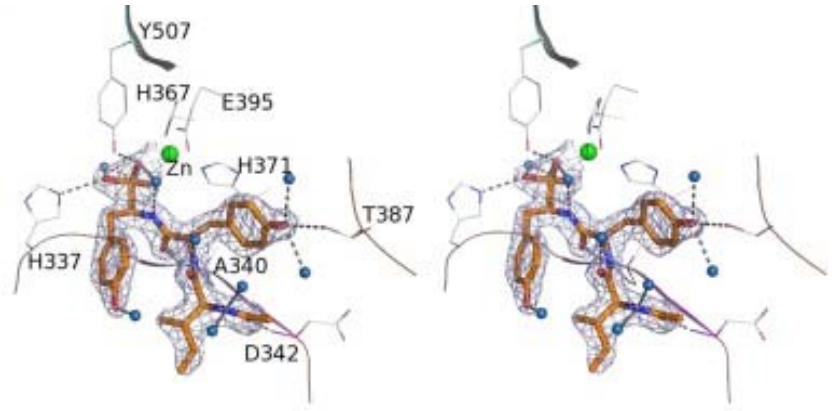


Figure I-4. Binding Mechanism of K-26 in AnCE. A recent crystal structure of K-26 bound to AnCE, an ACE-like protein produced by *D. melanogaster,* revealed interactions between the phosphonate of AHEP and the active site Zn^{2+} of AnCE. Used with permission.[47]

Although completely characterized biosynthetic pathways of C-P bond containing natural products are rare, ongoing investigations have started to unravel the biosynthetic mysteries of several phosphonate natural products including AEP, fosfomycin and bialaphos.[48-53] These investigations have revealed a unified biosynthetic route for the formation of the C-P bond. The key enzyme involved in the generation of the phosphonate moiety in all previously studied C-P bond containing natural products is phosphonenolpyruvate mutase (PEP mutase).[54-56] PEP mutase catalyzes the intramolecular rearrangement of phosphoenolpyruvate to phosphonopyruvate, which has been identified as a common biosynthetic antecedent of all PEP mutase derived phosphonate natural products.[57] (Figure I-5)

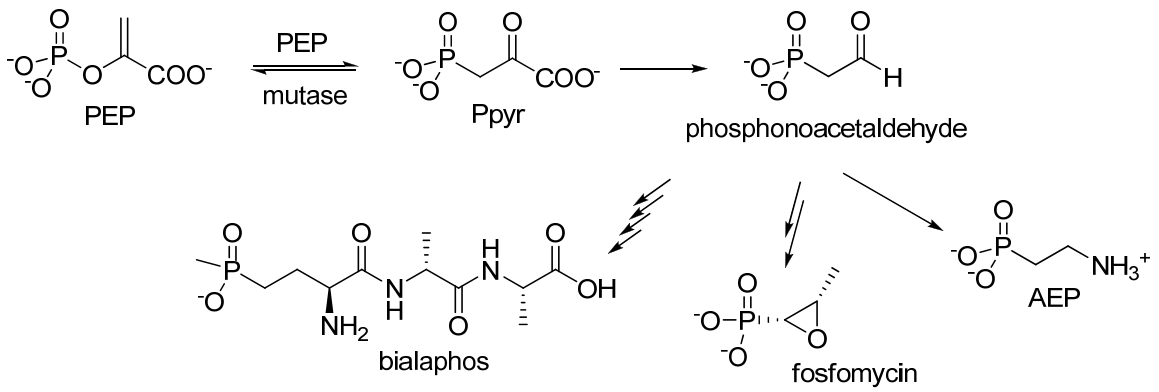


Figure I-5. Biosynthetic Route to Phosphonate Moiety. All previously studied phosphonate natural products derive their phosphonate moiety from the conversion of phosphoenolpyruvate (PEP) to phosphonopyruvate (Ppyr) by PEP mutase.

While the catalytic mechanism of PEP mutase has been widely studied, a unified mechanism has remained elusive. There are two possible mechanisms consistent with the experimentally confirmed intramolecular phosphoryl transfer retaining absolute stereochemistry at the phosphorus.[54, 57-59] (Figure I-6) In mechanism I, the phosphoryl group is transferred to an active site residue of PEP mutase which is then attacked by the C(3) of the pyruvate enol intermediate. While in mechanism II, metaphosphate dissociates from PEP and is attacked by the C(3) of the pyruvate enol intermediate. It is prosposed that in mechanism II, interactions with active site residues hold the metaphosphate in place, while rotation around the C1-C2 bond of the pyruvate enolate to allow attack of the C(3) onto the metaphosphate, which would account for the retention of stereochemistry at the phosphorus as well as the stereochemistry of the C1 position of phosphonopyruvate.

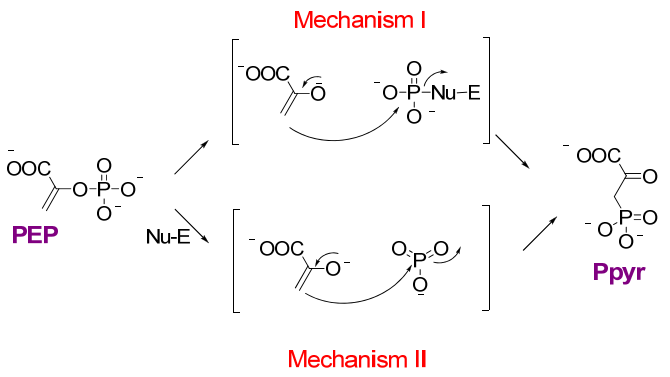


Figure I-6. Possible Mechanisms for PEP mutase. There are two plausible mechanisms for PEP mutase. In mechanism I, the phosphate of PEP is transferred to the active site of PEP mutase which is the attacked by the C(3) of the pyruvate enol intermediate. In mechanism II, the phosphate dissociates from PEP mutase as metaphosphate which is then attacked by the C(3) of the pyruvate enol intermediate.

Although PEP mutase is the biosynthetic route of all previously investigated C-P bond containing natural products, it is difficult to rationalize a possible biosynthetic pathway through phosphonopyruvate to generate AHEP. Comparison of AHEP to known C-P bond containing natural products makes it apparent that the mechanism of AHEP formation may be radically different than that of PEP mutase or an analogous enzyme. Initial labeled precursor incorporation investigations into K-26 biosynthesis have revealed AHEP is a discrete intermediate of K-26 and is derived from tyrosine without the formation of a β-elimination intermediate.[22, 60]

Case Study II: Anthramycin

Anthramycin is representative of a class of peptide natural products characterized by a 1,4-benzodiazepine ring system.[15, 61] This class of compounds is composed of four pharmacophore groups: 1,4-benzodiazepine-2-ones, 1,4-benzodiazepines-2,5-diones, dibenzodiazepinones and

pyrrolobenzodiazepines (PBDs). (Figure I-7) Over 50 1,4-benzodiazepine-2-ones are currently in clinically use for treatment of psychological diseases including insomnia, alcohol dependency and anxiety.[61] The most prominent member of this group is diazepam (Valium). 1,4-benzodiazepene-2,5-ones have been tested as anticonvulsants, peptidomemetic inhibitors and herbicides.[62-66] Within this group, only cyclopenin has been isolated from nature, identified as an intermediate to viridicatin, an inhibitor of HIV replication.[67] The sole member of the dibenzodiazepinone group is the naturally produced TLN-4601 (ECO-4601) and is characterized by broad spectrum antitumor activity and is currently under phase I/II clinical trials.[68, 69] The final group, PBDs, includes both natural and synthetic compounds. These compounds, including anthramycin, are potent alkylating agents reacting in the minor groove of DNA, imparting both antibacterial and antitumor activities.[70-74]

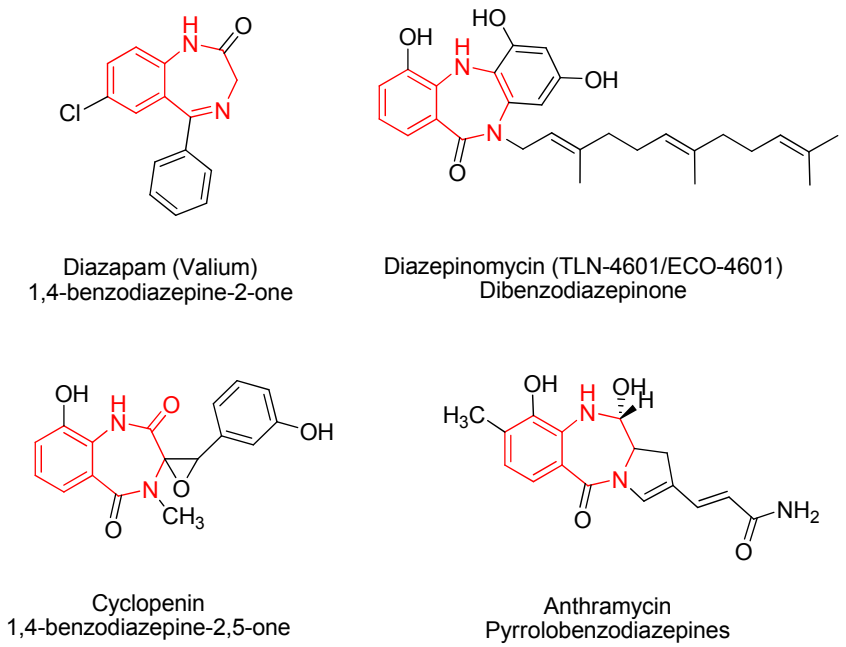


Figure I-7. Classes of 1,4-benzodiazepine. Representative members of the 1,4-benzodiazepine family of compounds. The core 1,4-benzodiazepine structure is highlighted in red.

PBDs are structurally defined by a tricyclic ring system consisting of an aromatic A-ring, a seven-membered diazepine B-ring and a hydropyrrolidine C-ring.[75] Chemical diversity of PBDs is introduced by varied substitution patterns of the A-ring and unsaturation at the C2-C3 bond and/or an exocyclic moiety at the C2 position of the C-ring (Figure I-8). Currently, 19 natural PBDs have been isolated from bacterial cultures, including anthramycin. While 16 natural PBDs show moderate antimicrobial properties, their antitumor activities are the most intriguing.[15]

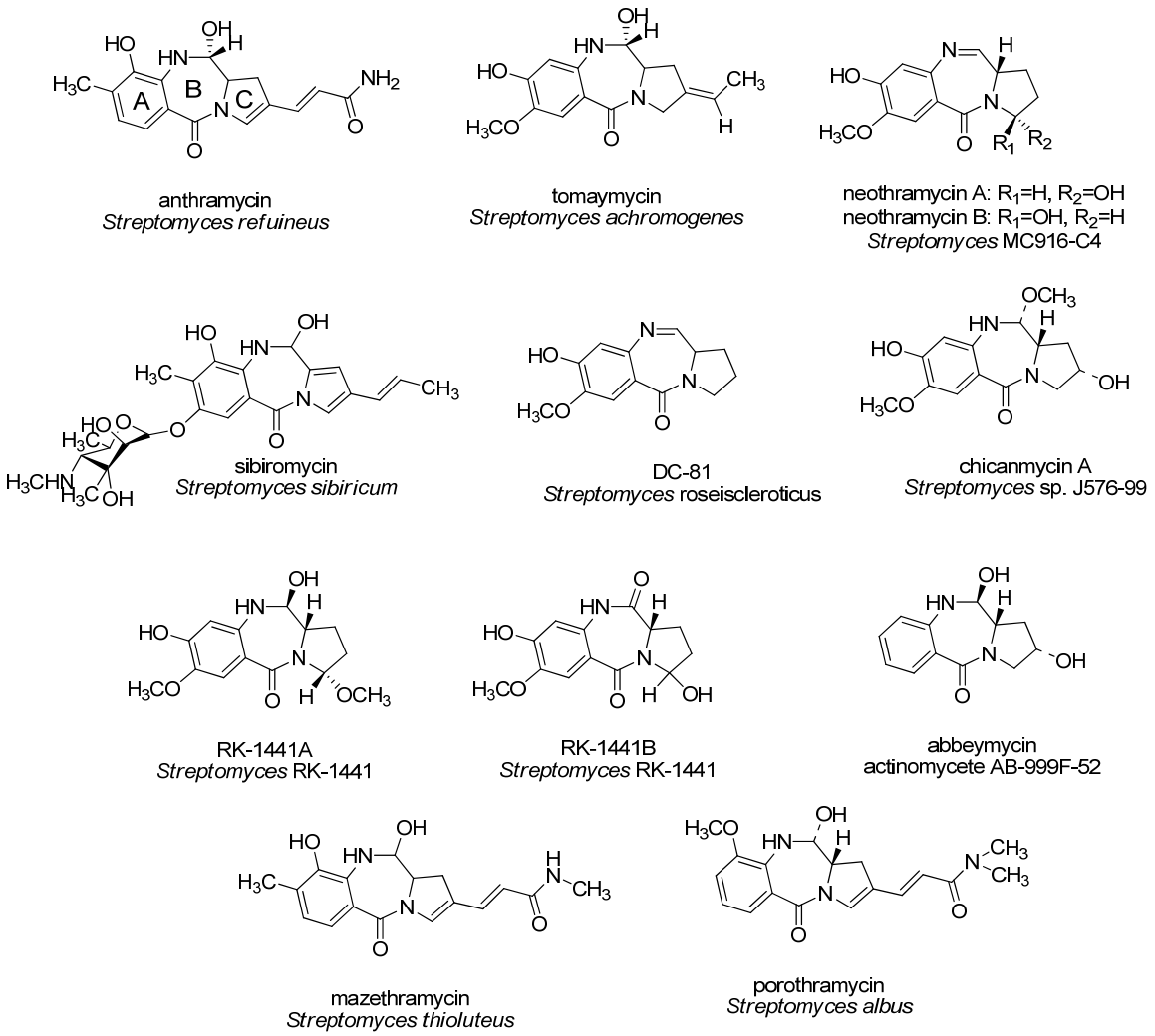


Figure I-8. Examples of Bacterial Pyrrolobenzodiazepines. Chemical diversity of PBDs is introduced by varied substitution patterns of the A-ring and unsaturation at the C2-C3 bond and/or an exocyclic moiety at the C2 position of the C-ring**.** Inactive PDBs have a carbonyl at the C11 of the B-ring, exemplified here by RK1441B

Anthramycin has shown broad spectrum antitumor activity *in vitro* including against HeLa, Sarcoma 180, Walker 256 carcinosarcoma and leukemia P388 cell lines.[76] In initial human studies, of 219 patients with unresponsive tumors, including lymphomas and sarcomas, approximately half saw a significant decrease in tumor size.[76] At the concentrations used, 0.02 mg/kg, no side effects were observed. However at concentrations between 0.1 and 0.5 mg/kg,

cardiotoxicity was observed in rats.[77] Structure-activity relationship studies suggest that the C9 hydroxyl of the A-ring may be the source of these cardiotoxic side effects and synthetic analogues have been created reducing this activity.[73, 74, 77, 78]

Initial experiments investigating the mechanism of inhibition of replication and transcription in cells by anthramycin revealed that anthramycin showed preferred binding to double stranded DNA over single stranded DNA and RNA.[79] Co-purification of anthramycin with double stranded DNA indicated covalent modification of DNA by anthramycin, while experiments with single stranded polydG suggested that guanine may be the reactive DNA residue.[80, 81] Structural studies including NMR and crystallography illuminated the role of a covalent bond between the C11 position of the B-ring and an exocyclic amino group of a guanine base within the DNA binding sequence[74, 82-88]. Anthramycin, sibiromycin and tomaymycin alkylate DNA in a site-specific manner preferring the reactive guanine to be situated between two purine nucleobases rather than pyrimidines.[85, 89, 90] In terms of DNA affinity, the binding order is as follows: sibiromycin > anthramycin > tomaymycin > DC-81 > neothramycin.[61, 72] The conformational flexibility of the PBD and degree of right-handed twist may be responsible for the differences in binding affinity.[91]

Currently, there are two plausible mechanisms for the bioactivity of PBDs consistent with experimental data.[92, 93] In solution, PBDs can exist in three interconverting forms: the imine form, the carbinolamine form and in methanol, the carbinolamine methyl ether form. In mechanism I, a direct S_N2 reaction of the

guanine amine on the carbinolamine carbon with water elimination leads to alkylation. In mechanism II, alkylation results from a direct reaction of the guanine amine with the imine form. (Figure I-9)

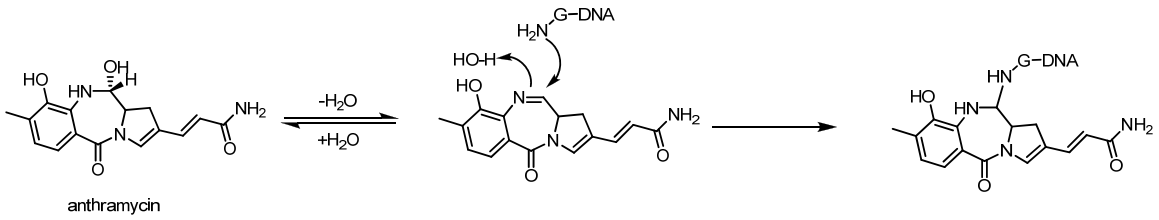


Figure I-9. Proposed Mechanism for PDB Bioactivity. Research strongly suggests that DNA alkylation occurs from a direction reaction of a guanine amine within the minor groove of B-DNA with the imine form of PBDs.

Mechanism II is strongly supported by the formation of two diastereomers of tomaymycin-DNA adducts (11*S,* 11a*S* and11*R*, 11a*S*) in a proportion different than while free in solution and increased reactivity of PBDs when electron-donating substitutions are present on the A-ring.[82, 87, 94, 95] The *S* configuration at the C11a position allows the PBD scaffold to adopt a right-handed twist permits binding within the minor groove of B-DNA. While there are four possible binding configurations of PBDs (two for each diasteromer: 5' to 3' or 3' to 5' orientations), anthramycin mainly binds with the C-ring on the 5'-end of the modified DNA strand resulting only in the 11*S*, 11a*S* adduct, likely due to the 35^{o} twist of anthramycin causing steric clashing in the other three binding configurations.[87, 96] (Figure I-10) Hydrogen bonding of the A-ring substituents and the C-ring acrylamide tail may increase the stability of this adduct.

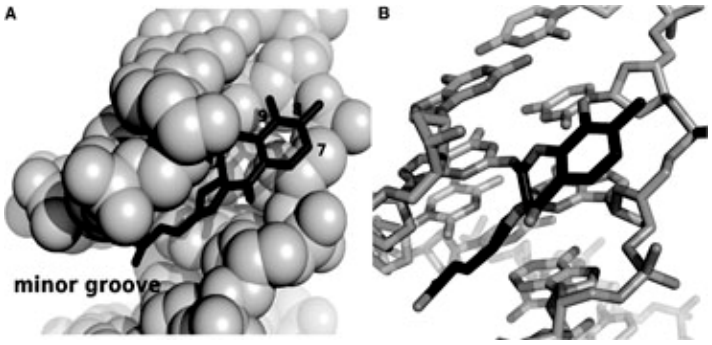


Figure I-10 Anthramycin Bound in the Minor Groove of DNA. The duplex is represented in CPK in (A) and sticks in (B). Used with permission.[15]

Despite the potent biological activities of PBDs, little is known about their biosynthesis. The structural similarities of PBDs indicate that they may be derived from common precursors through parallel biosynthetic pathways. Labeled precursor feeding studies with the producers of anthramycin, sibiromycin and tomaymycin by Hurley and coworkers identified the amino acid precursors of these PBDs[97-99]. (Figure I-11) The anthrinilate A-ring of all three PBDs is derived from L-tryptophan with retention of the indole nitrogen, while the dehydroproline C-ring is derived from L-tyrosine. Labeled L-methionine was incorporated into the 4-position of the A-ring and the C-1 position of the exocyclic side chain of the C-ring. Glucose provides the precursor of the sugar incorporated into sibiromycin.

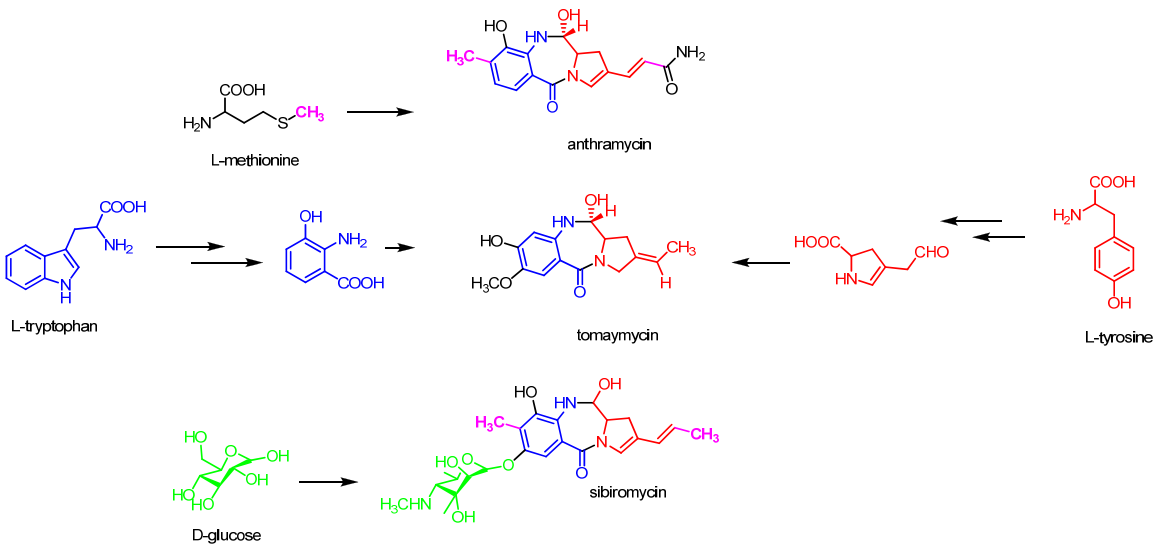
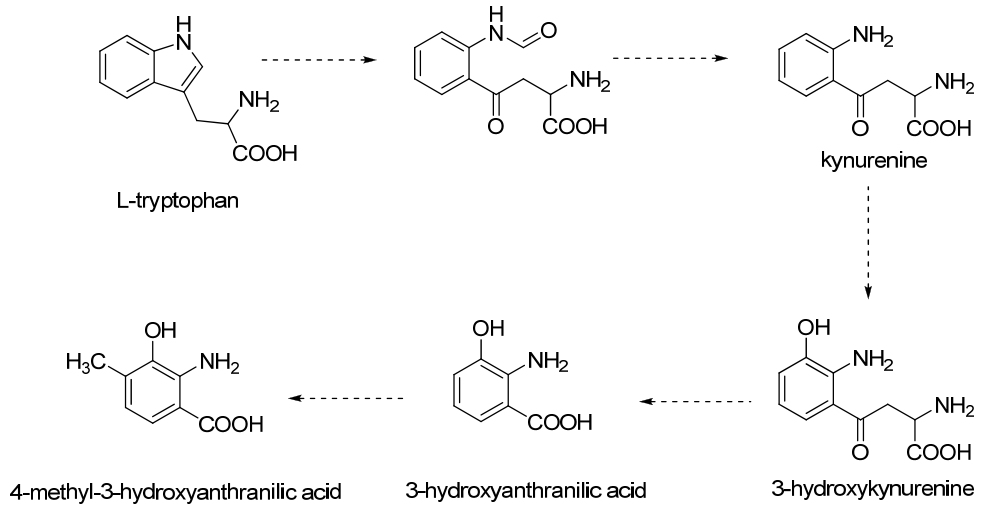


Figure I-11. Precursor Incorporation Studies with Bacterial PBDs. Hurley and coworkers established the amino acid precursors to bacterial PDBs. The A-ring is derived from tryptophan and the B-ring is derived from tyrosine, while methylations are methionine dependent.

Based on the structural similarity of the A-ring of these PBDs to anthranilic acid, a primary metabolite, additional investigations with ^{14}C-labeled anthranilic acid were performed.[97, 98, 100, 101] However, efficient incorporation of anthranilic acid was only detected in tomaymycin. The A-ring of anthramycin and sibiromycin share a similar core: 4-methyl-3-hydroxyanthranilic acid (MHA). However, feeding studies with 3-hydroxy-4-methyl-[2-^{14}C]-anthranilic acid resulted in incorporation solely into sibiromycin. Lack of labeled MHA incorporation into anthramycin may be due to the impermeability of *Streptomyces refuineus* to anthranilate analogues. The incorporation of L-tryptophan and anthranilic acid derivatives into tomaymycin and sibiromycin suggest that the A-ring of PBDs may be from oxidative ring opening of L-tryptophan to L-kynurenine followed by the primary kynurenine pathway to give the anthranilic acid derivatives. (Scheme I-1) Competition experiments identified four possible

intermediates: L-kynurenine, 3-hydroxykynurenine, 3-hydroxy-4-methylkynurenine and MHA for the biosynthesis of the sibiromycin A-ring supporting the role of the kynurenine pathway in the formation of the A-ring.

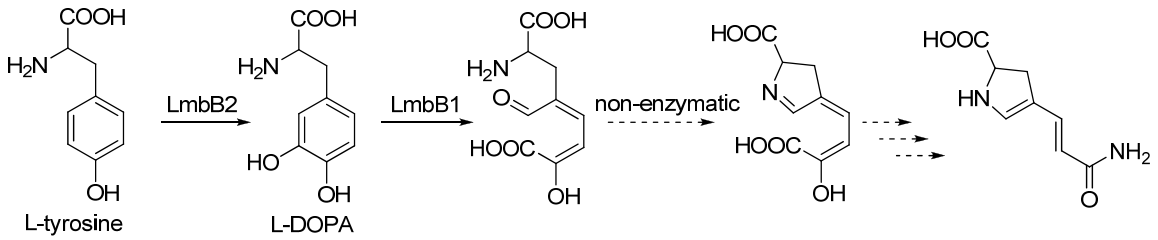


Scheme I-1. Proposed MHA Biosynthetic Pathway. Hurley *et al* proposed the derivation of MHA from the primary metabolic kynurenine pathway based on radiolabeled precursor incorporation studies.

The biosynthesis of the dehydroproline C-ring of anthramycin, sibiromycin and tomaymycin may proceed through a pathway similar to lincomycin biosynthesis in *Streptomyces lincolnensis*.[102, 103] In order to form the C-ring from L-tyrosine, an oxidative cleavage was proposed. The biotransformation of L-tyrosine to the five membered proline ring structure would first proceed through conversion of L-3,4-dihydroxyphenylalanine (L-DOPA) followed by subsequent oxidative cleavage. Feeding experiments performed by Hurley and coworkers support a route in which the pi-bond between C5 and C6 in tyrosine is broken prior to the formation of the five-membered ring. After pentacyclic ring closing, insertion of the exocyclic moiety would occur. (Scheme I-2)

Characterization of a strain of *S. lincolnensis* unable to produce lincomycin led to the isolation and identification of 4-propylidene-3,4-dihydropyrrole-2-carboxylic acid, a proposed intermediate in the biosynthetic pathway of lincomycin.[103] Genomic analysis of the lincomycin gene cluster revealed that this intermediate is due to the inability of this strain to synthesize a cofactor required for an F450 reductase, a participant in lincomycin biosynthesis.[104] Limited functional assignment could be generated for other genes within this biosynthetic cassette due to little to no similarity to characterized proteins within the Protein Data Base (PDB).

Currently, only the activities of two enzymes within this pathway have been assigned.[105, 106] LmbB1 contains the catechol-2,3-dioxygenase signature sequence Hx7FYx2DPxGx3E and catalyzes the conversion of L-DOPA to 4-(3-carboxy-3-oxopropenyl)2,3-dihydropyrrole-2-carboxylic acid through a transient intermediate with an absorbance of 378nm.[107] Structural studies support the enzymatic formation of a semialdehyde followed by non-enzymatic cyclization to the α-hydroxyacid. The activity of LmbB2 was indirectly assigned by observing to 4-(3-carboxy-3-oxopropenyl)2,3-dihydropyrrole-2-carboxylic acid only in cells transformed with a plasmid containing both *lmbB1* and *lmbB2*.[108] Based on the structural similarity of PDBs to lincomycin, the C-ring of the PBDs may follow a similar biosynthetic route.



Scheme I-2. Proposed Dehydroproline Acrylamine Biosynthesis. Precursor incorporation studies identified tyrosine as the amino acid precursor to the dehydroproline acrylamide hemisphere of anthramycin. Subsequent work with the analogous pathway in lincomycin confirmed the first two steps in the proposed biosynthesis.

Peptide Bond Biosynthesis

K-26 and anthramycin are unified by their peptidic nature and the incorporation of non-proteinogenic amino acids. Currently, there are three known paradigms for peptide biosynthesis: ribosomal, non-ribosomal and ligase-based.

Non-ribosomal Peptide Natural Product Biosynthesis

As the most prevalent biosynthetic paradigm for peptide natural products, non-ribosomal peptide synthetases (NRPS) have been shown to incorporate some 500 different monomers including the 20 proteinogenic amino acids, non-proteinogenic amino acids, fatty acids and α-hydroxy acids.[109, 110] The structural diversity imparted by the precursor assortment may contribute substantially to the broad clinical applications of non-ribosomal peptide natural products. Usually, the biosynthetic machinery of non-ribosomal peptides is clustered together in a large cassette. In addition to the genes encoding NRPSs, genes for amino acid biosynthesis, glycosylation, acetylation, transport and resistance may be found within the biosynthetic cassette.[110] NRPS are very large (>100 kD) proteins comprised of distinct modules, each responsible for the

incorporation of one discrete monomer into the final peptide.[111-114] Each module minimally consists of at least three domains: an adenylation (A)-domain, a thiolation (T)-domain and a condensation (C)-domain.[110, 111]

The A-domain selects the cognate amino acid and catalyzes the formation of a reactive aminoacyl adenylate intermediate utilizing ATP and releasing pyrophosphate.[115] (Figure I-12) Using biochemical, structural and phylogenic information, the amino acids of the A-domain involved in amino acid substrate binding were revealed.[116, 117] The crystal structure of PheA identified the amino acids involved in L-phenylalanine binding. This 8- to 10- amino acid specificity-conferring code can be extracted from primary sequence data of A-domains by sequence alignment to PheA. It is possible to predict the substrate specificity of a given A-domain based on identity of the specificity-conferring code to previously characterized A-domains.[118]

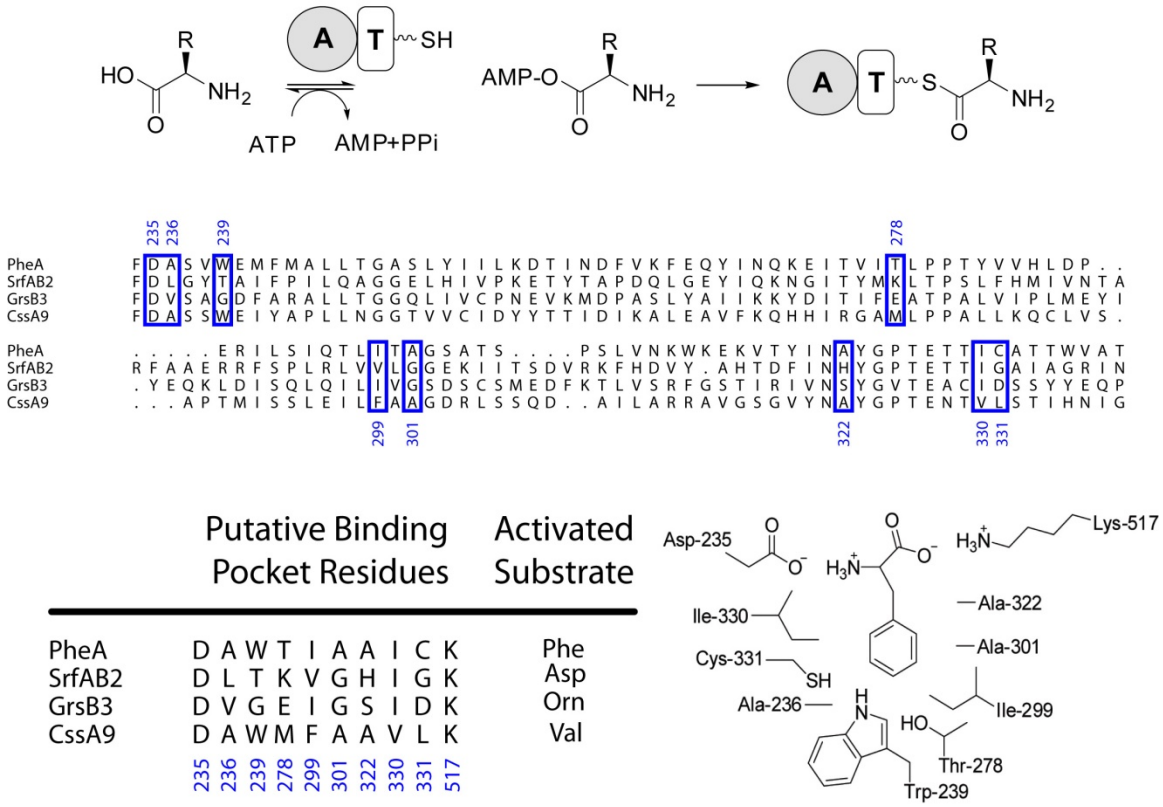


Figure I-12. A-domain Reactivity and Specificity. Top: An NRPS A-domain catalyzes the formation of an aminoacyl adenylate by utilizing ATP and subsequently reacts the aminoacyl adenylate with the free thiol of the phosphopantetheinyl moiety of the T-domain. Middle: Based on the crystal structure of PheA, 10 amino acids were identified involved in substrate discrimination. These amino acids confer a substrate specificity code (Lysine 517 not shown). Bottom Left: Representative examples A-domains and their respective specificity code. Bottom Right: A schematic of substrate binding within the amino acid binding pocket of PheA.

After formation of the reactive aminoacyl adenylate intermediate, the A-domain tethers the amino acid to a post-translationally modified T-domain by thioesterification.[119, 120] (Figure I-13) All T-domains must be post-translationally modified prior to fulfilling their role as transport proteins during peptide bond formation. Formation of the active *holo*-T-domain requires transfer of the 4'phosphopantetheinylate of coenzyme A (CoA) to an active site serine residue of the *apo*-T-domain by a phosphopantetheinyl transferase (PPTase). The activated aminoacyl adenylate reacts with the terminating sulfhydryl group of the

phosphopantetheinyl moiety resulting in an energy-rich thioester primed for condensation into the growing peptide product.

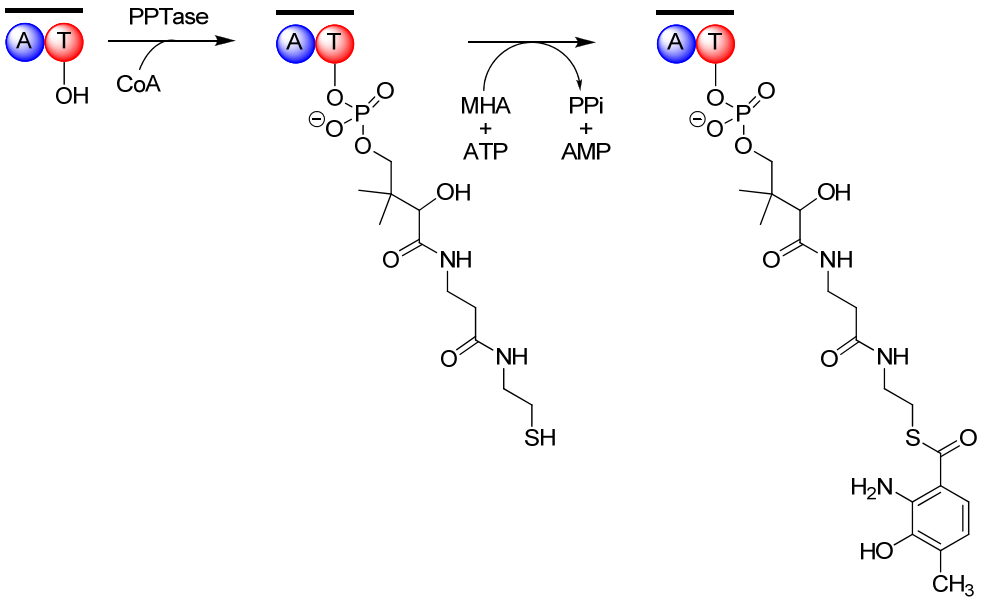


Figure I-13. The Role of the T-Domain in NRPS Systems. After post-translational modification of an active site serine residue with phosphopantethiene, the A-domain activates the cognate amino acid substrate as the aminoacyl adenylate and catalyzes the thioesterification of the phosphopantetheinyl moiety with the aminoacyl adenylate to give the tethered amino acid primed for condensation into the peptide.

Peptide bond formation is catalyzed by the C-domain by promoting nucleophilic attack of the amino group of the activated amino acid bound to the downstream (of the C-domain) module onto the acyl group of the amino acid tethered to the upstream module.[121, 122] (Figure I-14) Based on the recent crystal structure of a T-C didomain from tyrocidine biosynthesis, the C-domain has a V-shaped architecture, possessing sites for both the upstream and downsteam T-domain tethered amino acids.[123, 124] Additional substrate specificity for the nucleophilic amino acid substrate is imparted within the C-domain by the

structure of the respective binding pocket preventing incorporation of an incorrect amino acid into the peptide.

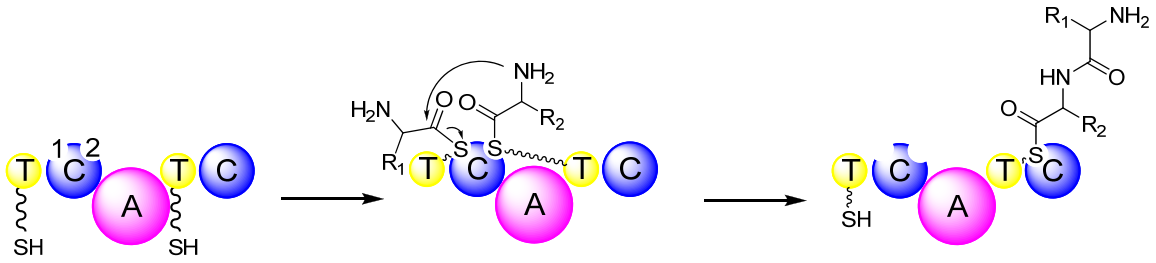


Figure I-14. C-Domain Reaction Mechanism. Two T-domains of neighboring modules are loaded with their cognate amino acid. The electrophile amino acid enters the acceptor site (1) and the nucleophile amino acid enters the donor site (2) of the C-domain. After peptide bond formation, the resulting dipeptide can now act as the electrophile in the active site of the subsequent C-domain.

In most NRPS, the peptide is released by the activity of a thioesterase (Te)-domain.[125, 126] The growing peptide chain is handed from one module to the next until it reaches the last T-domain on the final module. Release of the final peptide is catalyzed by a two-step process through an acyl-*O*-Te-enzyme intermediate that is attacked by either an internal nucleophile on the peptide or water, resulting in either a cyclic or linear peptide, respectively.[127, 128] Alternatively, a reductase domain can release the peptide with a terminal aldehyde or alcohol.

Additional structural diversity is incorporated into non-ribosomal peptides by tailoring domains.[129] These domains fall into two categories: integral parts of the NRPS acting in *cis* on the growing peptide or distinct enzymes acting in *trans*. These domains include cyclization (Cy)-domains, which form thiazoline and oxazoline rings from cysteine, serine and threonine residues; oxidation (Ox)-domains, responsible for oxidation of the thiazoline and oxadoline rings to

thiazoles and oxazoles, respectively; epimerization (E)-domains epimerize an L amino acid to a D amino acid and methyltransferase (MT)-domains insert a methyl group on an amino acid substrate using *S*-adenosyl methionine as the methyl donor.

Initial investigations into the biosynthesis of non-ribosomal peptides identified a direct link between the order of the modules of the NRPS, including the domains within those modules, and the incorporation of the amino acid into the growing peptide.[112, 130-132] This hierarchy of reactions was coined the co-linearity rule. (Figure I-15) For example, the cyclic lipopeptide surfactin follows this biosynthetic route.[133] A total of 24 domains are organized in seven modules over three NRPS that act linearly, starting at module 1 and ending at module 7. However, as more gene clusters were identified, deviations from the co-linearity rule were identified. Two types of deviations are iterative, exemplified by gramicidin S, and non-linear, exemplified by vibriobactin.[112, 134] In gramicidin S biosynthesis, a pentapeptide monomer is formed and stalled on the final module.[127, 135] The regenerated NRPS then engages in a second round of synthesis and is released by a head-to-tail cyclization to give the final product. In non-linear biosynthesis of an NRPS, the domains are arranged non-linearly, often missing expected domains, repeat use of single domains or include inactive domains.[112]

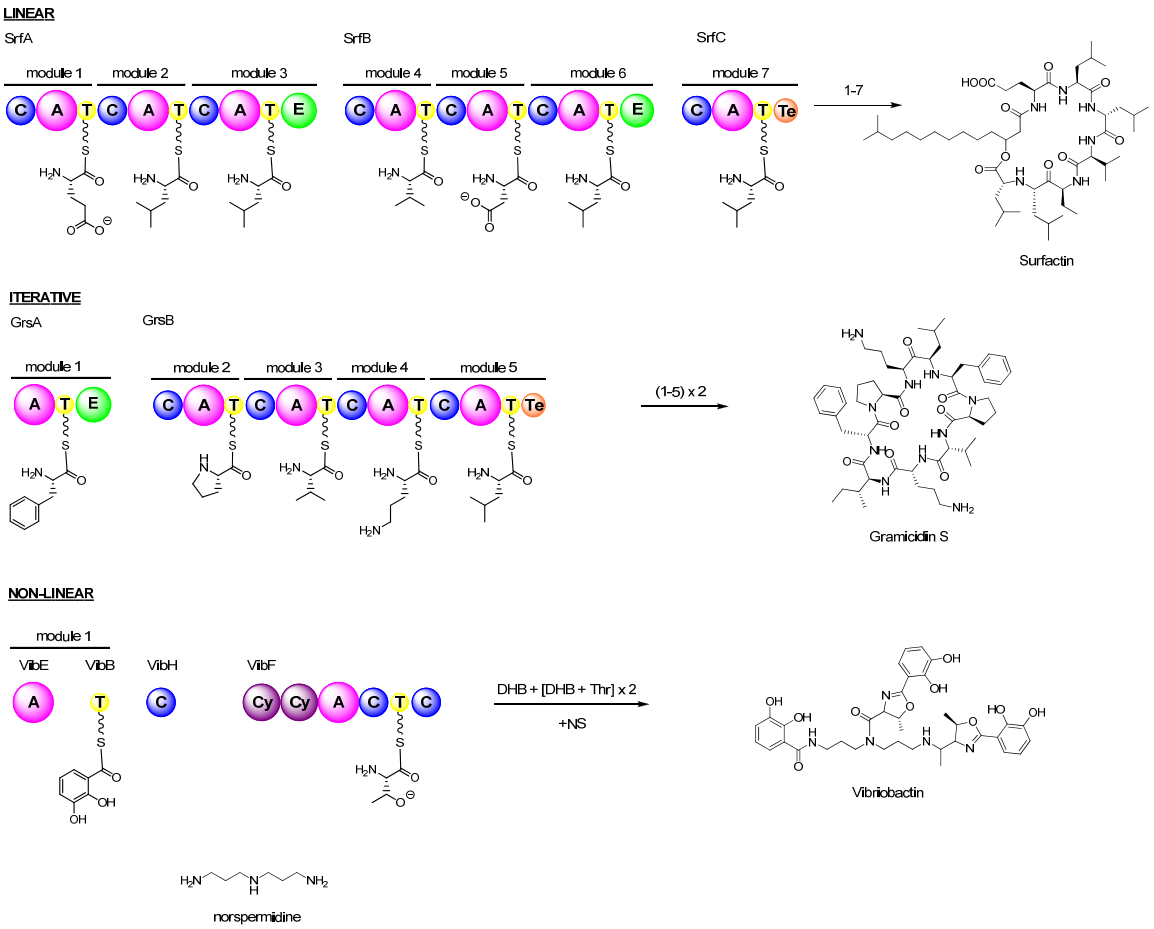


Figure I-15. NRPS Biosynthetic Strategies. Top: In linear NRPSs, synthesis proceeds from module 1 to module 7 until the final product is released by the Te-domain of the last module. Middle: In iterative NRPSs, the modules of the NRPS are used multiple times in succession and the Te-domain catalyzes the dimerization of the two peptides to gives a homodimeric product. Bottom: Non-linear NRPSs, do not follow the co-linearity rule.

Ribosomal Peptide Natural Product Biosynthesis

While NRPSs are the most common biosynthetic route for peptide natural products, in recent years, a growing number of ribosomally encoded cyclic peptides have been identified.[136-147] Briefly, three basic enzymes are required for ribosomal production of peptides. The aminoacyl-tRNA synthetase first selects the cognate amino acid and loads it onto either the 2' or 3' hydroxyl group of the corresponding tRNA, depending on whether the synthetase is classified as a

Class I or II, respectively.[148, 149] The ribosome is the catalytic machine responsible for generating the amino acid sequence of a peptide, based on the sequence of the bound mRNA.[150] The ribosome selects the correct aminoacyl-tRNA, with the help of elongation factor (EF)-Tu. The aminoacyl-tRNA, EF-Tu and GTP enter the acceptor (A)-site of the ribosome. The large ribosomal subunit stimulates hydrolysis of GTP committing the aminoacyl-tRNA to peptide bond formation, releasing EF-Tu. The amino acid-tRNA complex is bound via codon-anticodon recognition to the ribosome. Peptide bond formation occurs through nucleophilic attack of the amino group of the aminoacyl-tRNA in the A-site on the carbonyl of the aminoacyl-tRNA, with the attached peptide, in the peptidyl (P)-site. The tRNA with the attached polypeptide is now pushed from the A-site to the P-site, while the unloaded tRNA in the P-site shifts to the exit (E)-site. Subsequent cycles elongate the polypeptide.

Unlike non-ribosomal peptides, ribosomal peptides are limited by the 20 proteinogenic amino acids. However, extensive posttranslational modification of the precursor peptide imparts structural diversity to the peptide scaffold. For a majority of ribosomal peptide natural products, a series of genes encode a precursor peptide containing an N-terminal leader peptide and a C-terminal core peptide that is processed into the final compound.[151] (Figure I-16) Variably, the N-terminus of the leader peptide may be appended with a signal peptide governing subcellular localization and/or a recognition sequence may be attached to the C-terminal end of the core peptide. After post-translational modification, the precursor peptide is usually cleaved by at least one protease

freeing the core peptide. The gene cluster containing the precursor peptide also encompasses genes encoding modifying enzymes, resistance and secretion.

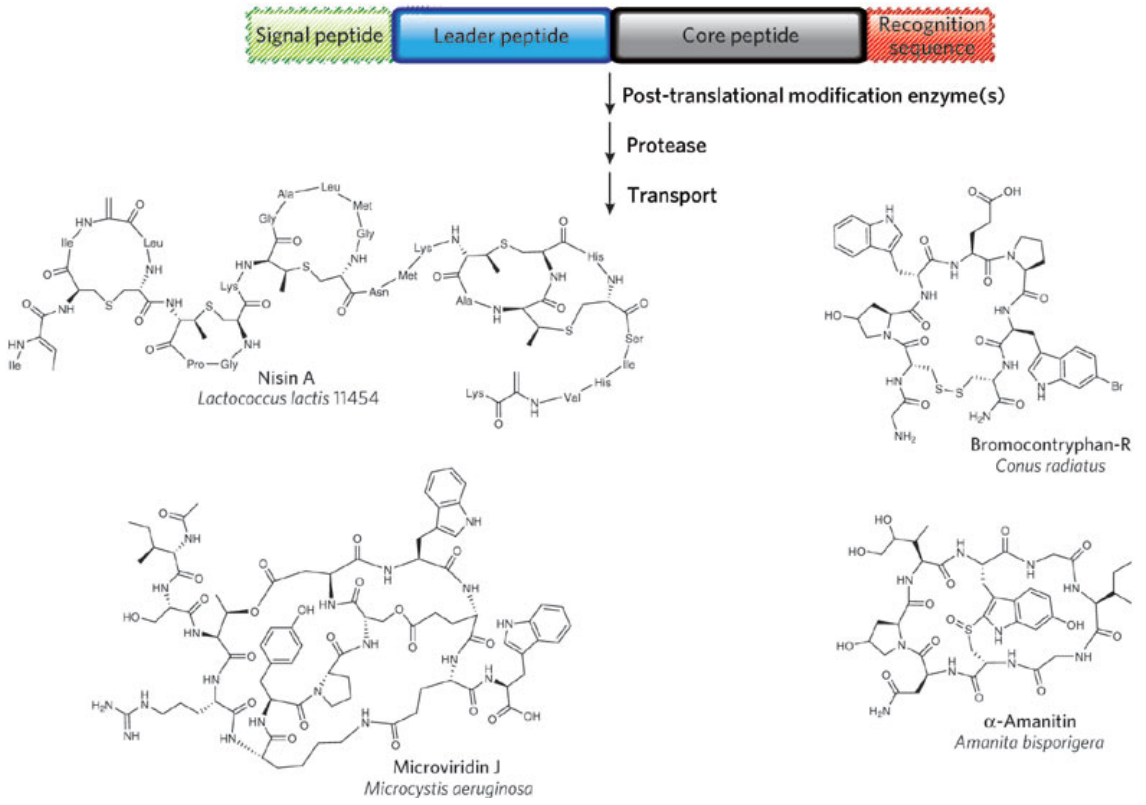


Figure I-16. General Ribosomal Peptide Biosynthesis. The precursor peptide usually consists of an N-terminal leader peptide and a C-terminal core peptide. A signal peptide and/or recognition sequence may be appended to the main peptide structure. The precursor peptide is ribosomally synthesized and post-translationally modified to the active compounds. Used with permission[151]

Several roles have been proposed for the N-terminal leader peptide, though none have been completely substantiated experimentally.[151] In fact, the role of the leader peptide may differ between families of ribosomal peptides or even within an individual class. Most often, it is proposed that the leader peptide plays a role in signaling secretion of the core peptide; however, most of the leader peptides in these gene clusters display no homology to proteins involved in bacterial secretion pathways. Additional proposed roles for the leader peptide

include acting as a recognition motif for post-translational modification enzymes including perhaps acting as chaperones for enzymes assisting in folding of the precursor peptide and stabilizing the peptide against degradation and premature activation of its biological function. Supporting the role of the leader peptide as a preventative to premature activation, the protease involved in cleaving the precursor peptide resides on the outside of the producing cell.

Despite the variability of the precursor peptides, both between and within classes, they remain substrates for post-translational modification. Both the side chains and main peptide chain can undergo post-translational modifications. Herein, I provide an overview of several families of ribosomal peptide natural products produced by bacteria and the modifications of those peptides as representative examples of their biosynthesis. These four families were chosen based on their prevalence in nature and the inclusion of post-translational modifications within their structure.

Lantibiotics

Lantibiotics are divided into two classes based on their leader peptide sequences and biosynthetic enzymes. Class I lantibiotics have leader peptides approximately 25 amino acids in length, rich in aspartic acid residues.[152] In this class, the leader peptide greatly reduces the antimicrobial activities of the core peptide and appears to be involved in the recruitment of enzymes for peptide modification and excretion.[153, 154] Class II lantibiotics have leader peptides rich in aspartate and glutamate residues, contain an ELXXBX motif (where B = V, L or I)

and usually end in two glycine residues.[155] The double glycine motif is required for proteolytic cleavage of the precursor peptide, while the leader peptide itself may be involved in secretion and guidance of additional biosynthetic enzymes involved in producing the final natural products.[156-158] A total of 15 post-translational modifications have been documented within the lantibiotic family.[159] (Figure I-17) Structurally, the lantibiotic family of ribosomal peptide natural products is distinguished by thioether crosslinks introduced by the dehydration of serine and threonine to dehydroalanine and dehydrobutyrine, respectively, followed by addition of the thiol of cysteine.

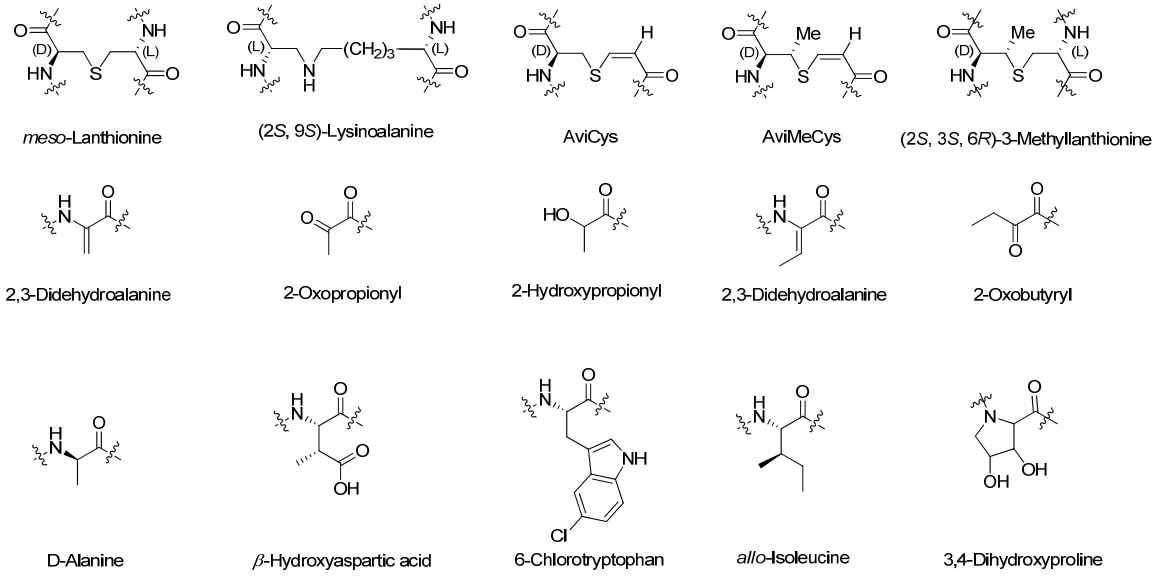


Figure I-17. Post-translational Modifications Found in Lantibiotics. A total of 15 post-translational modifications have been identified in the lantibiotics family. This family of ribosomal peptide natural products is distinguished by the incorporation of a variety of thioether crosslinks.

Microcins

Microcins are small, structurally diverse peptides produced by enterobacteria.[160] As opposed to the lantibiotics, only a few different types of

post-translational modifications have been reported for microcins.[161-164] One class of peptides within this family is exemplified by microcin B17, which contains oxazole and thiazole heterocycles.[161] (Figure I-18) The formation of these heterocycles requires three proteins including a cyclodehydratase, which generates the oxazoline and thiazoline structures from Gly-Ser and Gly-Cys sequences; a flavin-dependent dehydrogenase, responsible for oxidizing these intermediates to the oxazole and thiazole structures and a scaffolding protein, required for catalysis.[165] The leader peptide for these types of microcins has been shown to interact with the scaffolding proteins and reduce the antimicrobial activity of the core peptide.[166, 167]

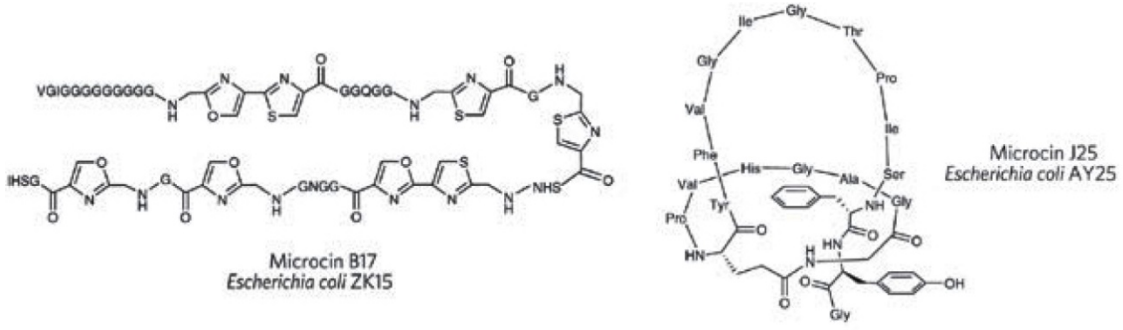


Figure I-18. Microcin Ribosomal Peptides. Two classes found within the Microcin family of peptide natural products are the thiazole and oxazole containing peptides and the lasso peptides. Used with permission.[151]

Another class of peptides within the microcin family is the lasso peptides, exemplified by microcin J25.[162-164] (Figure I-18) In these peptides, the distinguishing feature is a lactam bond. The N-terminal amino group reacts with the carboxylate side chain of an aspartate or glutamate residue at position 8, essentially trapping the C-terminus of the peptide chain threaded through the ring. The leader peptide of microcin J25 has been shown to be required for the

recruitment of two enzymes involved in post-translational modification of the core peptide.[146, 168] One protein catalyzes the adenylation of the aspartate/glutamate carboxylate at position 8, while a second cleaves off the leader peptide and installs the new lactam bond.

Cyanobactins

Produced by diverse cyanobacteria, over 100 unique cyclic peptides belong to the cyanobactin family, including the patellamides.[146] (Figure I-19) Members of this family of ribosomal peptide natural products share similar structural features, like oxazoles and thiazoles, as well as, similar biosynthetic origins. Two enzymes, a cyclodehydratase and a bifunctional enzyme with an oxidase domain, are required for the cyclization of serine, threonine and cysteine.[169] Proteolytic cleavage of the precursor peptide releases the core peptide which is then cyclized to give the final product.

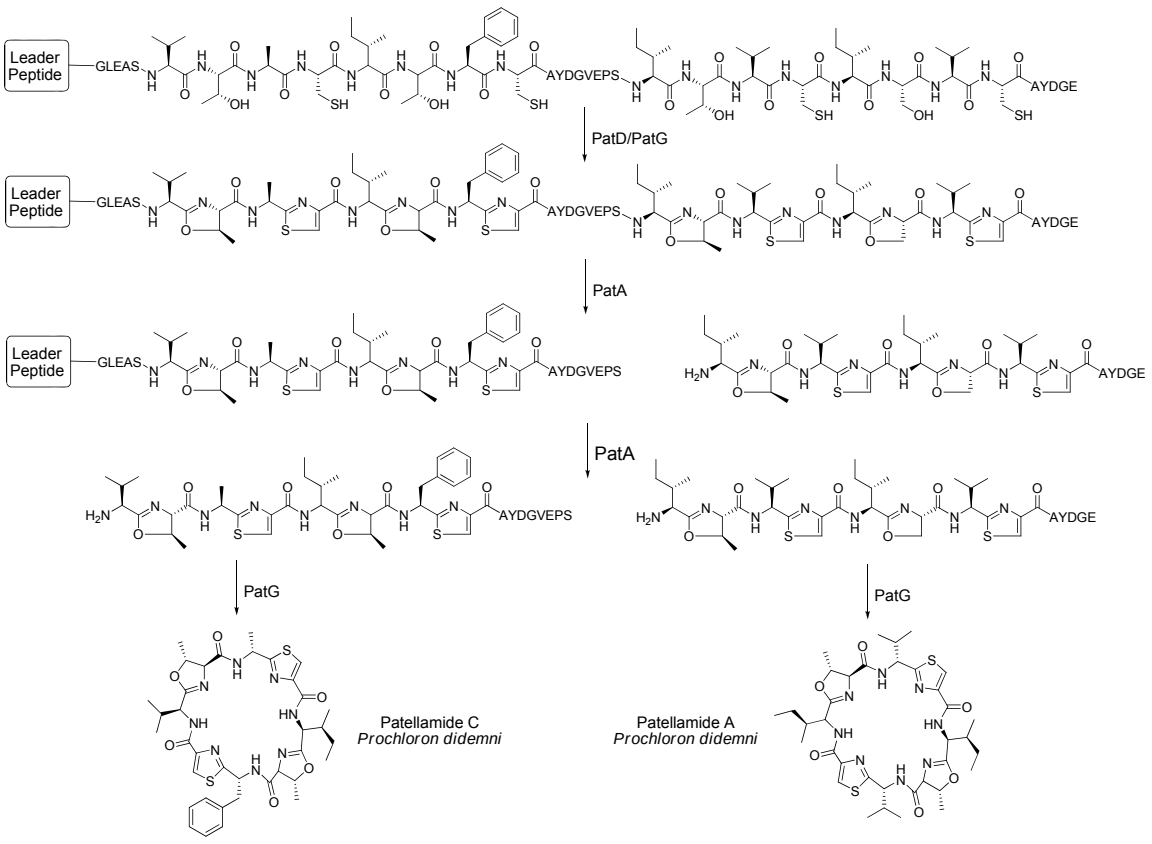


Figure I-19. Proposed Biosynthesis of Patellamides A and C. Cyanbactins differ from other types of ribosomal peptide natural products by incorporating two core peptides in the precursor peptide.

Interestingly, most cyanobactin precursor peptides consist of a leader peptide, whose activities have not been substantiated *in vitro*, and two highly variable core peptides flanked between recognition sequences responsible for directing post-translational modification of the core peptides.[146] The N-terminal region flanking the first core peptide typically contains a conserved G(L/V)E(A/P)S sequence, while the C-terminal region contains a conserved AYDG(E) sequence, which act as recognition sequences for the excision and cyclization of the core peptides.[169]

Thiopeptides

Thiopeptides are another large class of ribosomally encoded natural products containing a string of thiazoles/thiazolines which are conjoined by a six-member heterocyclic ring.[136, 139, 142, 144, 170] (Figure I-20) This heterocycle can be a pyridine, hydroxypyridine or a dehydropiperadine. Common modifications within the thiopeptide family include dehydrations of threonine and serine, thiazole and thiazoline formation and proposed cyclization of two dehydroalanines to form the six-membered ring structure.

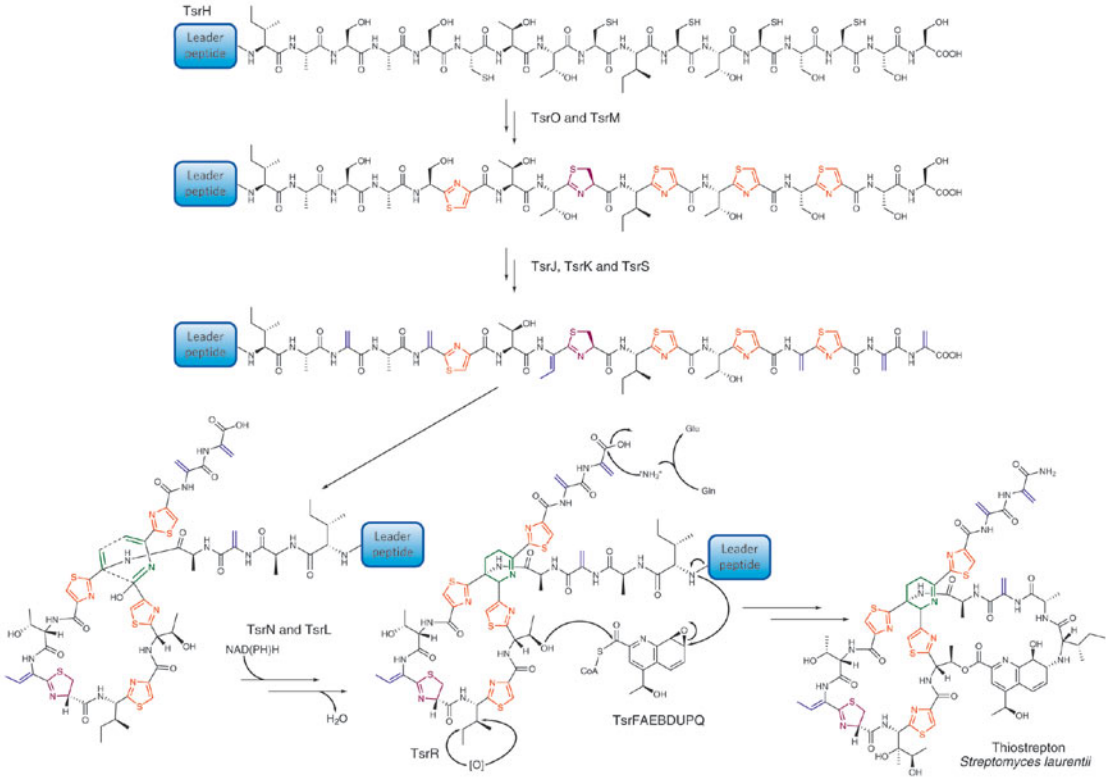


Figure I-20. Biosynthesis of Thiostrepton. Thiostrepton is initially encoded as the precursor peptide TsrH. Subsequent post-translational modifications install the heterocycle rings and dehydrated amino acid. This conformationally restrained peptide is further modified to give the characteristic central hexacycle of thiopeptides. Although represented in a particular order, the timing of the biosynthetic steps is not known. Used with permission.[151]

Ligase-based Peptide Biosynthesis

While non-ribsomal and ribosomal natural products make up the majority of peptide natural products, increasing examples of ligase-based biosyntheses have been reported in the literature, including non-ribosomal independent pathways, aSerRS homologues amino acid:[carrier protein] ligases and the recently identified ATP-independent strategy.[171-175] The biosynthesis of the dapdiamide antibiotics provides one example of the role of ATP-dependent ligases in peptide natural product biosynthesis.[172] (Figure I-21) ATP-dependent ligases are soluble enzymes that couple amino acid monomers to the growing peptide chain. There are two types of ATP-dependent ligases involved in dapdiamide biosynthesis. DdaG activates fumarate, as the aminoacyl adenylate using ATP and releasing pyrophosphate, and forms the first peptide bond between fumarate and 2,3-diaminopropionate (DAP) to create the N_β-fumaroyl-DAP dipeptide intermediate. Activity similar to DdaG occurs during the ATP-dependent biosynthesis of the siderophore ferrioxamine E by the non-ribosomal synthetase independent pathway.[173, 174] After amination of N_β-fumaroyl-DAP to N_β-fumaramoyl-DAP, DdaF activates the dipeptide as the acyl phosphate by using ATP and releasing ADP, similar to the ligases involved in glutathione and bacterial cell wall pentapeptide biosynthesis.[176, 177]

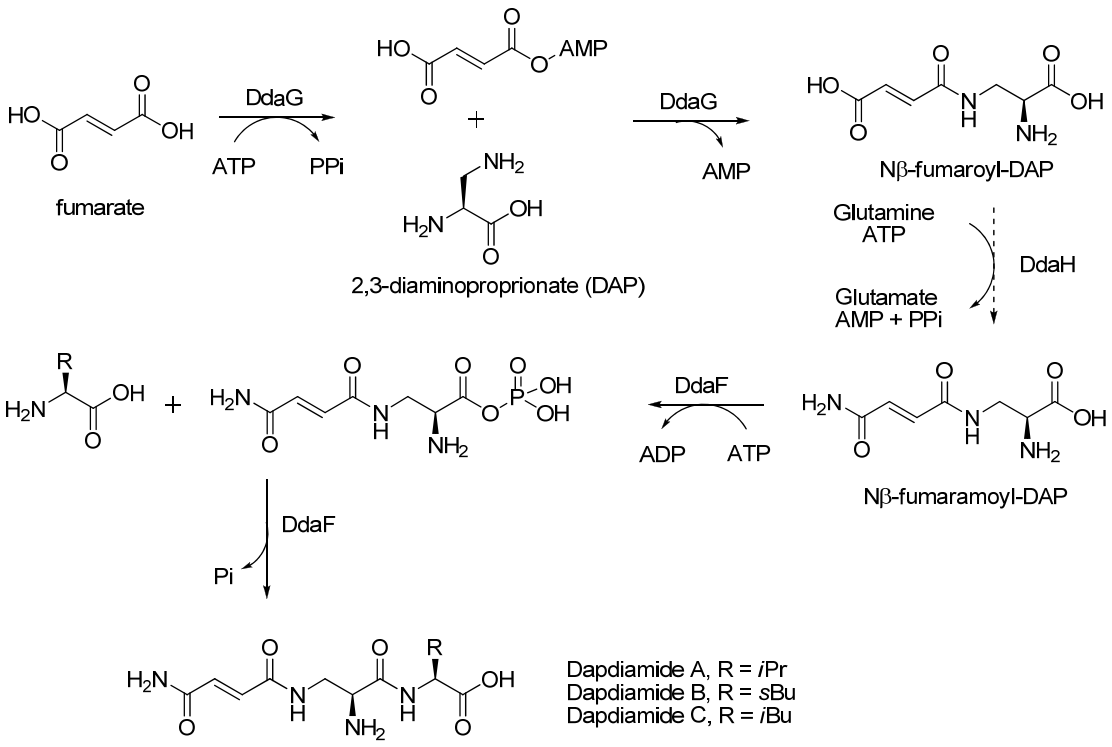


Figure I-21. Dapdiamide Biosynthesis. The biosynthesis of the dapdiamides requires two ATP-dependent ligases, DdaG, which employs an aminoacyl adenylate and DdaF, which utilizes an aminoacylphosphate, to activate amino acids for peptide bond formation.

More recently, two other biosynthetic pathways for peptides have been identified. Homologues to serine tRNA synthetase (aSerRS) have been identified in a variety of bacterial species including *Alpha*- and *Betaproteobacteria*, *Clostridium*, *Bacillus* and *Streptomyces* species.[175] While showing homology to serine tRNA synthetases, these proteins, exemplified by two homologues from *Bradyrhizobium japonicum* (BII0957 and BII6282) and one homologue from *Agrobacterium tumefaciens* (Atu2573), activate amino acids in the same manner as tRNA synthetases, but do not load amino acids onto tRNA. As the genes for these proteins are often near genes for small hypothetical proteins with a 4'phosphopantetheine attachment site, it was proposed that

aSerRS may be capable of loading these T-domain-like proteins called carrier proteins (CP). The three aSerRS (BII0957, BII6282 and Atu2573) were shown to tether alanine or glycine to these neighboring CPs; however, the peptide product of this route of biosynthesis has not been identified.

Unlike the other ligase-based biosynthetic pathways, another recently identified route for peptide bond formation is ATP-independent.[171] Capuramycins, produced by *Streptomyces* sp SANK 62799 and *Streptomyces griseus* SANK 60196, are nucleotide containing peptide natural products. (Figure I-22) The structure of the capuramycins consists of three distinct portions: a 5'-*C*-carbamoyl-uridine, an unsaturated hexuronic acid and an aminocaprolactam ring. The gene cluster for capuramycin A-503083 (A and B) contains two NRPS genes believed to be involved in the biosynthesis of the L-aminocaprolactam ring, while, two additional genes are required for the installation of the aminocaprolactam ring into the final capuramycin structure. CapS activates the carboxylic acid of the hexuronic acid-carbamoyl uridine complex by *S*-adenosyl methionine-dependent methylation. CapW is responsible for the condensation of the hexuronic acid-carbamoyl uridine complex with the amino caprolactam ring.

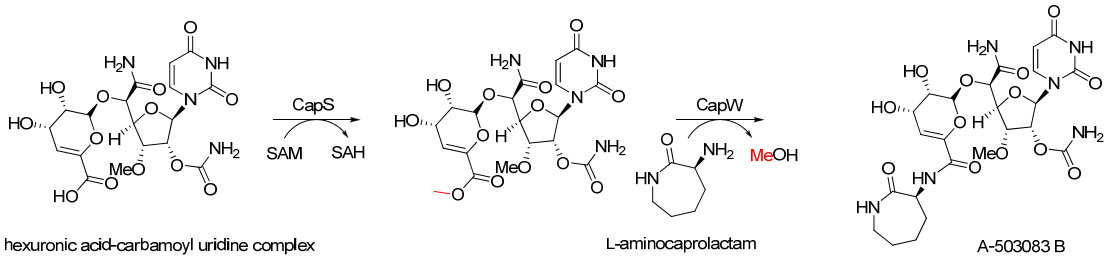


Figure I-22. ATP-independent Peptide Bond Formation. The biosynthetic pathway of the capuramycins includes a unique method for peptide bond formation. CapS installs a methyl on the free carboxylate of the hexuronic acid-carbamoyl uridine complex. This activated complex is then condensed with L-aminocaprolactam by CapW, releasing methanol and creating the final peptide.

Biosynthetic Hypotheses for K-26 and Anthramycin

Based on the chemical structures, we propose both K-26 and anthramycin are produced by NRPS:

K-26

The intermediacy of AHEP suggests that K-26 may be biosynthesized by the NRPS paradigm. In this case, L-isoleucine, L-tyrosine and AHEP would be activated by individual A-domains, loaded onto the corresponding T-domains and peptide bond formation would be catalyzed by the C-domain resulting in the formation of des-acetyl-K-26 (dK-26). (Figure I-23) Then either on the final T-domain or after release of the peptide by the Te-domain, N-acetylation would occur resulting in formation of K-26. Our proposed gene cluster would consist of an N-acetyltransferase, NRPS with a domain string of A-T-C-A-T-C-A-T-Te and hypothetical proteins of unknown function responsible for AHEP biosynthesis.

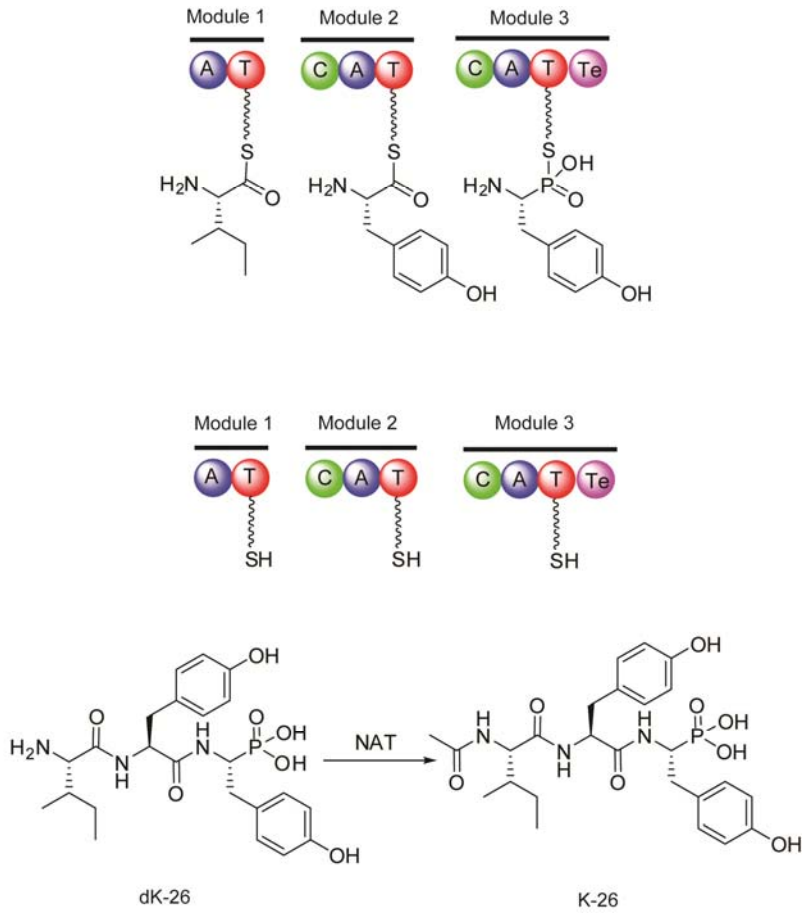


Figure I-23. Proposed Peptide Bond Formation in K-26. Based on the non-proteinogenic nature of AHEP, the most logical biosynthetic route for K-26 formation is via the NRPS paradigm. In this case, the discrete precursors would be activated and loaded onto individual T-domains by the corresponding A-domains. Peptide bond formation would be catalyzed by the C-domains and the peptide would be released by a thioesterase. *N*-acetylation could occur prior to peptide bond formation or afterward as depicted.

Anthramycin

Based on the non-proteinogenic nature of the two amino acids proposed in the formation anthramycin, evidenced by the aforementioned isotopic incorporation experiments, we propose an NRPS based biosynthetic hypothesis with a domain string of A-T-C-A-T-Re. (Figure I-24) The first A-domain of the NRPS would activate either 3-hydroxyanthranilic acid (HA) or MHA and the

second A-domain would activate the dehydroproline acrylamide hemisphere for loading on to their respective T-domains. A specialized reductase domain would provide the final cyclization to form the final benzodiazepine ring of anthramycin.

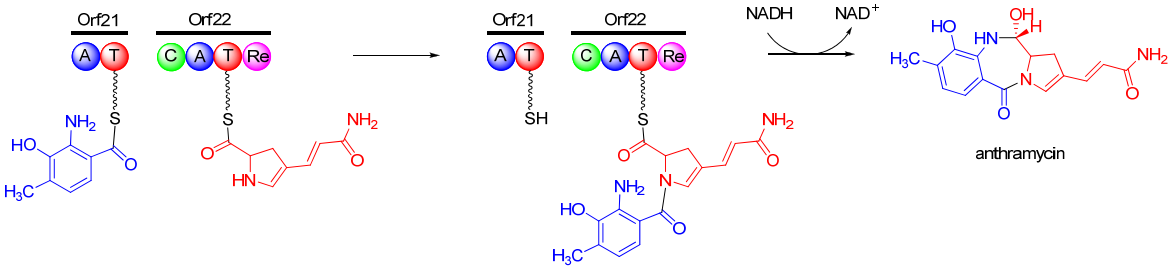


Figure I-24. Proposed Peptide Bond Formation in Anthramycin. Based on the non-proteinogenic nature of the two proposed precursors of anthramycin, peptide bond formation by an NRPS is the most logical hypothesis. In this case, the discrete amino acid precursors would be activated by individual A-domains and loaded onto the corresponding T-domains. The lone C-domain would catalyze peptide bond formation and a unique reductase domain would release anthramycin from the megasynthetase by installing the benzodiazepine ring.

Dissertation Goals

Within this dissertation the first biosynthetic investigations of two natural product classes are described.

Chapter II describes the development of the mass-based pyrophosphate exchange assay for characterization of the proteins involved in the biosynthesis of K-26 and anthramycin.

Chapter III describes the biosynthetic investigations of K-26, including our attempts at isolating the peptide bond forming machinery. This work is currently the only research being performed investigating this unique phosphonate containing natural product.

Chapter IV describes the biosynthetic investigation of anthramycin, including the first biochemical characterization of the enzymes involved in anthramycin biosynthesis. The work described herein is the first identification and characterization of a PBD gene cluster. It is important to note that after publication of our investigations into the biosynthesis of anthramycin, biosynthetic investigations of both sibiromycin and tomaymycin were published by the Gerratana laboratory.

Chapter V describes the possible future biosynthetic investigations of K-26 and anthramycin.

REFERENCES

1. Newman, D. J. and Cragg, G. M. (2007) Natural products as sources of new drugs over the last 25 years. *Journal of Natural Products*, **70**, 461-477.

2. Newman, D. J. and Cragg, G. M. (2009) Microbial antitumor drugs: natural products of microbial origin as anticancer agents. *Current Opinion in Investigational Drugs*, **10**, 1280-1296.

3. Koehn, F. E. and Carter, G. T. (2005) The evolving role of natural products in drug discovery. *Nature Reviews Drug Discovery*, **4**, 206-220.

4. Newman, D. and Cragg, G. (2009) Natural products in medicinal chemistry. *Bioorganic and Medicinal Chemistry*, **17**, 2120.

5. Cragg, G. M., Grothaus, P. G. and Newman, D. J. (2009) Impact of natural products on developing new anti-cancer agents. *Chemical Reviews*, **109**, 3012-3043.

6. Gross, H. (2010) Genomic mining - a concept for the discovery of new bioactive natural products. *Planta Medica*, **76**, 1169-1169.

7. Bode, H. B. and Muller, R. (2005) The impact of bacterial genomics on natural product research. *Angewandte Chemie-International Edition*, **44**, 6828-6846.

8. Pace, N. R. (1997) A molecular view of microbial diversity and the biosphere. *Science*, **276**, 734-740.

9. Mira, A., Martin-Cuadrado, A. B., D'Auria, G. and Rodriguez-Valera, F. (2010) The bacterial pan-genome: a new paradigm in microbiology. *International Microbiology*, **13**, 45-57.

10. Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S. (2003) Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. *Nature Biotechnology*, **21**, 526-531.

11. Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M. (2001) Genome sequence of an industrial microorganism Streptomyces avermitilis: Deducing the ability of producing secondary metabolites. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 12215-12220.

12. Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J. and Hopwood, D. A. (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature*, **417**, 141-147.

13. Keasling, J. D. (2010) Manufacturing Molecules Through Metabolic Engineering. *Science*, **330**, 1355-1358.

14. Yamato, M., Koguchi, T., Okachi, R., Yamada, K., Nakayama, K., Kase, H., Karasawa, A. and Shuto, K. (1986) K-26, a novel inhibitor of angiotensin I converting enzyme produced by an actinomycete K-26. *Journal of Antibiotics (Tokyo)*, **39**, 44-52.

15. Gerratana, B. (2010) Biosynthesis, synthesis, and biological activities of pyrrolobenzodiazepines. *Medicinal Research Reviews*.

16. Kafarski, P. and Lejczak, B. (2001) Aminophosphonic acids of potential medical importance. *Current Medicinal Chemistry - Anticancer Agents*, **1**, 301-312.

17. Wanke, C. and Amrhein, N. (1993) Evidence that the reaction of the UDP-N-acetylglucosamine 1-carboxyvinyltransferase proceeds through the O-phosphothioketal of pyruvic acid bound to Cys115 of the enzyme. *European Journal of Biochemistry*, **218**, 861-870.

18. Horiguchi, M. and Kandatsu, M. (1959) Isolation of 2-aminoethane phosphonic acid from rumen protozoa. *Nature*, **184(Suppl 12)**, 901-902.

19. Hilderbrand, R. L. *The Role of phosphonates in living systems*, CRC Press: Boca Raton, Fla., 1983.

20. Moschidis, M. C. (1984) Phosphonolipids. *Progress in Lipid Research*, **23**, 223-246.

21. Metcalf, W. W. and van der Donk, W. A. (2009) Biosynthesis of Phosphonic and Phosphinic Acid Natural Products. *Annual Review of Biochemistry*, **78**, 65-94.

22. Ntai, I., Manier, M. L., Hachey, D. L. and Bachmann, B. O. (2005) Biosynthetic origins of C-P bond containing tripeptide K-26. *Organic Letters*, **7**, 2763-2765.

23. Koguchi, T., Yamada, K., Yamato, M., Okachi, R., Nakayama, K. and Kase, H. (1986) K-4, a novel inhibitor of angiotensin I converting enzyme produced by Actinomadura spiculosospora. *Journal of Antibiotics (Tokyo)*, **39**, 364-371.

24. Kido, Y., Hamakado, T., Anno, M., Miyagawa, E., Motoki, Y., Wakamiya, T. and Shiba, T. (1984) Isolation and characterization of I5B2, a new phosphorus containing inhibitor of angiotensin I converting enzyme produced by Actinomadura sp. *Journal of Antibiotics (Tokyo)*, **37**, 965-969.

25. Acharya, K. R., Sturrock, E. D., Riordan, J. F. and Ehlers, M. R. (2003) Ace revisited: a new target for structure-based drug design. *Nature Reviews Drug Discovery*, **2**, 891-902.

26. Ntai, I. and Bachmann, B. O. (2008) Identification of ACE pharmacophore in the phosphonopeptide metabolite K-26. *Bioorganic & Medicinal Chemistry Letters*, **18**, 3068-3071.

27. Riviere, G., Michaud, A., Corradi, H. R., Sturrock, E. D., Acharya, K. R., Cogez, V., Bohin, J. P., Vieau, D. and Corvol, P. (2007) Characterization of the first angiotensin-converting like enzyme in bacteria: Ancestor ACE is already active. *Gene*, **399**, 81-90.

28. Appleford, P. J., Griffiths, M., Yao, S. Y. M., Ng, A. M. L., Chomey, E. G., Isaac, R. E., Coates, D., Hope, I. A., Cass, C. E., Young, J. D. and Baldwin, S. A. (2004) Functional redundancy of two nucleoside transporters of the ENT family (CeENT1, CeENT2) required for development of Caenorhabditis elegans. *Molecular Membrane Biology*, **21**, 247-259.

29. Coates, D., Siviter, R. and Isaac, R. E. (2000) Exploring the Caenorhabditis elegans and Drosophila melanogaster genomes to understand neuropeptide and peptidase function. *Biochemical Society Transactions*, **28**, 464-469.

30. Coates, K. J., Haines, H., Bobbitt, J., Kondekas, N. and Isaac, J. (2000) Investigation of the pork supply chain into Singapore. *Asian-Australasian Journal of Animal Sciences*, **13**, 87-90.

31. Cornel, M. J., Williams, T. A., Lamango, N. S., Coates, D., Corvol, P., Soubrier, F., Hoheisel, J., Lehrach, H. and Isaac, R. E. (1995) Cloning and Expression of an Evolutionary Conserved Single-Domain Angiotensin-Converting Enzyme from Drosophila-Melanogaster. *Journal of Biological Chemistry*, **270**, 13613-13619.

32. Ekbote, U., Coates, D. and Isaac, R. E. (1999) A mosquito (Anopheles stephensi) angiotensin I-converting enzyme (ACE) is induced by a blood meal and accumulates in the developing ovary. *Febs Letters*, **455**, 219-222.

33. Houard, X., Williams, T. A., Michaud, A., Dani, P., Isaac, R. E., Shirras, A. D., Coates, D. and Corvol, P. (1998) The Drosophila melanogaster-related angiotensin-I-converting enzymes Acer and Ance - Distinct enzymic characteristics and alternative expression during pupal development. *European Journal of Biochemistry*, **257**, 599-606.

34. Isaac, R. E., Ekbote, U., Coates, D. and Shirras, A. D. (1999) Insect angiotensin-converting enzyme - A processing enzyme with broad substrate specificity and a role in reproduction. *Neuropeptides: Structure and Function in Biology and Behavior*, **897**, 342-347.

35. Isaac, R. E., Macgregor, D. and Coates, D. (1996) Metabolism and inactivation of neurotransmitters in nematodes. *Parasitology*, **113**, S157-S173.

36. Isaac, R. E., Michaud, A., Keen, J. N., Williams, T. A., Coates, D., Wetsel, W. C. and Corvol, P. (1999) Hydrolysis by somatic angiotensin-I converting enzyme of basic dipeptides from a cholecystokinin gastrin and a LH-RH peptide extended at the C-terminus with Gly-Arg/Lys-Arg, but not from diarginyl insulin. *European Journal of Biochemistry*, **262**, 569-574.

37. Isaac, R. E., Schoofs, L., Williams, T. A., Corvol, P., Veelaert, D., Sajid, M. and Coates, D. (1998) Toward a role for angiotensin-converting enzyme in insects. *Trends in Comparative Endocrinology and Neurobiology*, **839**, 288-292.

38. Isaac, R. E., Schoofs, L., Williams, T. A., Veelaert, D., Sajid, M., Corvol, P. and Coates, D. (1998) A novel peptide-processing activity of insect peptidyl-dipeptidase A (angiotensin I-converting enzyme): the hydrolysis of lysyl-arginine and arginyl-arginine from the C-terminus of an insect prohormone peptide. *Biochemical Journal*, **330**, 61-65.

39. Isaac, R. E., Siviter, R. J., Stancombe, P., Coates, D. and Shirras, A. D. (2000) Conserved roles for peptidases in the processing of invertebrate neuropeptides. *Biochemical Society Transactions*, **28**, 460-464.

40. Isaac, R. E., Williams, T. A., Sajid, M., Corvol, P. and Coates, D. (1997) Cleavage of arginyl-arginine and lysyl-arginine from the C-terminus of pro-hormone peptides by human germinal angiotensin I-converting enzyme (ACE) and the C-domain of human somatic ACE. *Biochemical Journal*, **328**, 587-591.

41. Laurent, V., Brooks, D. R., Coates, D. and Isaac, R. E. (2001) Functional expression and characterization of the cytoplasmic aminopeptidase P of Caenorhabditis elegans. *European Journal of Biochemistry*, **268**, 5430-5438.

42. Lee, A. J., Huntley, J., Van den Broek, A., Coates, D. and Isaac, R. E. (2002) Expression and characterisation of a Psoroptes ovis glutathione S-transferase. *Veterinary Parasitology*, **105**, 49-63.

43. Lee, A. J., Isaac, R. E. and Coates, D. (1999) The construction of a cDNA expression library for the sheep scab mite Psoroptes ovis. *Veterinary Parasitology*, **83**, 241-252.

44. Meyer, D. J., Muimo, R., Thomas, M., Coates, D. and Isaac, R. E. (1996) Purification and characterization of prostaglandin-H E-isomerase, a sigma-class glutathione S-transferase, from Ascaridia galli. *Biochemical Journal*, **313**, 223-227.

45. Siviter, R. J., Coast, G. M., Winther, A. M. E., Nachman, R. J., Taylor, C. A. M., Shirras, A. D., Coates, D., Isaac, R. E. and Nassel, D. R. (2000) Expression and functional characterization of a Drosophila neuropeptide precursor with homology to mammalian preprotachykinin A. *Journal of Biological Chemistry*, **275**, 23273-23280.

46. Turner, A. J., Isaac, R. E. and Coates, D. (2001) The neprilysin (NEP) family of zinc metalloendopeptidases: genomics and function. *Bioessays*, **23**, 261-269.

47. Akif, M., Ntai, I., Sturrock, E. D., Isaac, R. E., Bachmann, B. O. and Acharya, K. R. (2010) Crystal structure of a phosphonotripeptide K-26 in complex with angiotensin converting enzyme homologue (AnCE) from Drosophila melanogaster. *Biochemical and Biophysical Research Communications*, **398**, 532-536.

48. Seto, H. and Kuzuyama, T. (1999) Bioactive natural products with carbon-phosphorus bonds and their biosynthesis. *Natural Products Reports*, **16**, 589-596.

49. Warren, W. A. (1968) Biosynthesis of phosphonic acids in Tetrahymena. *Biochimica Et Biophysica Acta*, **156**, 340-346.

50. Hidaka, T., Goda, M., Kuzuyama, T., Takei, N., Hidaka, M. and Seto, H. (1995) Cloning and Nucleotide-Sequence of Fosfomycin Biosynthetic Genes of Streptomyces-Wedmorensis. *Molecular & General Genetics*, **249**, 274-280.

51. Thompson, C. J. and Seto, H. (1995) Bialaphos. *Biotechnology*, **28**, 197-222.

52. Schwartz, D., Berger, S., Heinzelmann, E., Muschko, K., Welzel, K. and Wohlleben, W. (2004) Biosynthetic gene cluster of the herbicide phosphinothricin tripeptide from Streptomyces viridochromogenes Tu494. *Applied and Environmental Microbiology*, **70**, 7093-7102.

53. Blodgett, J. A., Thomas, P. M., Li, G., Velasquez, J. E., van der Donk, W. A., Kelleher, N. L. and Metcalf, W. W. (2007) Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide. *Nature Chemical Biology*, **3**, 480-485.

54. Seidel, H. M., Freeman, S., Seto, H. and Knowles, J. R. (1988) Phosphonate biosynthesis: isolation of the enzyme responsible for the formation of a carbon-phosphorus bond. *Nature*, **335**, 457-458.

55. Bowman, E. D., McQueney, M. S., Scholten, J. D. and Dunaway-Mariano, D. (1990) Purification and characterization of the Tetrahymena pyriformis P-C bond forming enzyme phosphoenolpyruvate phosphomutase. *Biochemistry*, **29**, 7059-7063.

56. Hidaka, T., Seto, H. and Imai, S. (1989) Biosynthetic Mechanism of C-P Bond Formation - Isolation of Carboxyphosphonoenolpyruvate and Its Conversion to Phosphinopyruvate. *Journal of the American Chemical Society*, **111**, 8012-8013.

57. Bowman, E., Mcqueney, M., Barry, R. J. and Dunaway-mariano, D. (1988) Catalysis and Thermodynamics of the Phosphoenolpyruvate Phosphonopyruvate Rearrangement - Entry into the Phosphonate Class of Naturally-Occurring Organo-Phosphorus Compounds. *Journal of the American Chemical Society*, **110**, 5575-5576.

58. Freeman, S., Seidel, H. M., Schwalbe, C. H. and Knowles, J. R. (1989) Phosphonate Biosynthesis - the Stereochemical Course of Phosphoenolpyruvate Phosphomutase. *Journal of the American Chemical Society*, **111**, 9233-9234.

59. Mcqueney, M. S., Lee, S. L., Swartz, W. H., Ammon, H. L., Mariano, P. S. and Dunaway-mariano, D. (1991) Evidence for an Intramolecular, Stepwise Reaction Pathway for Pep Phosphomutase Catalyzed P-C Bond Formation. *Journal of Organic Chemistry*, **56**, 7121-7130.

60. Ntai, I., Phelan, V. V. and Bachmann, B. O. (2006) Phosphonopeptide K-26 biosynthetic intermediates in Astrosporangium hypotensionis. *Chemical Communications*, 4518-4520.

61. Puvvada, M. S., Forrow, S. A., Hartley, J. A., Stephenson, P., Gibson, I., Jenkins, T. C. and Thurston, D. E. (1997) Inhibition of bacteriophage T7 RNA polymerase in vitro transcription by DNA-binding pyrrolo[2,1-c][1,4]benzodiazepines. *Biochemistry*, **36**, 2478-2484.

62. Cummings, M. D., Schubert, C., Parks, D. J., Calvo, R. R., LaFrance, L. V., Lattanze, J., Milkiewicz, K. L. and Lu, T. (2006) Substituted 1,4-benzodiazepine-2,5-diones as alpha-helix mimetic antagonists of the HDM2-p53 protein-protein interaction. *Chemistry and Biolology of Drug Design*, **67**, 201-205.

63. Demartino, G., Massa, S., Corelli, F., Pantaleoni, G., Fanini, D. and Palumbo, G. (1983) Cns Agents - Neuropsychopharmacological Effects of 5h-Pyrrolo[2,1-C] [1,4]Benzodiazepine Derivatives. *European Journal of Medicinal Chemistry*, **18**, 347-350.

64. Joseph, C. G., Wilson, K. R., Wood, M. S., Sorenson, N. B., Phan, D. V., Xiang, Z., Witek, R. M. and Haskell-Luevano, C. (2008) The 1,4-benzodiazepine-2,5-dione small molecule template results in melanocortin receptor agonists with nanomolar potencies. *Journal of Medicinal Chemistry*, **51**, 1423-1431.

65. Karp, G. M., Manfredi, M. C., Guaciaro, M. A., Ortlip, C. L., Marc, P. and Szamosi, I. T. (1997) Synthesis and herbicidal activity of 1H-1,4-benzodiazepine-2,5-diones. *Journal of Agricultural and Food Chemistry*, **45**, 493-500.

66. Wright, W. B., Jr., Brabander, H. J., Greenblatt, E. N., Day, I. P. and Hardy, R. A., Jr. (1978) Derivatives of 1,2,3,11a-tetrahydro-5H-pyrrolo[2,1-c][1,4]benzodiazepine-5,11(10H)-dione as anxiolytic agents. *Journal of Medicinal Chemistry*, **21**, 1087-1089.

67. Heguy, A., Cai, P., Meyn, P., Houck, D., Russo, S., Michitsch, R., Pearce, C., Katz, B., Bringmann, G., Feineis, D., Taylor, D. L. and Tyms, A. S. (1998) Isolation and characterization of the fungal metabolite 3-O-methylviridicatin as an inhibitor of tumour necrosis factor alpha-induced human immunodeficiency virus replication. *Antiviral Agents and Chemotherapy*, **9**, 149-155.

68. Gourdeau, H., McAlpine, J. B., Ranger, M., Simard, B., Berger, F., Beaudry, F., Farnet, C. M. and Falardeau, P. (2008) Identification, characterization and potent antitumor activity of ECO-4601, a novel peripheral benzodiazepine receptor ligand. *Cancer Chemotherary and Pharmacology*, **61**, 911-921.

69. McAlpine, J. B., Bachmann, B. O., Piraee, M., Tremblay, S., Alarco, A. M., Zazopoulos, E. and Farnet, C. M. (2005) Microbial Genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example. *Journal of Natural Products*, **68**, 493-496.

70. Kumar, R. and Lown, J. W. (2003) Recent developments in novel pyrrolo[2,1-c][1,4]benzodiazepine conjugates: synthesis and biological evaluation. *Mini Reviews in Medicinal Chemistry*, **3**, 323-339.

71. Pepper, C. J., Hambly, R. M., Fegan, C. D., Delavault, P. and Thurston, D. E. (2004) The novel sequence-specific DNA cross-linking agent SJG-136 (NSC 694501) has potent and selective in vitro cytotoxicity in human B-cell chronic lymphocytic leukemia cells with evidence of a p53-independent mechanism of cell kill. *Cancer Research*, **64**, 6750-6755.

72. Thurston, D. E., Bose, D. S., Howard, P. W., Jenkins, T. C., Leoni, A., Baraldi, P. G., Guiotto, A., Cacciari, B., Kelland, L. R., Foloppe, M. P. and Rault, S. (1999) Effect of A-ring modifications on the DNA-binding behavior and cytotoxicity of pyrrolo[2,1-c][1,4]benzodiazepines. *Journal of Medicinal Chemistry*, **42**, 1951-1964.

73. Petrusek, R. L., Anderson, G. L., Garner, T. F., Fannin, Q. L., Kaplan, D. J., Zimmer, S. G. and Hurley, L. H. (1981) Pyrrol[1,4]benzodiazepine antibiotics. Proposed structures and characteristics of the in vitro deoxyribonucleic acid adducts of anthramycin, tomaymycin, sibiromycin, and neothramycins A and B. *Biochemistry*, **20**, 1111-1119.

74. Petrusek, R. L., Uhlenhopp, E. L., Duteau, N. and Hurley, L. H. (1982) Reaction of Anthramycin with DNA - Biological Consequences of DNA Damage in Normal and Xeroderma Pigmentosum Cell-Lines. *Journal of Biological Chemistry*, **257**, 6207-6216.

75. Hurley, L. H. (1977) Pyrrolo(1,4)benzodiazepine antitumor antibiotics. Comparative aspects of anthramycin, tomaymycin and sibiromycin. *Journal of Antibiotics (Tokyo)*, **30**, 349-370.

76. Korman, S. and Tendler, M. D. (1965) Clinical investigation of cancer chemotherapeutic agents for neoplastic disease. *Journal of New Drugs*, **5**, 275-285.

77. Cargill, C., Bachmann, E. and Zbinden, G. (1974) Effects of daunomycin and anthramycin on electrocardiogram and mitochondrial metabolism of the rat heart. *Journal of the National Cancer Institute*, **53**, 481-486.

78. Lubawy, W. C., Dallam, R. A. and Hurley, L. H. (1980) Protection against Anthramycin-Induced Toxicity in Mice by Coenzyme-Q10. *Journal of the National Cancer Institute*, **64**, 105-109.

79. Kohn, K. W. and Spears, C. L. (1970) Reaction of Anthramycin with Deoxyribonucleic Acid. *Journal of Molecular Biology*, **51**, 551-&.

80. Kohn, K. W., Bono, V. H. and Kann, H. E. (1968) Anthramycin a New Type of DNA-Inhibiting Antibiotic - Reaction with DNA and Effect on Nucleic Acid Synthesis in Mouse Leukemia Cells. *Biochimica Et Biophysica Acta*, **155**, 121-&.

81. Kohn, K. W., Glaubige.D and Spears, C. L. (1974) Reaction of Anthramycin with DNA .2. Studies of Kinetics and Mechanism. *Biochimica Et Biophysica Acta*, **361**, 288-302.

82. Barkley, M. D., Cheatham, S., Thurston, D. E. and Hurley, L. H. (1986) Pyrrolo[1,4]benzodiazepine antitumor antibiotics: evidence for two forms of tomaymycin bound to DNA. *Biochemistry*, **25**, 3021-3031.

83. Boyd, F. L., Stewart, D., Remers, W. A., Barkley, M. D. and Hurley, L. H. (1990) Characterization of a Unique Tomaymycin-D(Cicgaattcicg)2 Adduct Containing 2 Drug Molecules Per Duplex by Nmr, Fluorescence, and Molecular Modeling Studies. *Biochemistry*, **29**, 2387-2403.

84. Graves, D. E., Pattaroni, C., Krishnan, B. S., Ostrander, J. M., Hurley, L. H. and Krugh, T. R. (1984) The Reaction of Anthramycin with DNA - Proton and Carbon Nuclear Magnetic-Resonance Studies on the Structure of the Anthramycin-DNA Adduct. *Journal of Biological Chemistry*, **259**, 8202-8209.

85. Hertzberg, R. P., Hecht, S. M., Reynolds, V. L., Molineux, I. J. and Hurley, L. H. (1986) DNA-Sequence Specificity of the Pyrrolo[1,4]Benzodiazepine Antitumor Antibiotics - Methidiumpropyl-Edta-Iron(Ii) Footprinting Analysis of DNA-Binding Sites for Anthramycin and Related Drugs. *Biochemistry*, **25**, 1249-1258.

86. Hurley, L. H. and Petrusek, R. (1979) Proposed Structure of the Anthramycin-DNA Adduct. *Nature*, **282**, 529-531.

87. Kopka, M. L., Goodsell, D. S., Baikalov, I., Grzeskowiak, K., Cascio, D. and Dickerson, R. E. (1994) Crystal structure of a covalent DNA-drug adduct: anthramycin bound to C-C-A-A-C-G-T-T-G-G and a molecular explanation of specificity. *Biochemistry*, **33**, 13593-13610.

88. Krugh, T. R., Graves, D. E. and Stone, M. P. (1989) 2-Dimensional Nmr-Studies on the Anthramycin-D(Atgcat)2 Adduct. *Biochemistry*, **28**, 9988-9994.

89. Kizu, R., Draves, P. H. and Hurley, L. H. (1993) Correlation of DNA-Sequence Specificity of Anthramycin and Tomaymycin with Reaction-Kinetics and Bending of DNA. *Biochemistry*, **32**, 8712-8722.

90. Konrad, F., Schwalbe, B., Heeg, K., Wagner, H., Wiedeck, H., Kilian, J. and Ahnefeld, F. W. (1989) Frequency of Bacterial-Colonization and Respiratory-Tract Infections and Resistance Behavior in Patient Subjected to Long-Term Ventilation Patients with Selective Decontamination of the Digestive Tract. *Anaesthesist*, **38**, 99-109.

91. Arora, S. K. (1979) Structural Investigations of Mode of Action of Drugs.1. Molecular-Structure of Mitomycin-C. *Life Sciences*, **24**, 1519-1526.

92. Hurley, L. H. (1977) Pyrrolo(1,4)Benzodiazepine Antitumor Antibiotics - Comparative Aspects of Anthramycin, Tomaymycin and Sibiromycin. *Journal of Antibiotics*, **30**, 349-370.

93. Hurley, L. H., Gairola, C. and Zmijewski, M. (1977) Pyrrolo(1,4)Benzodiazepine Antitumor Antibiotics - Invitro Interaction of Anthramycin, Sibiromycin and Tomaymycin with DNA Using Specifically Radiolabeled Molecules. *Biochimica Et Biophysica Acta*, **475**, 521-535.

94. Hurley, L. H., Reck, T., Thurston, D. E., Langley, D. R., Holden, K. G., Hertzberg, R. P., Hoover, J. R., Gallagher, G., Jr., Faucette, L. F., Mong, S. M. and et al. (1988) Pyrrolo[1,4]benzodiazepine antitumor antibiotics: relationship of DNA alkylation and sequence specificity to the biological activity of natural and synthetic compounds. *Chemical Research in Toxicology*, **1**, 258-268.

95. Vargiu, A. V., Ruggerone, P., Magistrato, A. and Carloni, P. (2006) Anthramycin-DNA binding explored by molecular simulations. *Journal of Physical Chemistry B*, **110**, 24687-24695.

96. Boyd, F. L., Cheatham, S. F., Remers, W., Hill, G. C. and Hurley, L. H. (1990) Characterization of the Structure of the Anthramycin-D(Atgcat)2 Adduct by Nmr and Molecular Modeling Studies - Determination of the Stereochemistry of the Covalent Linkage Site, Orientation in the Minor Groove of DNA, and Effect on Local DNA-Structure. *Journal of the American Chemical Society*, **112**, 3279-3289.

97. Hurley, L. H., Zmijewski, M. and Chang, C. J. (1975) Biosynthesis of Anthramycin - Determination of Labeling Pattern by Use of Radioactive and Stable Isotope Techniques. *Journal of the American Chemical Society*, **97**, 4372-4378.

98. Hurley, L. H. and Gairola, C. (1979) Pyrrolo (1,4) benzodiazepine antitumor antibiotics: Biosynthetic studies on the conversion of tryptophan to the anthranilic acid moieties of sibiromycin and tomaymycin. *Antimicrobial Agents and Chemotherapy*, **15**, 42-45.

99. Hurley, L. H., Lasswell, W. L., Malhotra, R. K. and Das, N. V. (1979) Pyrrolo[1,4]benzodiazepine antibiotics. Biosynthesis of the antitumor antibiotic sibiromycin by Streptosporangium sibiricum. *Biochemistry*, **18**, 4225-4229.

100. Hurley, L. H., Zmijewski, M. and Chang, C. J. (1975) Biosynthesis of anthramycin. Determination of the labeling pattern by the use of radioactive and stable isotope techniques. *Journal of the American Chemical Society*, **97**, 4372-4378.

101. Hurley, L. H. (1980) Elucidation and Formulation of Novel Biosynthetic Pathways Leading to the Pyrrolo[1,4]Benzodiazepine Antibiotics Anthramycin, Tomaymycin, and Sibiromycin. *Accounts of Chemical Research*, **13**, 263-269.

102. Brahme, N. M., Gonzalez, J. E., Rolls, J. P., Hessler, E. J., Mizsak, S. and Hurley, L. H. (1984) Biosynthesis of the Lincomycins .1. Studies Using Stable Isotopes on the Biosynthesis of the Propyl-L-Hygric and Ethyl-L-Hygric Acid Moieties of Lincomycin-a and Lincomycin-B. *Journal of the American Chemical Society*, **106**, 7873-7878.

103. Kuo, M. S., Yurek, D. A., Coats, J. H., Chung, S. T. and Li, G. P. (1992) Isolation and Identification of 3-Propylidene-Delta(1)-Pyrroline-5-Carboxylic Acid, a Biosynthetic Precursor of Lincomycin. *Journal of Antibiotics*, **45**, 1773-1777.

104. Peschke, U., Schmidt, H., Zhang, H. Z. and Piepersberg, W. (1995) Molecular Characterization of the Lincomycin-Production Gene-Cluster of Streptomyces Lincolnensis-78-11. *Molecular Microbiology*, **16**, 1137-1156.

105. Neusser, D., Schmidt, H., Spizek, J., Novotna, J., Peschke, U., Kaschabeck, S., Tichy, P. and Piepersberg, W. (1998) The genes lmbB1 and lmbB2 of Streptomyces lincolnensis encode enzymes involved in the conversion of L-tyrosine to propylproline during the biosynthesis of the antibiotic lincomycin A. *Archives of Microbiology*, **169**, 322-332.

106. Harayama, S. and Rekik, M. (1989) Bacterial Aromatic Ring-Cleavage Enzymes Are Classified into 2 Different Gene Families. *Journal of Biological Chemistry*, **264**, 15328-15333.

107. Novotna, J., Honzatko, A., Bednar, P., Kopecky, J., Janata, J. and Spizek, J. (2004) L-3,4-Dihydroxyphenyl alanine-extradiol cleavage is followed by intramolecular cyclization in lincomycin biosynthesis. *European Journal of Biochemistry*, **271**, 3678-3683.

108. Colabroy, K. L., Hackett, W. T., Markham, A. J., Rosenberg, J., Cohen, D. E. and Jacobson, A. (2008) Biochemical characterization of L-DOPA 2,3-dioxygenase, a single-domain type I extradiol dioxygenase from lincomycin biosynthesis. *Archives of Biochemistry and Biophysics*, **479**, 131-138.

109. Caboche, S., Pupin, M., Leclere, V., Fontaine, A., Jacques, P. and Kucherov, G. (2008) NORINE: a database of nonribosomal peptides. *Nucleic Acids Research*, **36**, D326-331.

110. Fischbach, M. A. and Walsh, C. T. (2006) Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chemical Reviews*, **106**, 3468-3496.

111. Marahiel, M. A. (1997) Protein templates for the biosynthesis of peptide antibiotics. *Chemistry & Biology*, **4**, 561-567.

112. Mootz, H. D., Schwarzer, D. and Marahiel, M. A. (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem*, **3**, 490-504.

113. Schwarzer, D. and Marahiel, M. A. (2001) Multimodular biocatalysts for natural product assembly. *Naturwissenschaften*, **88**, 93-101.

114. vonDohren, H., Keller, U., Vater, J. and Zocher, R. (1997) Multifunctional peptide synthetases. *Chemical Reviews*, **97**, 2675-2705.

115. Dieckmann, R., Lee, Y. O., Vanliempt, H., Vondohren, H. and Kleinkauf, H. (1995) Expression of an Active Adenylate-Forming Domain of Peptide Synthetases Corresponding to Acyl-Coa-Synthetases. *Febs Letters*, **357**, 212-216.

116. Challis, G. L., Ravel, J. and Townsend, C. A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry & Biology*, **7**, 211-224.

117. Stachelhaus, T., Mootz, H. D. and Marahiel, M. A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology*, **6**, 493-505.

118. Du, L. C., Sanchez, C., Chen, M., Edwards, D. J. and Shen, B. (2000) The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chemistry & Biology*, **7**, 623-642.

119. Ehmann, D. E., Shaw-Reid, C. A., Losey, H. C. and Walsh, C. T. (2000) The EntF and EntE adenylation domains of Escherichia coli enterobactin synthetase: Sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 2509-2514.

120. Stachelhaus, T., Huser, A. and Marahiel, M. A. (1996) Biochemical characterization of peptides carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chemistry & Biology*, **3**, 913-921.

121. Stachelhaus, T., Mootz, H. D., Bergendahl, V. and Marahiel, M. A. (1998) Peptide bond formation in nonribosomal peptide biosynthesis - Catalytic role of the condensation domain. *Journal of Biological Chemistry*, **273**, 22773-22781.

122. Bergendahl, V., Linne, U. and Marahiel, M. A. (2002) Mutational analysis of the C-domain in nonribosomal peptide synthesis. *European Journal of Biochemistry*, **269**, 620-629.

123. Keating, T. A., Marshall, C. G., Walsh, C. T. and Keating, A. E. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nature Structural Biology*, **9**, 522-526.

124. Stein, T., Vater, J., Kruft, V., Otto, A., WittmannLiebold, B., Franke, P., Panico, M., McDowell, R. and Morris, H. R. (1996) The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. *Journal of Biological Chemistry*, **271**, 15428-15435.

125. Sieber, S. A. and Marahiel, M. A. (2003) Learning from nature's drug factories: Nonribosomal synthesis of macrocyclic peptides. *Journal of Bacteriology*, **185**, 7036-7043.

126. Schneider, A. and Marahiel, M. A. (1998) Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in Bacillus subtilis. *Archives of Microbiology*, **169**, 404-410.

127. Kohli, R. M., Trauger, J. W., Schwarzer, D., Marahiel, M. A. and Walsh, C. T. (2001) Generality of peptide cyclization catalyzed by isolated thioesterase domains of nonribosomal peptide synthetases. *Biochemistry*, **40**, 7099-7108.

128. Miller, D. A., Luo, L. S., Hillson, N., Keating, T. A. and Walsh, C. T. (2002) Yersiniabactin synthetase: A four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis. *Chemistry & Biology*, **9**, 333-344.

129. Walsh, C. T., Chen, H. W., Keating, T. A., Hubbard, B. K., Losey, H. C., Luo, L. S., Marshall, C. G., Miller, D. A. and Patel, H. M. (2001) Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Current Opinion in Chemical Biology*, **5**, 525-534.

130. Guenzi, E., Galli, G., Grgurina, I., Pace, E., Ferranti, P. and Grandi, G. (1998) Coordinate transcription and physical linkage of domains in surfactin synthetase are not essential for proper assembly and activity of the multienzyme complex. *Journal of Biological Chemistry*, **273**, 14403-14410.

131. Keating, T. A. and Walsh, C. T. (1999) Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Current Opinion in Chemical Biology*, **3**, 598-606.

132. Marahiel, M. A., Stachelhaus, T. and Mootz, H. D. (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chemical Reviews*, **97**, 2651-2673.

133. Peypoux, F., Bonmatin, J. M. and Wallach, J. (1999) Recent trends in the biochemistry of surfactin. *Applied Microbiology and Biotechnology*, **51**, 553-563.

134. Gehring, A. M., Mori, I. and Walsh, C. T. (1998) Reconstitution and characterization of the Escherichia coli enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry*, **37**, 2648-2659.

135. Shaw-Reid, C. A., Kelleher, N. L., Losey, H. C., Gehring, A. M., Berg, C. and Walsh, C. T. (1999) Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E-coli EntF catalyzes both elongation and cyclolactonization. *Chemistry & Biology*, **6**, 385-400.

136. Brown, L. C. W., Acker, M. G., Clardy, J., Walsh, C. T. and Fischbach, M. A. (2009) Thirteen posttranslational modifications convert a 14-residue peptide into the antibiotic thiocillin. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 2549-2553.

137. Cotter, P. D., Draper, L. A., Lawton, E. M., Daly, K. M., Groeger, D. S., Casey, P. G., Ross, R. P. and Hill, C. (2008) Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I Listeria monocytogenes. *Plos Pathogens*, **4**, -.

138. Hallen, H. E., Luo, H., Scott-Craig, J. S. and Walton, J. D. (2007) Gene family encoding the major toxins of lethal Amanita mushrooms. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 19097-19101.

139. Kelly, W. L., Pan, L. and Li, C. X. (2009) Thiostrepton Biosynthesis: Prototype for a New Family of Bacteriocins. *Journal of the American Chemical Society*, **131**, 4327-4334.

140. Lawton, E. M., Cotter, P. D., Hill, C. and Ross, R. P. (2007) Identification of a novel two-peptide lantibiotic, Haloduracin, produced by the alkaliphile Bacillus halodurans C-125. *FEMS Microbiology Letters*, **267**, 64-71.

141. Lee, S. W., Mitchell, D. A., Markley, A. L., Hensler, M. E., Gonzalez, D., Wohlrab, A., Dorrestein, P. C., Nizet, V. and Dixon, J. E. (2008) Discovery of a widely distributed toxin biosynthetic gene cluster. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 5879-5884.

142. Liao, R., Duan, L., Lei, C., Pan, H., Ding, Y., Zhang, Q., Chen, D., Shen, B., Yu, Y. and Liu, W. (2009) Thiopeptide biosynthesis featuring ribosomally synthesized precursor peptides and conserved posttranslational modifications. *Chemistry & Biology*, **16**, 141-147.

143. McClerren, A. L., Cooper, L. E., Quan, C., Thomas, P. M., Kelleher, N. L. and van der Donk, W. A. (2006) Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 17243-17248.

144. Morris, R. P., Leeds, J. A., Naegeli, H. U., Oberer, L., Memmert, K., Weber, E., LaMarche, M. J., Parker, C. N., Burrer, N., Esterow, S., Hein, A. E., Schmitt, E. K. and Krastel, P. (2009) Ribosomally synthesized thiopeptide antibiotics targeting elongation factor Tu. *Journal of the American Chemical Society*, **131**, 5946-5955.

145. Philmus, B., Christiansen, G., Yoshida, W. Y. and Hemscheidt, T. K. (2008) Post-translational Modification in Microviridin Biosynthesis. *Chembiochem*, **9**, 3066-3073.

146. 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. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 7315-7320.

147. Ziemert, N., Ishida, K., Liaimer, A., Hertweck, C. and Dittmann, E. (2008) Ribosomal synthesis of tricyclic depsipeptides in bloom-forming cyanobacteria. *Angewandte Chemie-International Edition*, **47**, 7756-7759.

148. Eriani, G., Delarue, M., Poch, O., Gangloff, J. and Moras, D. (1990) Partition of Transfer-Rna Synthetases into 2 Classes Based on Mutually Exclusive Sets of Sequence Motifs. *Nature*, **347**, 203-206.

149. Sankaranarayanan, R., Dock-Bregeon, A. C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B. and Moras, D. (1999) The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell*, **97**, 371-381.

150. Ogle, J. M., Carter, A. P. and Ramakrishnan, V. (2003) Insights into the decoding mechanism from recent ribosome structures. *Trends in Biochemical Sciences*, **28**, 259-266.

151. Oman, T. J. and van der Donk, W. A. (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nature Chemical Biology*, **6**, 9-18.

152. Lubelski, J., Rink, R., Khusainov, R., Moll, G. N. and Kuipers, O. P. (2008) Biosynthesis, immunity, regulation, mode of action and engineering of the model lantibiotic nisin. *Cellular and Molecular Life Sciences*, **65**, 455-476.

153. Li, B., Yu, J. P., Brunzelle, J. S., Moll, G. N., van der Donk, W. A. and Nair, S. K. (2006) Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. *Science*, **311**, 1464-1467.

154. van der Meer, J. R., Rollema, H. S., Siezen, R. J., Beerthuyzen, M. M., Kuipers, O. P. and de Vos, W. M. (1994) Influence of amino acid substitutions in the nisin leader peptide on biosynthesis and secretion of nisin by Lactococcus lactis. *The Journal of Biological Chemistry*, **269**, 3555-3562.

155. Havarstein, L. S., Holo, H. and Nes, I. F. (1994) The leader peptide of colicin V shares consensus sequences with leader peptides that are common among peptide bacteriocins produced by gram-positive bacteria. *Microbiology*, **140 (Pt 9)**, 2383-2389.

156. Chen, P., Qi, F. X., Novak, J., Krull, R. E. and Caufield, P. W. (2001) Effect of amino acid substitutions in conserved residues in the leader peptide on biosynthesis of the lantibiotic mutacin II. *FEMS Microbiology Letters*, **195**, 139-144.

157. Furgerson Ihnken, L. A., Chatterjee, C. and van der Donk, W. A. (2008) In vitro reconstitution and substrate specificity of a lantibiotic protease. *Biochemistry*, **47**, 7352-7363.

158. 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 synthetase activity. *Science*, **303**, 679-681.

159. Willey, J. M. and van der Donk, W. A. (2007) Lantibiotics: peptides of diverse structure and function. *Annual Review of Microbiology*, **61**, 477-501.

160. Duquesne, S., Destoumieux-Garzon, D., Peduzzi, J. and Rebuffat, S. (2007) Microcins, gene-encoded antibacterial peptides from enterobacteria. *Natural Products Reports*, **24**, 708-734.

161. Bayer, A., Freund, S., Nicholson, G. and Jung, G. (1993) Posttranslational Backbone Modifications in the Ribosomal Biosynthesis of the Glycine-Rich Antibiotic Microcin-B17. *Angewandte Chemie-International Edition in English*, **32**, 1336-1339.

162. Bayro, M. J., Mukhopadhyay, J., Swapna, G. V. T., Huang, J. Y., Ma, L. C., Sineva, E., Dawson, P. E., Montelione, G. T. and Ebright, R. H. (2003) Structure of antibacterial peptide microcin J25: A 21-residue lariat protoknot. *Journal of the American Chemical Society*, **125**, 12382-12383.

163. Rosengren, K. J., Clark, R. J., Daly, N. L., Goransson, U., Jones, A. and Craik, D. J. (2003) Microcin J25 has a threaded sidechain-to-backbone ring structure and not a head-to-tail cyclized backbone. *Journal of the American Chemical Society*, **125**, 12464-12474.

164. Wilson, K. A., Kalkum, M., Ottesen, J., Yuzenkova, J., Chait, B. T., Landick, R., Muir, T., Severinov, K. and Darst, S. A. (2003) Structure of microcin J25, a peptide inhibitor of bacterial RNA polymerase, is a lassoed tail. *Journal of the American Chemical Society*, **125**, 12475-12483.

165. 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.

166. 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.

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

168. Duquesne, S., Destoumieux-Garzon, D., Zirah, S., Goulard, C., Peduzzi, J. and Rebuffat, S. (2007) Two enzymes catalyze the maturation of a lasso peptide in Escherichia coli. *Chemistry & Biology*, **14**, 793-803.

169. Lee, J., McIntosh, J., Hathaway, B. J. and Schmidt, E. W. (2009) Using Marine Natural Products to Discover a Protease that Catalyzes Peptide Macrocyclization of Diverse Substrates. *Journal of the American Chemical Society*, **131**, 2122-+.

170. Bagley, M. C., Dale, J. W., Merritt, E. A. and Xiong, A. (2005) Thiopeptide antibiotics. *Chemical Reviews*, **105**, 685-714.

171. Funabashi, M., Yang, Z. Y., Nonaka, K., Hosobuchi, M., Fujita, Y., Shibata, T., Chi, X. L. and Van Lanen, S. G. (2010) An ATP-independent strategy for amide bond formation in antibiotic biosynthesis. *Nature Chemical Biology*, **6**, 581-586.

172. Hollenhorst, M. A., Clardy, J. and Walsh, C. T. (2009) The ATP-dependent amide ligases DdaG and DdaF assemble the fumaramoyl-dipeptide scaffold of the dapdiamide antibiotics. *Biochemistry*, **48**, 10467-10472.

173. Kadi, N., Oves-Costales, D., Barona-Gomez, F. and Challis, G. L. (2007) A new family of ATP-dependent oligomerization-macrocyclization biocatalysts. *Nature Chemical Biology*, **3**, 652-656.

174. Oves-Costales, D., Kadi, N. and Challis, G. L. (2009) The long-overlooked enzymology of a nonribosomal peptide synthetase-independent pathway for virulence-conferring siderophore biosynthesis. *Chemical Communications*, 6530-6541.

175. Mocibob, M., Ivic, N., Bilokapic, S., Maier, T., Luic, M., Ban, N. and Weygand-Durasevic, I. (2010) Homologs of aminoacyl-tRNA synthetases acylate carrier proteins and provide a link between ribosomal and nonribosomal peptide synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 14585-14590.

176. Mooz, E. D. and Meister, A. (1967) Tripeptide (glutathione) synthetase. Purification, properties, and mechanism of action. *Biochemistry*, **6**, 1722-1734.

177. Walsh, C. T. (1989) Enzymes in the D-Alanine Branch of Bacterial-Cell Wall Peptidoglycan Assembly. *Journal of Biological Chemistry*, **264**, 2393-2396.

CHAPTER II

γ-$^{18}O_4$-ATP PYROPHOSPHATE EXCHANGE ASSAY

Most peptide natural products, especially those containing non-proteinogenic amino acids, such as those proposed in the biosynthesis of K-26 (AHEP) and anthramycin (MHA and (*E*)-4-(3-amino-3-oxoprop-1-enyl)-2,3-dihydro-1H-pyrrole-2-carboxylic acid), are formed via the NRPS paradigm.[1] In order to investigate the biosynthesis of the peptide natural products K-26 and anthramycin, we required an assay to test for amino acid activation necessary for peptide bond formation in NRPS based peptide biosynthesis. First developed by Lee and Lipmann in 1975, the pyrophosphate exchange assay takes advantage of the non-catalytic reversible nature of isolated synthetases (both tRNA synthetases and NRPS A-domains) to reveal the substate specificity of the synthetase of interest.[2] These isolated synthetases perform half reactions of cognate amino acid with ATP creating tightly bound aminoacyl adenylates releasing pyrophosphate. Subsequent amino acid esterification to the respective tRNA or T-domain prepares the amino acid substrate for peptide bond formation (Figure II-1).

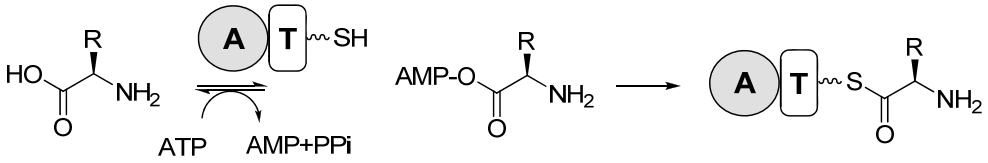


Figure II-1. Role of A-domains in NRPS. NRPS A-domains catalyze the reversible formation of aminoacyl adenylates for subsequent thiolation reactions to tether the amino acid on T-domains. The reversible nature of aminoacyl adenylate formation can be exploited to identify the amino acid substrate of the A-domain.

Classically, pyrophosphate exchange activity is assayed using the ATP-^{32}PPi isotope exchange assay where the synthetase of interest is incubated with an excess of ATP, amino acid and ^{32}PPi and the reversible back-exchange of labeled ^{32}PPi into ATP is monitored.[2, 3] In order to monitor the exchange activity, the ATP is captured on activated charcoal and the amount of radioactivity incorporated is determined by scintillation counting. This assay has provided a highly sensitive assay for identifying the substrate specificity of a variety of synthetases and has been developed for high-throughput screening and kinetic measurements.[4] However, it is limited by the relatively high amount of radioactive substrate used in each experiment, extensive liquid handling of radioactive materials and speed of analysis. In order to circumvent the use of radioactive materials and to increase the speed of analysis, we developed a mass-based pyrophosphate exchange assay allowing for direct exchange analysis and reduction of background interference. To validate this method, we tested a panel of previously characterized synthetases including TycA, an A-domain responsible for L-phenylalanine activation during tyrocidine biosynthesis; ValA, an orphan A-domain responsible for valine activation and tRNA synthetases for tryptophan (TrpRS) and lysine (LysRS).[5-8]

Results

Hypothetically, mass spectrometry (MS) based pyrophosphate exchange can detect mass shifts on either side of the exchange equation. We decided to introduce a heavy-atom label into the ATP starting material as γ-$^{18}O_4$-ATP based on its commercial availablility and literature precedent for its chemical synthesis.[9] An additional advantage of incorporating the label into ATP is that back exchange is favored by incubating the synthetase with an excess of unlabeled pyrophosphate, reducing the amount of label required for each reaction. Conditions for this assay were based on previously reported radioactive assay methods and simultaneously developed for both MALDI-TOF and ESI-LC mass spectrometry analyses.[4, 10] Adenylation enzymes (200 nM) were incubated with 1 mM γ-$^{18}O_4$-ATP, 1 mM amino acid, 5 mM $MgCl_2$ and 5 mM pyrophosphate. For MALDI-TOF MS analysis, reactions were quenched by mixing with an equal volume of 9-aminoacridine in acetone, identified as the optimal matrix for the detection of ATP in negative mode.[11] For ESI-LC/MS analysis, reactions were quenched with an equal volume of acetone and separated from contaminating salts by using a Hypercarb graphitic matrix column (Thermo, Inc.), which dependably retains charged metabolites by charge quadrupole and hydrophobic interactions.[12, 13] ATP was eluted from the column with an isocratic gradient of 17.5% ACN in 20 mM ammonium acetate buffer (pH 6) and detected in negative ion mode.

Observation of the 8 Da mass shift, caused by back exchange of unlabeled pyrophosphate leading to the formation of γ-$^{16}O_4$-ATP and

consumption by γ-$^{18}O_4$-ATP, allows for the quantification of enzyme activity. (Figure II-2) The enzyme activity was quantified as the integrated peak ratio of γ-$^{16}O_4$-ATP to all ATP species in the reaction mixture. As the species being compared in the reaction mixture are isotopologues of ATP, peak integration and the subsequent signal ratio is quantitative in mass spectrometric analysis, as there should be no difference in ionization efficiency due to differences in chemical composition. The observed mass ratios correlated to the fraction of ATP-pyrophosphate exchange activity. Using this strategy, 6 μL reaction incubations can be analyzed via MALDI-TOF MS in as little as 30 seconds and, with higher sensitivity, ESI-LC/MS in 5 minutes.

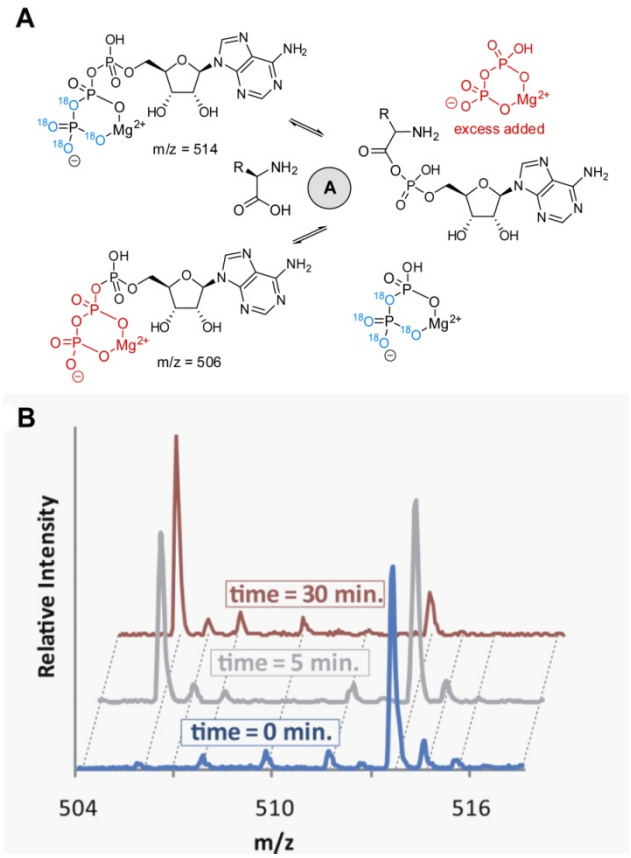


Figure II-2. ATP-PP$_i$ exchange. A: The exchange reaction, performed in the absence of thiolation activity, measures equilibrium exchange of γ-$^{18}O_4$-ATP with $^{16}O_4$-pyrophosphate. B: Time dependent formation of γ-$^{16}O_4$-ATP and disappearance of γ-$^{18}O_4$-ATP with TycA, measured by MALDI-TOFMS. Intermediary masses correspond to $^{18}O_3$, $^{18}O_2$, and $^{18}O_1$ peaks.[14]

In order to show the utility of the mass based pyrophosphate exchange assay, we determined the limit of detection (LOD) for both MALDI-TOF and ESI-LC mass spectrometry (MS) analysis. (Figure II-3) Known amounts of labeled and unlabeled ATP were added to TycA reaction mixtures with no amino acid present. For ESI-LC/MS, the LOD was determined to be 3.4 μM (0.34% exchange), and for MALDI-TOF MS, a higher LOD of 10 μM (1% exchange) was observed. A possible source for the higher LOD in MALDI-TOF MS samples is the ion suppression effects in MALDI-TOF MS caused by a large amount of salt

within the MALDI samples, which is eliminated in the ESI-LC/MS sample preparation by the use of the Hypercarb column. Background levels of γ-$^{16}O_4$-ATP found within commercial γ-$^{18}O_4$-ATP were estimated to be 3.4 μM by ESI-LC/MS and fall below the LOD threshold for MALDI-TOF MS analysis. Both detection methods showed comparable sensitivity to reported radioactive pyrophosphate exchange assays. Under optimized radioactive conditions, as little as 50 pmol exchange (0.01%) has been detected.[15] With the rapid MALDI-TOF method, 1% (60 pmol) exchange was detected, while with full scan ESI-LC/MS detection 0.1% (6 pmol) exchange was observed and with selected reaction monitoring ESI-LC/MS detection, the sensitivity was enhance by monitoring as little as 0.01% (600 fmol) exchange.

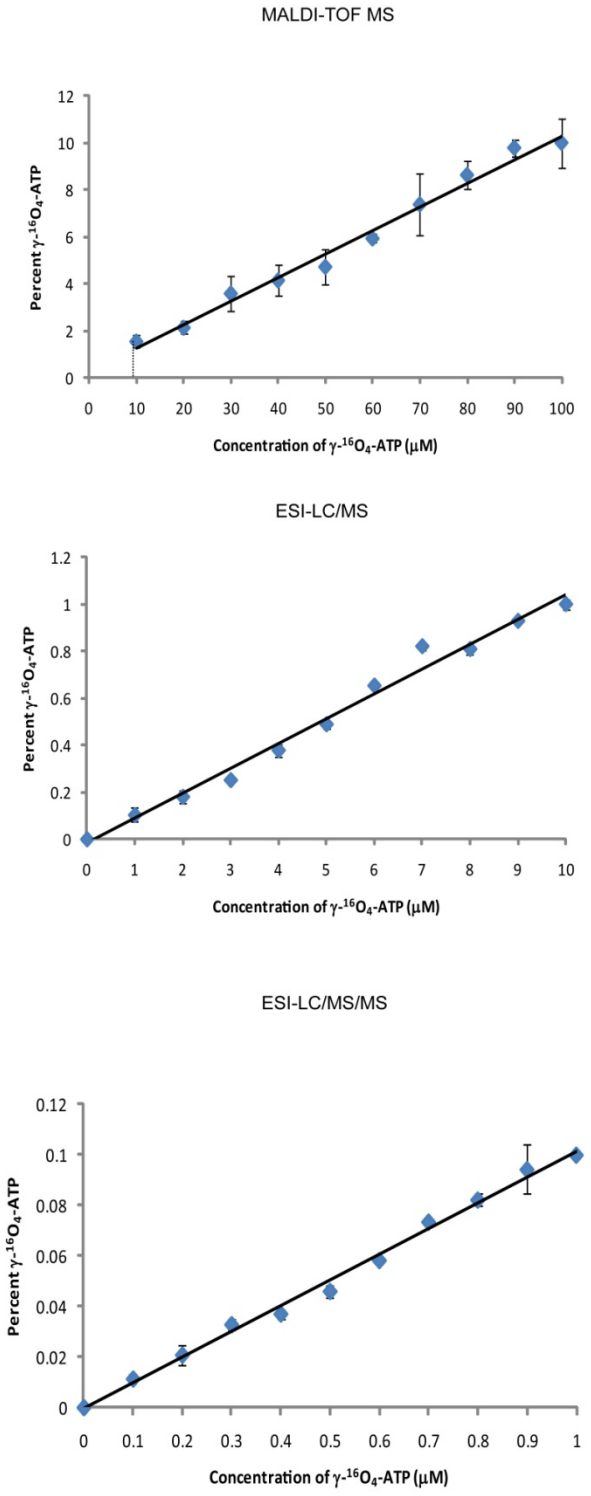


Figure II-3. Limit of Exchange Detection for MALDI-TOFMS, ESI-LC/MS and ESI-LC/MS/MS detection methods. Top: MALDI-TOFMS detection (in triplicate) of γ-$^{16}O_4$-ATP in 1 mM γ-$^{18}O_4$-ATP in reaction mixture demonstrated reliable detection above 1 % exchange. Middle: Similar measurement using ESI-LC/MS detection permitted detection levels above 0.1% exchange. Bottom: Similar measurement using ESI-LC/MS/MS detection permitted detection levels above 0.01% exchange.[14]

As some NRPS A-domains are not particularly soluble or active when heterologously expressed, longer incubation times may be necessary to identify the correct amino acid substrate. Of particular concern in these cases is the hydrolytic stability of γ-$^{18}O_4$-ATP under typical assay conditions. To test this, γ-$^{18}O_4$-ATP was incubated under assay conditions with 1) no enzyme (TycA), 2) no amino acid and 3) incorrect amino acid for up to 15 hours. No loss of label was observed at up to 15 hours in assay buffer in the absence of enzyme, consistent with literature reports showing the relative stability of ^{18}O-substituted phosphates in buffered solutions.[16] (Figure II-4) However, in the presence of enzyme or incorrect amino acid, slow exchange of the ^{18}O labeled was observed after 2-5 hours. Incubation of γ-$^{18}O_4$-ATP with enzyme for 5 hours resulted in a decrease in γ-$^{18}O_4$-ATP of 45 - 60 %, and an increase of partially labeled ATP, whereas only a 14 - 24% increase in the formation of γ-$^{16}O_4$-ATP, corresponding to the loss of the bridging β–γ ^{18}O, was observed under the same conditions. This indicates that non β–γ–bridging ^{18}O atoms exchanged more rapidly than bridging ^{18}O atoms. As pyrophosphate exchange is only indicated by the complete loss of the bridging β–γ ^{18}O, the slow loss of non-bridging labels can be compensated for by calculating the exchange as the ratio of unlabeled ATP divided by the sum of all ATP species normalized to the theoretical equilibrium $^{16}O/^{18}O$ molar ratio (83.33% apparent exchange = 100% exchange). Therefore, % exchange = $(100/0.833) * {}^{16}O/({}^{18}O + {}^{16}O)$.

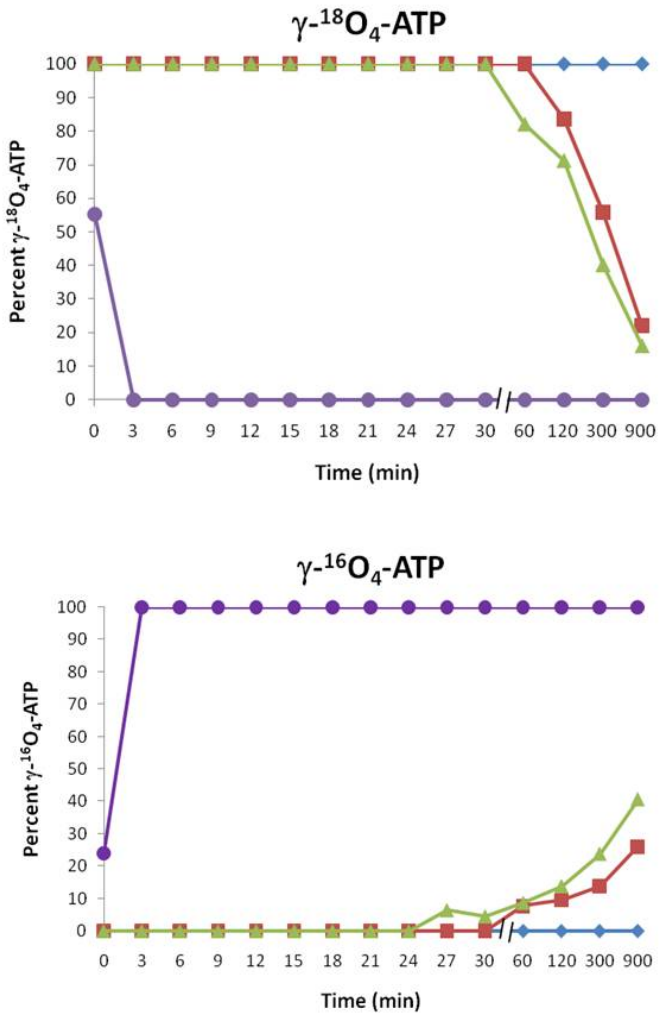


Figure II-4. ^{18}O ATP lability in 15 hour time course with No TycA (◆), no Substrate (■), glutamate (▲) and phenylalanine (●). Over 15 hours γ-$^{18}O_4$-ATP is stable when no enzyme is present. γ-$^{18}O_4$-ATP only becomes significantly depleted after two hours when γ-$^{18}O_4$-ATP is incubated with TycA without amino acid or with incorrect amino acid present. However, γ-$^{16}O_4$-ATP created by loss of all four ^{18}O labels is still only 25% while exchange of 1, 2 or 3 ^{18}O labels accounts for the decrease of γ-$^{18}O_4$-ATP detected. The loss of ^{18}O labels can be corrected for by calculating the ratio of the area of γ-$^{16}O_4$-ATP to the area of total ATP and subtracting the substrate free control.[14]

TycA, from tyrocidine biosynthesis, was chosen as the model synthetase to validate the mass spectrometry based pyrophosphate exchange assay. TycA specificity has been extensively studied by the traditional radioactive pyrophosphate exchange assay.[2, 4] To demonstrate amino acid selectivity, we

tested the full panel of proteinogenic amino acids with the addition of D-phenylalanine, a previously identified substrate of TycA. Both L- and D-phenylalanine were identified as substrates of TycA, showing 70 – 100% exchange using both MALDI-TOF and ESI-LC/MS analysis. (Figure II-5) While lower rates of exchange were effectively measured by ESI-LC/MS analysis, due to the LOD for MALDI-TOF analysis (1% exchange), no exchange was observed for all other amino acids tested. However, the signal could be enhanced by increasing the enzyme concentration and/or incubation times.

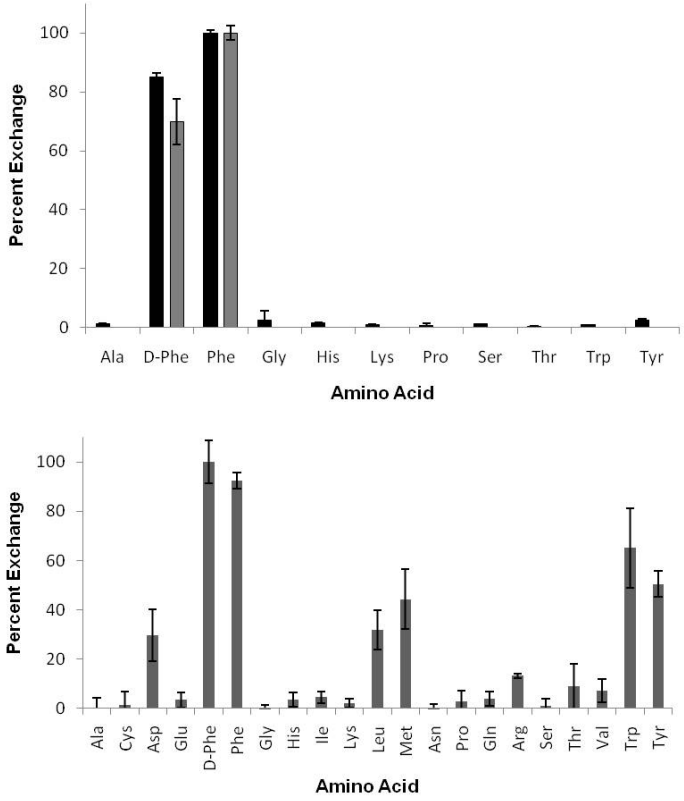


Figure II-525. Amino Acid Activation by TycA. Top: ESI-LC/MS (■) and MALDI-TOFMS (■) detection of TycA substrate activation by ATP-PP_i exchange. Bottom: Enhanced sensitivity of MALDI-TOFMS detection of TycA substrate activation by γ–$^{18}O_4$-ATP-PP_i exchange. Activities were monitored via MALDI-TOFMS by incubating 1μM TycA for two hours in assay reaction mixture.[14]

To further validate the mass-based assay, an additional 'orphan' synthetase, ValA, a valine activating A-domain, identified and verified by Shen and coworkers, was tested.[5] The MALDI-TOF assay demonstrated appreciable levels of exchange for valine in accordance with previous studies. (Table II-1) Again, for all other amino acids tested, levels of exchange fell below the LOD for MALDI-TOF MS analysis, but low levels of exchange were easily detectable using either ESI-LC/MS analysis or longer incubation times with higher enzyme concentrations for the MALDI-TOF method.

Table II-1. Activity of Amino Acid Activating Enzymes[14]

Enzyme	Amino Acid	% γ-$^{18}O_4$-ATP exchange
TycA	L-Phe	100 ±2.5
	D-Phe	69.8 ±7.8
ValA[a]	Val	5.5 ± 0.8
TrpRS[a]	Trp	9.7 ± 1.6
LysRS[a]	Lys	13.0 ± 2.7

[a] Exchange measured by MALDI-TOFMS for a panel of amino acids. Exchange is only listed for active amino acids. All other amino acids tested fell below the threshold of detection for the MALDI-TOFMS based assay.

In addition to the aforementioned NRPS adenylation domains, two previously characterized tRNA synthetases from *E. coli*, TrpRS and LysRS, were also assayed using mass-based pyrophosphate exchange. As shown in Table II-1, TrpRS and LysRS activate their cognate amino acids under standard assay conditions. No activation was observed for non-substrates.

To demonstrate the applicability of the mass-based exchange assay for kinetics measurements, the exchange velocities for TycA were plotted versus

L-phenylalanine concentration. (Figure III-6) The apparent Michaelis-Menten kinetic parameters were calculated for the purpose of comparison to previously reported TycA apparent parameters. ESI-LC/MS analysis yielded an apparent K_M of 67 ± 2 μM and apparent k_{cat} of 92 ± 2 min^{-1}. These values are consistent with previously described measurements which report K_M values between 40 μM to 13 μM.[4, 7]

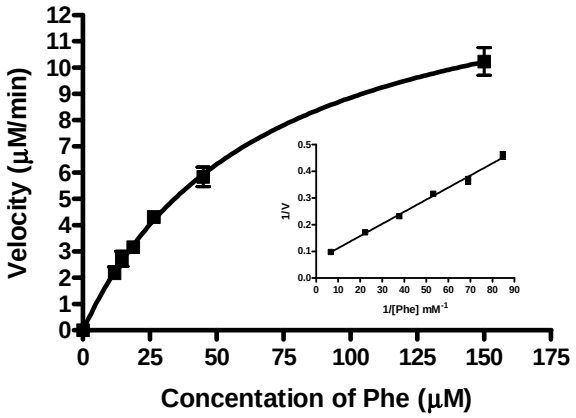


Figure II-6. TycA L-phenylalanine substrate dependence Substrate dependence of TycA with L-Phe. Data were measured in triplicate using the ESI-LC/MS method.

Discussion

The conventional radioactive pyrophosphate exchange assay has been an indispensible tool in identifying substrate specificity of isolated synthetases. The assay described herein improves on the existing assay as a complement to the ever expanding repertoire of technologies available to investigate the biosynthesis of peptide natural products.[3, 17-21] Mass based pyrophosphate exchange directly measures the mass isotopologue ratio and permits quantitative

analysis using both MALDI and ESI ionization allowing substrate identification and easy determination of exchange kinetic parameters.

Additionally, our method provides practical advantages compared to the conventional assay. The use of stable isotopes circumvents the labor and regulatory expenses related to the safe handling of radioactive materials while eliminating the potential for false results due to radioactive artifacts. The speed of the mass-based assay compares favorably to conventional exchange methods which employ solid phase capture, centrifugation or TLC steps followed by liquid scintillation counting, typically requiring continuous monitoring of β–emission for up to 48 hours for maximal sensitivity. γ-$^{18}O_4$-ATP is indefinitely stable at -80 °C in buffered solutions, while the low reaction volume used in our method (6 μL) permits over 3000 exchange reactions to be performed using 10 mg of γ-$^{18}O_4$-ATP, limiting the cost of the assay to 33 cents using commercial γ-$^{18}O_4$-ATP. Additionally, the assay was designed to allow easy scale up of the reaction mixture, if necessary. Mass-based detection can be performed in as little as 30 seconds per sample in our MALDI-TOF MS implementation, which has not yet been optimized for speed. Furthermore, the MALDI-based assay described herein requires little sample clean-up and handling prior to analysis.

Recently, microbial genomics initiatives have identified staggering numbers of gene clusters containing cryptic putative NRPS. NRPS A-domain substrate selectivity can be estimated by primary sequence analysis using homology modeling approaches or neural network algorithms.[22, 23] However, subsequent to these *in silico* analyses, it is often necessary to provide

biochemical evidence supporting A-domain selectivity. The γ–$^{18}O_4$-ATP-PPi exchange system provides a rapid, sensitive and reproducible means to measure adenylation domain specificity.

Materials and Methods

Chemicals and general methods

All chemicals were purchased from Sigma-Aldrich unless otherwise noted. γ–$^{18}O_4$-ATP was purchased from Cambridge Isotope Labs (Cambridge, MA). Protein concentrations were measured with Pierce® BCA Protein Assay Kit (Thermo Scientific).

Mass Spectrometry

MALDI-TOF MS analyses were performed on a Voyager-DE™ STR (Applied Biosystems, Inc.) using a nitrogen laser (337 nm). Prior to analysis, 1 μL of analyte/matrix mixture was spotted on to a stainless steel MALDI target. External mass calibration was performed in the reflectron mode using a mixture of γ-$^{18}O_4$-ATP and γ–$^{16}O_4$-ATP. Mass spectra were acquired in negative ion mode over a range of 450 to 1200 m/z. Mass spectra for ATP-PPi exchange analysis were obtained by averaging 100 consecutive laser shots. Data acquisition and quantitative spectral analysis was conducted using Applied Biosystems DataExplorer software, version 4.5.

ESI-LC/MS analyses were performed on a ThermoFinnigan LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, Waltham, MA) equipped with an ESI interface in negative ion mode. Nitrogen was used both for the auxiliary

and sheath gas. The auxiliary and sheath gases were set to 20 psi and 36 psi, respectively. The following instrumental parameters were used: capillary temperature 300 C; source voltage 4.5 kV; source current 100 μA capillary voltage -49.0 V; tube lens -148.30 V; skimmer offset 0.00 V; activation time 50 ms with an isolation width of 1 m/z. The sensitivity of the mass spectrometer was tuned by infusion of γ-$^{16}O_4$-ATP at a flowrate of 0.01 mL/min. Samples were introduced by a Waters Acquity UPLC system (Waters, Milford, MA) with an injection volume of 5 μL. γ-$^{18}O_4$-ATP was separated from contaminating salts on a 5 μm Hypercarb column (3 x 50mm, ThermoFisher Scientific) with an isocratic method of 82.5% 20 mM ammonium acetate pH 6 containing 0.1% diethylamine and 17.5% acetonitrile over 5 minutes with a flow rate of 0.2 mL/min. A divert valve was used for the first two minutes to avoid introducing salts into the source and the retention time of ATP species under these conditions was 3 minutes. For ESI-LC/MS/MS studies, the selection reaction monitoring mode was used with collision energy of 20eV. The mass transitions of (514 → 408, 410, 412, 414, 416), (512 → 408, 410, 412, 414), (510 → 408, 410, 414), (508 → 408. 410) and (506 → 408) corresponding to unlabeled, partially labeled and fully labeled ATP were monitored. Data acquisition and quantitative spectral analysis was conducted using the Thermo-Finnigan Xcaliber software, version 2.0 Sur 1.

Data Analysis

Monoisotopic peak area was determined using the respective software for each method. For MALDI-TOFMS analysis, the ratio of the area of γ–$^{16}O_4$-ATP (m/z 506) to the area of total ATP including unlabeled, partially labeled, fully

labeled and monosodium-coordinated ions (m/z 506, 508, 510, 512, 514, 528, 530, 532, 534, 536) was calculated using Microsoft excel and adjusted to reflect actual incorporation based on maximum theoretical incorporation to yield percent exchange. For ESI-LC/MS analysis, the ratio of the area of γ–$^{16}O_4$-ATP to the area of total ATP including unlabeled, partially labeled, fully labeled, monosodium-coordinated and sodium acetate adduct ions was calculated using Microsoft excel and adjusted to reflect actual incorporation based on maximum theoretical incorporation to yield percent exchange. For ESI-LC/MS/MS, the ratio of the area of γ–$^{16}O_4$-ATP, taken as the area of the product ion (m/z 408), to the area of total ATP including unlabeled, partially labeled, and fully labeled ions, taken as the sum of the area of their respective product ions ((514 → 408, 410, 412, 414, 416), (512 → 408, 410, 412, 414), (510 → 408, 410, 414), (508 → 408. 410) and (506 → 408)), was calculated using Microsoft excel and adjusted to reflect actual incorporation based on maximum theoretical incorporation to yield percent exchange.

Cloning, Expression and Purification of A-domains

TycA Expression and Purification

Overproduction of the 110kD His-tagged TycA PheATE was performed in *E. coli* BL21(DE3) bearing the vector pSU18-*tycAPheATE*-His. In a 2.8-L baffled flask, 500 mL LB medium containing 30 μg/mL chloramphenicol was inoculated (1:100) with an overnight culture made from a fresh colony of the above strain and grown at 37°C in a 2.8L baffled flask until OD_{600} ~ 0.6. IPTG was then added to give a final concentration of 0.5 mM and the culture was incubated for an

additional 20 h at 30 °C. The cells were pelleted (30 min, 3750 rpm, 4 °C) and resuspended in binding buffer (500 mM NaCl, 20 mM NaH_2PO_4, 20 mM imidazole, pH 7.4). For cell lysis, DNase I (NEB, 0.2 U/mL) was added, the cells were disrupted using a French pressure cell and filtered through a 0.45 μm filter. The protein was purified on a HisTrap FF column (GE Healthcare) on an ÄKTA chromatography system (GE Healthcare) using binding buffer with linearly increasing imidazole concentration (20–500 mM). The pure protein was then desalted with a HiTrap Desalting column using 20 mM Tris, pH 7.5 and stored in aliquots at –80°C in storage buffer containing 5% glycerol and 1 mM DTT.

ValA Expression and Purification

Production of the 56.6 kDa his-tagged Val-A was performed in *E. coli* BL21(DE3) bearing the vector pBS3. The conditions for expression and purification are similar to those described above except 50 μg/ml kanamycin was added to the medium instead of chloramphenicol.

Kinetic Parameters for TycA

For TycA kinetic measurements, the reactions were performed as described with the following modifications: % exchange was measured over a range of L-Phe concentrations (0.118 – 0.15 mM) and rates were measured as single time points at 3 minutes. In all cases the % exchange of reactions did not exceed 12%. The 'apparent' substrate dependence parameters, K_M and k_{cat}, were calculated from initial velocity time curves with varied L-Phe concentrations using a non-linear least square fit to the Michaelis-Menten equation.

REFERENCES

1. Caboche, S., Pupin, M., Leclere, V., Fontaine, A., Jacques, P. and Kucherov, G. (2008) NORINE: a database of nonribosomal peptides. *Nucleic Acids Research*, **36**, D326-331.

2. Lee, S. G. and Lipmann, F. (1975) Tyrocidine synthetase system. *Methods in Enzymology*, **43**, 585-602.

3. Linne, U. and Marahiel, M. A. (2004) Reactions catalyzed by mature and recombinant nonribosomal peptide synthetases. *Methods in Enzymology*, **388**, 293-315.

4. Otten, L. G., Schaffer, M. L., Villiers, B. R., Stachelhaus, T. and Hollfelder, F. (2007) An optimized ATP/PP(i)-exchange assay in 96-well format for screening of adenylation domains for applications in combinatorial biosynthesis. *Biotechnology Journal*, **2**, 232-240.

5. Du, L. and Shen, B. (1999) Identification and characterization of a type II peptidyl carrier protein from the bleomycin producer Streptomyces verticillus ATCC 15003. *Chemistry & Biology*, **6**, 507-517.

6. Joseph, D. R. and Muench, K. H. (1971) Tryptophanyl transfer ribonucleic acid synthetase of Escherichia coli. II. Molecular weight, subunit structure, sulfhydryl content, and substrate-binding properties. *The Journal of Biological Chemistry*, **246**, 7610-7615.

7. Pfeifer, E., Pavela-Vrancic, M., von Dohren, H. and Kleinkauf, H. (1995) Characterization of tyrocidine synthetase 1 (TY1): requirement of posttranslational modification for peptide biosynthesis. *Biochemistry*, **34**, 7450-7459.

8. Stern, R., DeLuca, M., Mehler, A. H. and McElroy, W. D. (1966) Role of sulfhydryl groups in activating enzymes. Properties of Escherichia coli lysine-transfer ribonucleic acid synthetase. *Biochemistry*, **5**, 126-130.

9. Hoard, D. E. and Ott, D. G. (1965) Conversion of Mono- and Oligodeoxyribonucleotides to 5-Triphosphates. *Journal of the American Chemical Society*, **87**, 1785-1788.

10. Linne, U., Schwarzer, D., Schroeder, G. N. and Marahiel, M. A. (2004) Mutational analysis of a type II thioesterase associated with nonribosomal peptide synthesis. *European Journal of Biochemistry*, **271**, 1536-1545.

11. Sun, G., Yang, K., Zhao, Z., Guan, S., Han, X. and Gross, R. W. (2007) Shotgun metabolomics approach for the analysis of negatively charged water-soluble cellular metabolites from mouse heart tissue. *Analytical Chemistry*, **79**, 6629-6640.

12. Xing, J., Apedo, A., Tymiak, A. and Zhao, N. (2004) Liquid chromatographic analysis of nucleosides and their mono-, di- and triphosphates using porous graphitic carbon stationary phase coupled with electrospray mass spectrometry. *Rapid Communications in Mass Spectrometry*, **18**, 1599-1606.

13. Hu, Y., Phelan, V., Ntai, I., Farnet, C. M., Zazopoulos, E. and Bachmann, B. O. (2007) Benzodiazepine biosynthesis in Streptomyces refuineus. *Chemistry & Biology*, **14**, 691-701.

14. Phelan, V. V., Du, Y., McLean, J. A. and Bachmann, B. O. (2009) Adenylation Enzyme Characterization Using gamma-O-18(4)-ATP Pyrophosphate Exchange. *Chemistry & Biology*, **16**, 473-478.

15. Eigner, E. A. and Loftfield, R. B. (1974) Kinetic techniques for the investigation of amino acid: tRNA ligases (aminoacyl-tRNA synthetases, amino acid activating enzymes). *Methods in Enzymology*, **29**, 601-619.

16. Cohn, M. and Hu, A. (1978) Isotopic (18O) shift in 31P nuclear magnetic resonance applied to a study of enzyme-catalyzed phosphate--phosphate exchange and phosphate (oxygen)--water exchange reactions. *Proceedings of the National Academy of Sciences of the United States of America*, **75**, 200-203.

17. Dorrestein, P. C., Bumpus, S. B., Calderone, C. T., Garneau-Tsodikova, S., Aron, Z. D., Straight, P. D., Kolter, R., Walsh, C. T. and Kelleher, N. L. (2006) Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. *Biochemistry*, **45**, 12756-12766.

18. Francklyn, C. S., First, E. A., Perona, J. J. and Hou, Y. M. (2008) Methods for kinetic and thermodynamic analysis of aminoacyl-tRNA synthetases. *Methods*, **44**, 100-118.

19. Hicks, L. M., O'Connor, S. E., Mazur, M. T., Walsh, C. T. and Kelleher, N. L. (2004) Mass spectrometric interrogation of thioester-bound intermediates in the initial stages of epothilone biosynthesis. *Chemistry & Biology*, **11**, 327-335.

20. La Clair, J. J., Foley, T. L., Schegg, T. R., Regan, C. M. and Burkart, M. D. (2004) Manipulation of carrier proteins in antibiotic biosynthesis. *Chemistry & Biology*, **11**, 195-201.

21. Zou, Y. and Yin, J. (2008) Alkyne-functionalized chemical probes for assaying the substrate specificities of the adenylation domains in nonribosomal peptide synthetases. *Chembiochem*, **9**, 2804-2810.

22. Challis, G. L., Ravel, J. and Townsend, C. A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry & Biology*, **7**, 211-224.

23. Stachelhaus, T., Mootz, H. D. and Marahiel, M. A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology*, **6**, 493-505.

CHAPTER III

BIOSYNTHETIC STUDIES OF K-26

K-26, isolated from *Astrosporangium hypotensionis* (NRRL 12379), is a member of a unique class of naturally occurring phosphonate containing peptides which incorporate a phosphonic acid analogue of tyrosine (Figure III-1).[1] K-26 was originally discovered via bioassay guided fractionation testing for inhibitors of angiotensin converting enzyme (ACE). K-26 possesses potent ACE inhibition with an IC_{50} of 14.4 nM, which is comparable to the antihypertensive drug Captopril.[2] NMR, mass spectrometry and synthetic studies have demonstrated that K-26 is a tripeptide consisting of *N*-acetylated L-isoleucine, L-tyrosine and the non-proteinogenic amino acid (*R*)-1-amino-2-(hydroxyphenyl)ethylphosphonic acid (AHEP).[1, 2] AHEP is shared among several other K-26 analogues isolated from *Streptosporangium* and *Actinomadura* species, all with some ACE inhibitory activity.[3] Despite the potent hypotensive capability of K-26 and related analogues, little is known about the biosynthetic pathway of K-26 and its AHEP moiety.

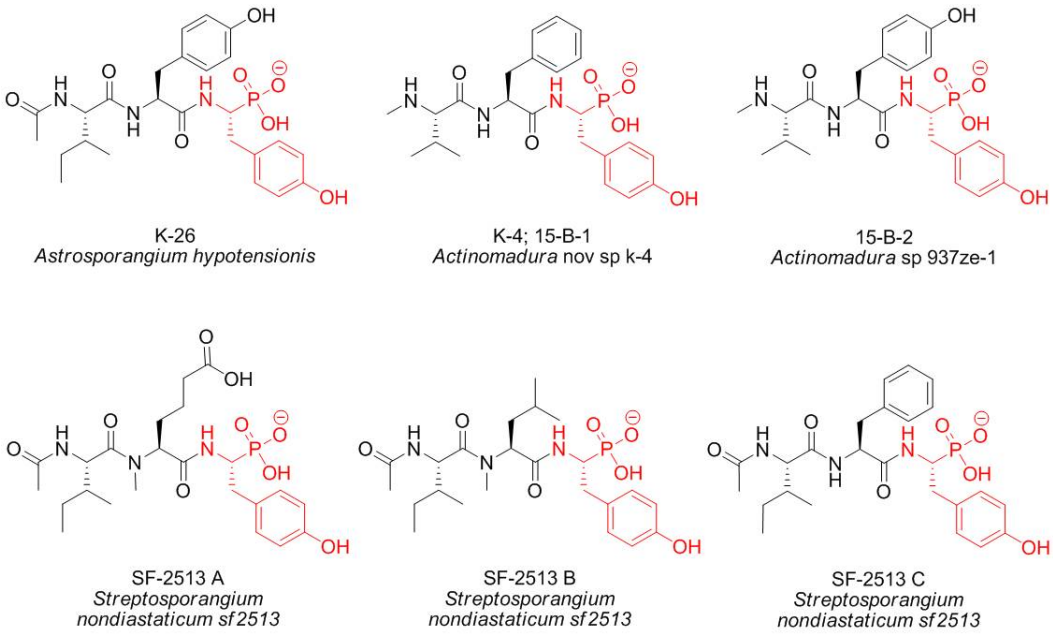


Figure 26. K-26 and Related Analogues. K-26 and its analogues share the non-proteinogenic amino acid AHEP, highlighted in red. These tripeptides all show inhibitory activity with ACE.

In all previously studied phosphonate containing natural products, the phosphonate moiety is derived from the catalytic activity of phosphoenolpyruvate (PEP) mutase.[4-8] PEP mutase catalyzes the intramolecular rearrangement of phosphoenolpyruvate to phosphonopyruvate; the biosynthetic precursor of all previously studied phosphonate natural products (Figure III-2). Structurally, AHEP is radically different from all other naturally occurring phosphonates. Additionally, it is not readily apparent how a precursor of AHEP could form the elimination intermediate necessary for carbon-phosphorus bond formation catalyzed by PEP mutase. In order to investigate the biosynthesis of K-26, it was necessary to verify its production in our hands, determine the structure and identify the biosynthetic precursors in order to reveal the genes required for K-26 biosynthesis.

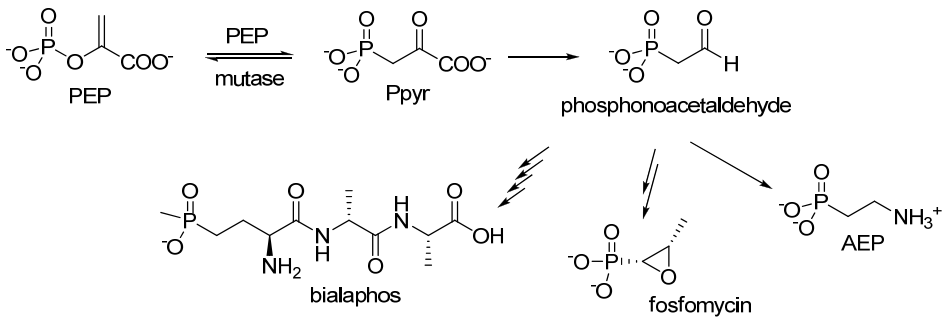


Figure 27. C-P Bond Forming Pathway. All previously investigated carbon-phosphorus bond containing natural products are derived from phosphonoacetaldehyde. Phosphonoacetaldehyde is formed subsequent to the PEP catalyzed transformation of PEP to phosphonopyruvate.

Results

Incorporation of isotopically-labeled precursors

From the original structural analysis performed by Seto and coworkers in 1986, K-26 was determined to consist of *N*-acetyl L-isoleucine, L-tyrosine and the non-proteinogenic amino acid AHEP.[1] To validate these findings and to shed light on the biosynthetic pathway of K-26 and AHEP, isotopically labeled precursors were pulse-fed to growing cultures of *A. hypotensionis*.[9] The purpose of these experiments was to reveal the timing of precursor incorporation. Initial studies by Ntai et al revealed that tyrosine is a precursor of AHEP and AHEP is a discrete precursor of K-26. Additional studies revealed that each amino acid is incorporated independently and *N*-acetylation occurs after at least the first peptide bond is formed (Figure III-3).

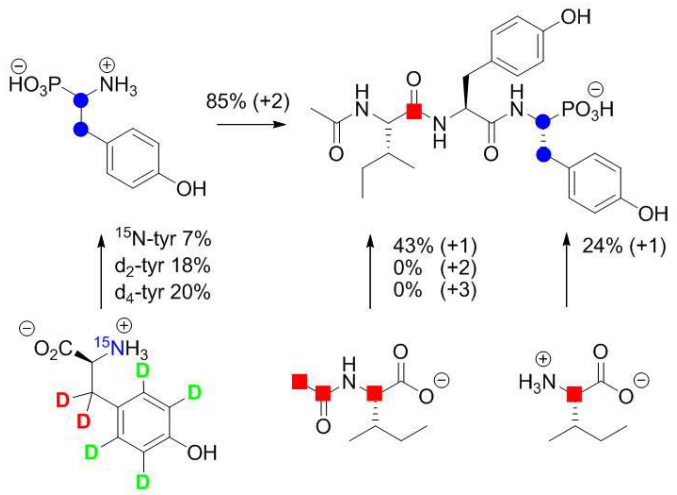


Figure 28. Precursor Incorporation into K-26. Incorporation of tyrosine into AHEP indicates that AHEP is derived from tyrosine. AHEP is incorporated intact into K-26 as is most likely a discrete precursor. *N*-acetylation occurs after the first peptide bond is formed.

Decarboxylation of tyrosine is formally required for AHEP formation, therefore, tyramine was proposed as a logical biosynthetic intermediate. D_4-tyramine was produced from D_4-tyrosine by enzymatic decarboxylation catalyzed by tyrosine decarboxylase from *Streptococcus faecalis*. Four different pulse-feeding experiments of 1 mM D_4-tyramine failed to show any incorporation into AHEP and K-26[10] (Figure III-4).

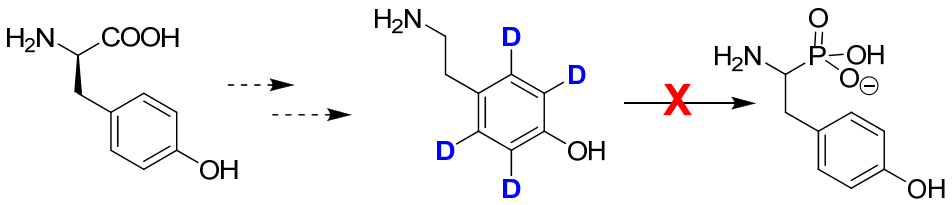
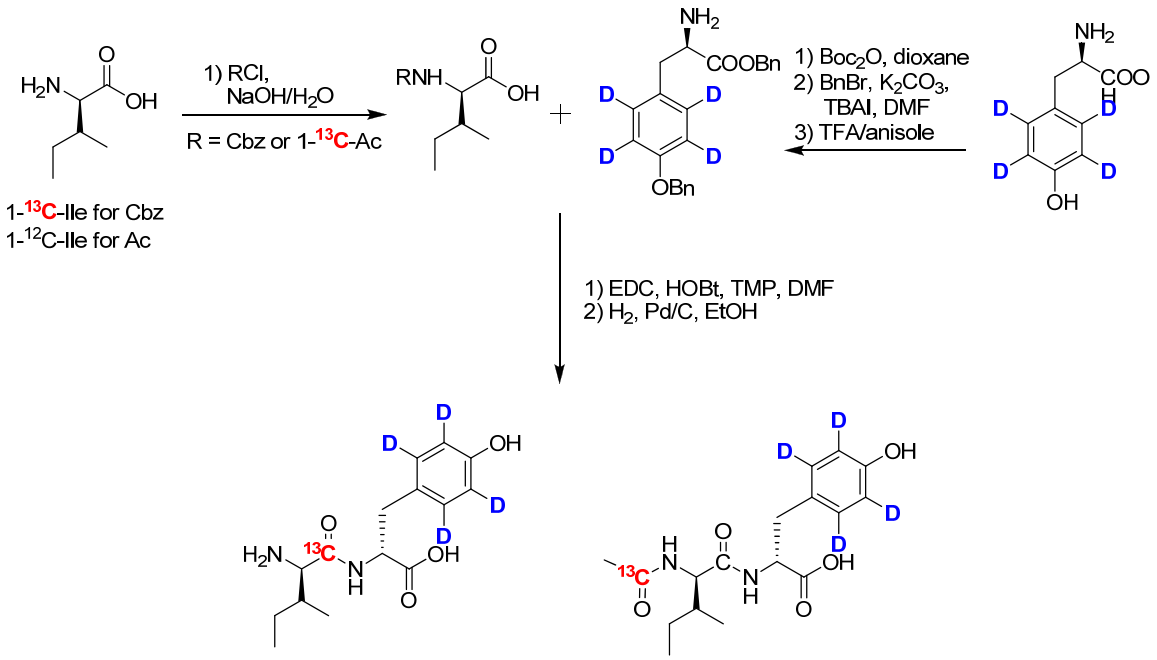


Figure 29. Tyramine Incorporation. Although decarboxylation is formally required for AHEP biosynthesis, D_4-tyramine was not incorporated into K-26.

One possible biosynthetic route for K-26 consistent with the precursor incorporation studies includes a series of ligases. To investigate this possibility, [1-^{13}C]-N-acetyl-L-isoleucine-[ring-D_4]-L-tyrosine and [1-^{13}C]-L-isoleucine-[ring-D_4]-L-tyrosine were synthesized using traditional peptide methods.[11-16] Briefly, the amine of [ring-D_4]-L-tyrosine was reacted with Boc anhydride to give the protected amine. Boc-[ring-D_4]-L-tyrosine was further protected by addition of benzyl bromide. Subsequent reaction with TFA yielded the free amine. The amine group of [1-^{13}C]-L-isoleucine was protected by Cbz and coupled with the Bn-protected [ring-D_4]-L-tyrosine under standard conditions. The protecting groups were then removed using catalytic hydrogenation resulting in the fully deprotected peptide, [1-^{13}C]-L-isoleucine-[ring-D_4]-L-tyrosine. L-isoleucine was reacted with [1-^{13}C]-acetyl chloride and coupled with the benzyl-protected [ring-D_4]-L-tyrosine as above. The benzyl groups were removed by catalytic hydrogenation to give [1-^{13}C]-*N*-acetyl-L-isoleucine-[ring-D_4]-L-tyrosine.



Scheme III-1. Synthesis of Stable Labeled Dipeptides

After synthetic preparation of the stable-labeled dipeptide precursors, separate production cultures of *A. hypotensionis* were supplemented with 1 mM of each dipeptide or with water vehicle for four days. *A. hypotensionis* was cultured in K26 production media following the production methods established by Yamato *et al.*[1] After six days of incubation, the production cultures were harvested by centrifugation, separating the culture broth supernantant and the cell paste. The culture broth was extracted by Diaion HP-20 polystyrene resin at pH 3. K-26 was eluted from the resin with a 50% methanol solution. The methanolic solution was evaporated *in vacuo* and the residue redissolved in methanol and subjected to mass spectrometric analysis. Cultures that were supplemented with only water were used for reference.

The specific incorporation of the labeled dipeptides into K-26 was determined using the Selected Reaction Monitoring (SRM) method described by Ntai *et al*[9]. This tandem mass spectrometry method fragments the [M-H] precursor ions of K-26 by collision-induced dissociation. By prudently selecting the precursor and product isotopomer masses in a given experiment, it is possible to quantitate the isotopic enrichment in both the charged product ion and the neutral loss fragments. K-26 (m/z 534) characteristically fragments to AHEP (m/z 216) and Tyr-AHEP (m/z 379) under tandem mass spectrometry conditions. Figure III-5 shows a chromatographic example of these mass spectrometry experiments. Incorporation of the labeled dipeptide, intact, would shift the precursor mass of K-26 by 5 amu. However, if the dipeptide undergoes proteolytic cleavage, the possibility exists for the precursor mass of K-26 to shift by 4 amu or 1 amu depending on the incorporation of [ring-D_4]-L-tyrosine or [1-^{13}C]-L-isoleucine/[1-^{13}C]-*N*-acetyl-L-isoleucine, respectively.

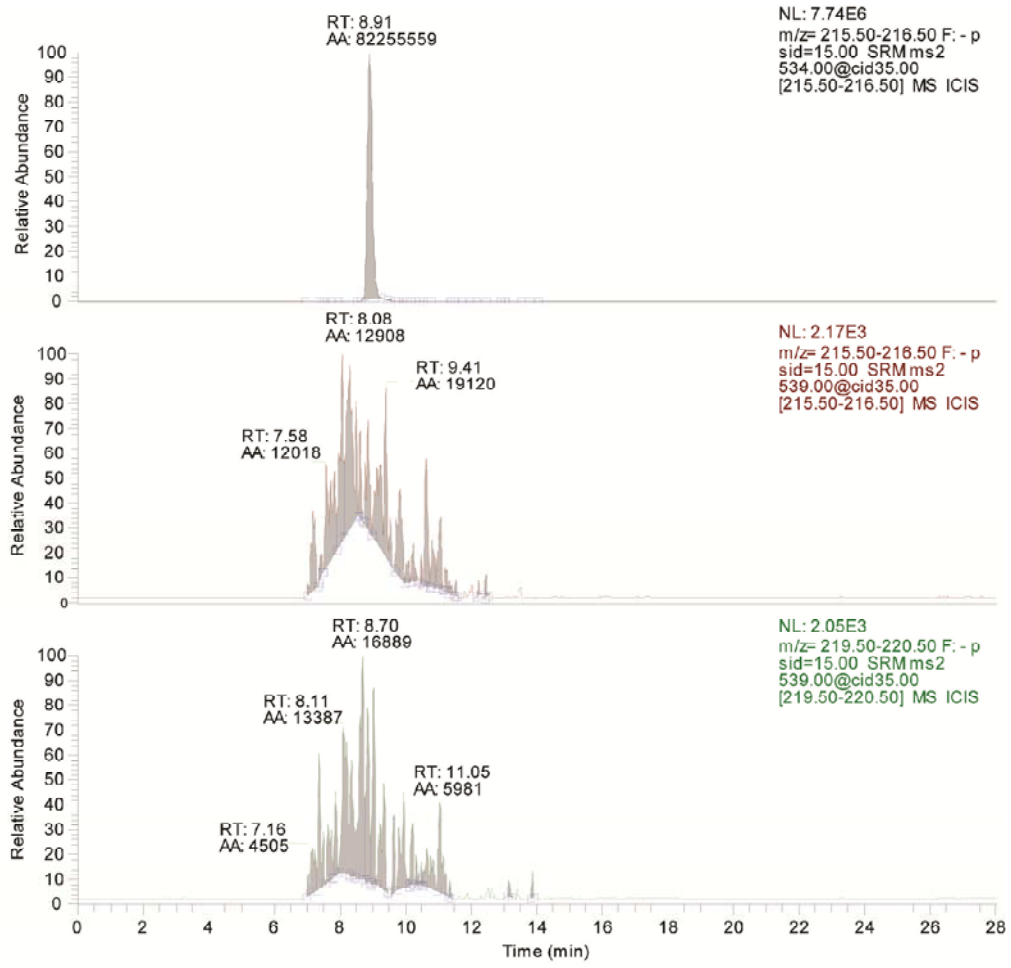


Figure 30. SRM Detection of Precursor Incorporation. To detect incorporation of a proposed biosynthetic precursor into K-26, an SRM method was used. Judicious selection of the precursor and product ion masses allows for calculation of label incorporation. This example is from a control sample from a culture of *A. hypotensionis* without administration of a labeled precursor. Top: K-26 (m/z 534) fragmentation to AHEP (m/z 216). Middle: K-26 + 5 (m/z 539) fragmentation to AHEP (m/z 216). Bottom: K-26 + 5 (m/z 539) fragmentation to AHEP + 4 (m/z 220).

To monitor incorporation of the labeled dipeptides, we selected a specific precursor mass for K-26 (m/z 534, 539) and monitored the characteristic product ions m/z 216 and m/z 220, which correspond to AHEP and AHEP+5. If the dipeptides are incorporated intact, we would expect to observe fragmentation of the K-26+5 peak (m/z 539) to AHEP (m/z 216). Fragmentation of the K-26+5 (m/z 539) peak to AHEP+4 (m/z 220) would indicate that the dipeptide was

hydrolyzed and D_4-tyrosine was incorporated into AHEP. Additional experiments monitored the possible incorporation of inidividual labeled amino acids, if hydrolysis occurred. Least squares fitting of labeled isotopomer data permitted the extraction of the relative isotopic enrichments of each amino acid in K-26. Based on the isotopic incorporation data, the dipeptides were not incorporated intact. (Figure III-6)

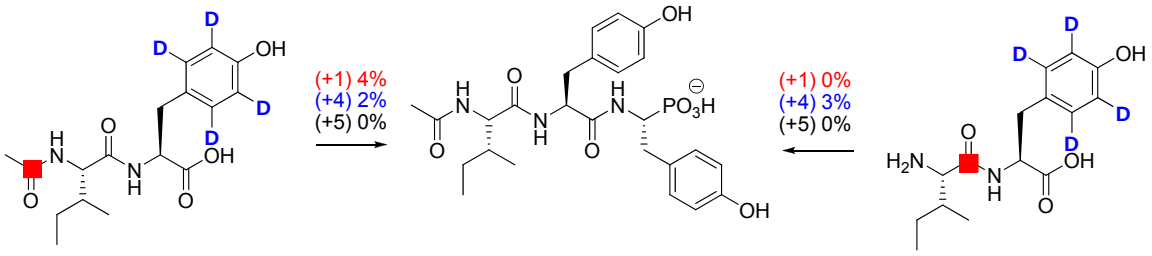


Figure 31. Dipeptide Incorporation Studies. Precursor Incorporation studies with stable-labeled didpeptide substrate indicate that the dipeptides were not incorporated intact.

The labeled tyrosine and AHEP incorporation studies revealed that tyrosine is a precursor to AHEP and AHEP is a discrete precursor to K-26. However, tyramine is not incorporated indicating that decarboxylation and phosphonylation of tyrosine may occur in concert or that the tyramine intermediate is not released before phosphonylation. Moreover, the labeled isoleucine and dipeptide experiments indicate that each amino acid is incorporated individually and that *N*-acetylation occurs after peptide bond formation.

Biosynthetic hypothesis

Based on the stable-labeled precursor incorporation studies, we hypothesized that K-26 is biosynthesized via the NRPS paradigm.[17] As described in Chapter I, NRPS are modular multidomain proteins capable of synthesizing small peptides containing non-proteinogenic amino acids. As feeding studies indicate that AHEP is a discrete precursor of K-26, we hypothesize that AHEP is biosynthesized prior to peptide bond formation. In this case, shown in Figure III-7, L-isoleucine, L-tyrosine and AHEP would be activated by individual A-domains, loaded onto the corresponding T-domains and peptide bond formation would be catalyzed by the C-domain resulting in the formation of des-acetyl-K-26 (dK-26). Then either on the final T-domain or after release of the peptide by the Te-domain, N-acetylation would occur resulting in formation of K-26. Our proposed gene cluster would consist of an N-acetyltransferase, NRPS with a domain string of A-T-C-A-T-C-A-T-Te and hypothetical proteins of unknown function responsible for AHEP biosynthesis.

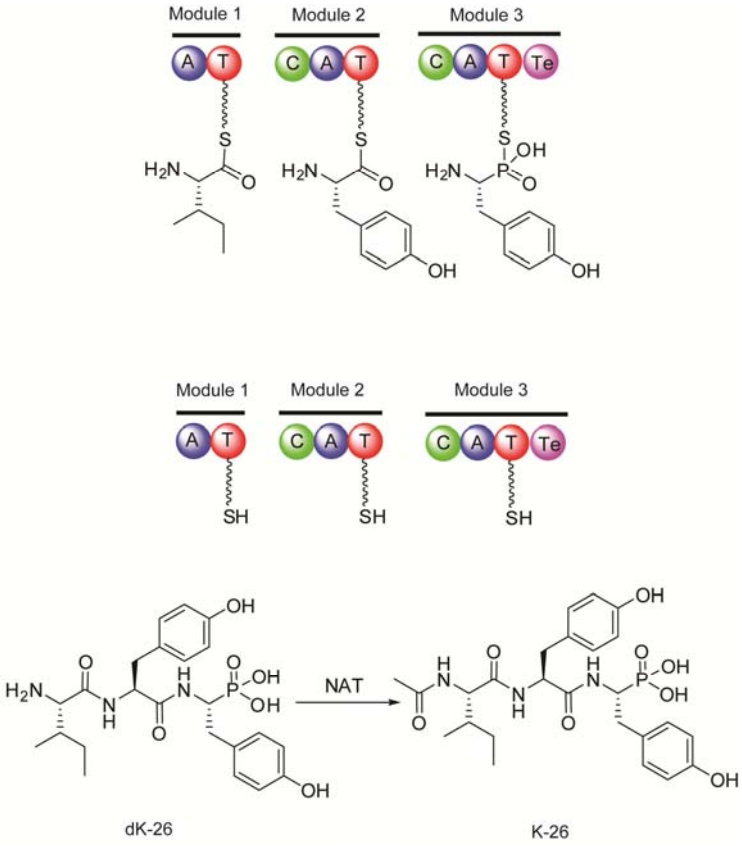


Figure 32 Possible biosynthetic route for K-26 formation. Isoleucine, tyrosine and AHEP would be loaded onto T-domains of an NRPS. Following peptide bond formation, dK-26 would be *N*-acetylated to give K-26

Genetic Analysis

With the results of the precursor incorporation studies and a biosynthetic hypothesis in hand, we turned to genetic analysis to identify the prospective gene cluster of K-26. In order to rapidly identify the gene cluster for K-26, the genome of *A. hypotensionis* was sequenced and annotated by The Institute of Genomic Research (TIGR) providing approximately 10 times coverage. From this data, we identified 11 NRPS gene clusters, four of which may potentially be involved in tripeptide biosynthesis: pps8, pps6, pps3 and pps1. Additionally, no PEP mutase

was found within the *A. hypotensionis* genome, supporting our hypothesis that AHEP is formed via a novel biosynthetic route for phosphonate formation.

In the hope that substrate specificity of the A-domains of prospective K-26 gene clusters would provide evidence to determine which tripeptide biosynthetic cassette is responsible for K-26 biosynthesis, we extracted the 8- to 10- amino acid specificity code involved in discrimination of amino acid binding within the A-domain binding pocket, as described by Challis *et al* and Stachelhaus *et al*.[18, 19] The prospective gene clusters and their A-domain specificity codes are shown in Table III-1. Out of these four gene clusters, pps3 seems like the most promising gene cluster, however, no *N*-acetyltransferase nor proteins of unknown function are clustered with this NRPS. As no clear gene cluster for K-26 biosynthesis was revealed by genomic evaluation, we turned to traditional biochemical methods to isolate of the protein machinery responsible for K-26 biosynthesis from cell-free extract of *A. hypotensionis*.

Table III-1. Prospective K-26 NRPS Gene Clusters

Cluster	Domains	Selectivity Code	Amino acid
pps6	AT-?AT-CATTe	A1: DAQDFGVVDK	Gln – 90% Identity
		A2: DALQMAGGFK	Unknown
		A3: DGMDSGFVEK	Unknown
pps8	CAT-AT-CATTe	A1: DAAHLGCMCK	Unknown
		A2: DILQLGMIW-	Gly - 100% (K517 missing)
		A3: DMPKVGEVAK	Unknown
pps3	AT-CAT-CATTe	A1: DGFFVGGAK	Ile – 70% Identity
		A2: DVFYLIGVIK	Unknown – 90% Id w/A3
		A3: DVFYLAGVIK	Iva, abu – 70% Identity
pps1	CAT-CACT-CATTe	A1: DILQLGVIWK	Gly – 90% Identity
		A2: DILQLGVIWK	Gly – 90% Identity
		A3: DFTKLGHVGK	Asp – 80% Identity

Reverse Genetic Analysis

As genomic analysis did not provide us with sufficient evidence to identify the biosynthetic gene cluster of K-26, we were forced to attempt to identify the gene cluster using traditional biochemical methods. Based on the structure of K-26, there are three possible handles for biochemical assay development: *N*-acetylation, peptide bond formation and AHEP biosynthesis. To begin biochemical investigations, *N*-acetylation of dK-26 was first investigated, as most acetyltransferases only require acetyl-Coenzyme A (Ac-CoA) and substrate for activity. Second, the mass-based pyrophosphate exchange assay was used to test for L-isoleucine, L-tyrosine and AHEP activation required for peptide bond formation in nonribosomal peptide biosynthesis.

N-acetyltransferase Isolation

Based on the precursor incorporation experiments, *N*-acetylation occurs after peptide bond formation in K-26 biosynthesis. To identify the *N*-acetyltransferase (NAT) involved in K-26 biosynthesis, Ioanna Ntai developed a mass spectrometry based assay.[20] Protein fractions generated from *A. hypotensionis* cells shown to produce K-26 were incubated with dK-26 and acetyl-CoA. After incubation, the reaction mixture was subjected to mass spectrometry utilizing selected reaction monitoring (SRM) specifically monitoring the fragmentation of the precursor, dK-26 ([M-H] 492); the product, K-26 ([M-H] 534) and the internal standard D_3-K26 ([M-H] 537) to AHEP ([M-H] 216). From proteomic analyses of SDS-PAGE gel slices of partially purified protein mixtures,

three NATs were identified: ORF1625, ORF5817 and ORF6213. After heterologous expression and purification, the purified proteins were tested for N-acetylation of dK-26, Ile-Tyr-(*S*)-AHEP, Ile-Tyr and Ile-Tyr-Tyr. The activities of these NATs are summarized in Table III-2.

Table III-2. Summary of NAT Activity

NAT	Substrate	Turnover
ORF1625	L-Ile-L-Tyr-(*R*)-AHEP	0.07%
	L-Ile-L-Tyr-(*S*)-AHEP	0.01%
ORF5817	L-Ile-L-Tyr-(*R*)-AHEP	0.01%
	L-Ile-L-Tyr-(*S*)-AHEP	ND*
ORF6213	L-Ile-L-Tyr-(*R*)-AHEP	5%
	L-Ile-L-Tyr-(*S*)-AHEP	5%
	L-Ile-L-Tyr	11%
	L-Ile-L-Tyr-L-Tyr	55%

Of these NATs, ORF6213 showed the most *N*-acetyltransferase activity on dK-26. ORF6213 is within the pps1 tripeptide gene cluster. However, the first two A-domains of the NRPS of pps1 have the same 8- to 10- amino acid specificity code within the A-domain amino acid binding pocket indicating that both A-domains are likely responsible for activating the same amino acid. To determine whether ORF6213 is the NAT responsible for acetylation of dK-26, possibly as the result of off-target activity, an attempt was made to measure the K_M. Protein fractions (30 μL) were incubated with Ac-CoA and dK-26 (0.5 mM – 5 mM) at 30 ^oC for 4 hours. The reaction was quenched by addition of an equal volume of methanol and subjected to mass spectrometry analysis.

The K_M of ORF6213 was determined to be higher than 5mM dK-26 (Figure III-8). Higher concentrations of dK-26 could not be tested due to the limited solubility of dK-26. These results indicate that the acetylation of dK26 may be due to promiscuous NATs and not a dedicated NAT clustered with the peptide bond forming machinery. Since these experiments did not unambiguously identify the gene cluster of K-26, we proceeded to use our second handle, peptide bond formation, in an attempt to isolate the proteins responsible for K-26 biosynthesis.

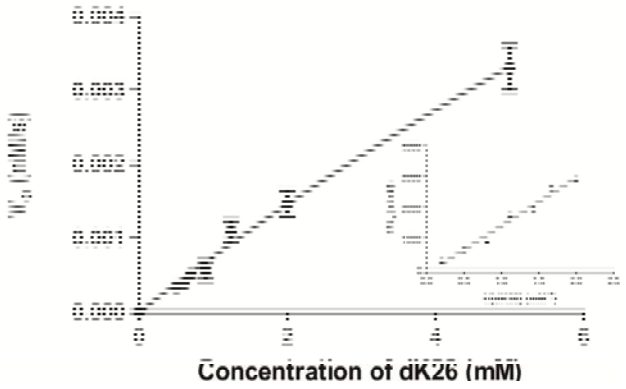


Figure 33. Kinetic Characterization of ORF6213. The K_M of Orf6213 is greater than 5mM. Higher dK-26 concentrations could not be tested due to the low solubility of dK-26.

Adenylation Enzyme Isolation

Based on our hypothesis, K-26 is biosynthesized via the NRPS paradigm where A-domains are involved in amino acid selectivity and promote the reaction of the cognate amino acid with ATP to create the aminoacyl adenylate and release of inorganic pyrophosphate.[17] This reaction is reversible, in most cases, and can be followed by the pyrophosphate exchange assay. To purify the

proteins responsible for amino acid activation in K-26 biosynthesis, protein fractions were tested for exchange activity with Ile, Tyr and AHEP, with a negative control performed concurrently.[21-25] To test for this activity, we used the mass-based pyrophosphate exchange assay as described in Chapter II. Briefly, partially purified protein fractions were incubated with 1 mM γ-$^{18}O_4$-ATP, 5 mM pyrophosphate and 1 mM amino acid substrate in the presence of magnesium. After incubation, the reaction was quenched and the mixture was subjected to mass spectrometry analysis. Activity was measured as the percentage of unlabeled ATP, from the back exchange of pyrophosphate, to total ATP, adjusted to reflect the probability of incorporation.

With a sensitive and accurate assay in hand, we began our initial purification of adenylation enzymes from *A. hypotensionis* cell-free extract with the goal of proteomic identification of the enzymes responsible for activity and subsequent correlation of the proteins to the biosynthetic gene cluster of K-26. Figure III-9 is a comprehensive chart of all the purification methods used. The following represents the final attempt at purification:

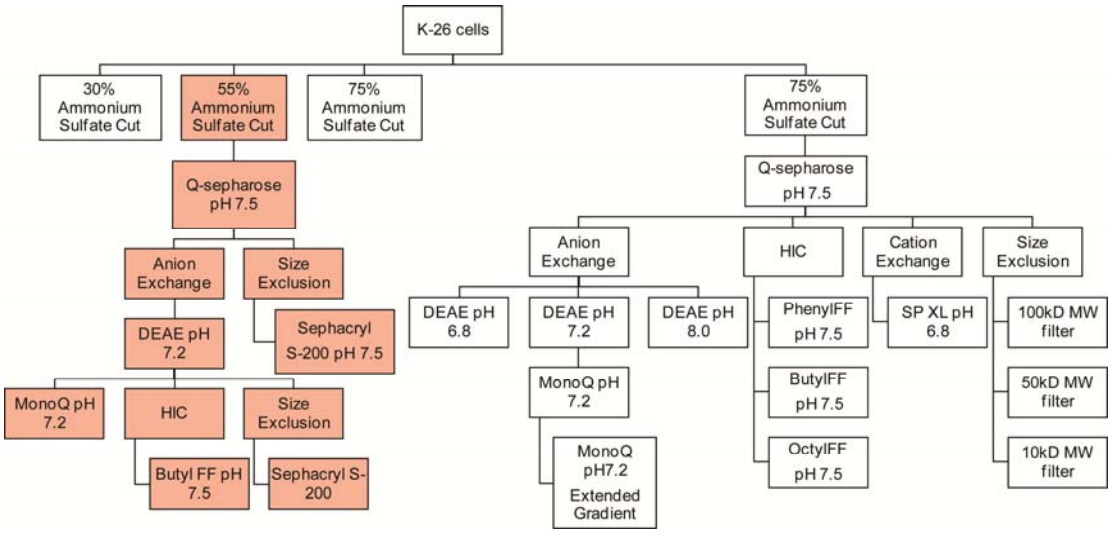


Figure 34. Overview of adenylation enzyme purification methods. The methods used for final purification are highlighted in light red.

A. hypotensionis was cultured in K26 production media following the production methods established by Yamato *et al.* The production cultures were harvested by centrifugation, separating the culture broth and the cell paste. The broth was adjusted to pH 3 and extracted with Diaion HP-20 resin. K-26 was eluted from the HP-20 resin with a 50% methanol solution. The methanolic solution was evaporated *in vacuo* and the residue redissolved in methanol and subjected to mass spectrometric analysis, confirming the production of K-26. The cell paste, of *A. hypotensionis* cultures verified to produce K-26, was washed with 20% glycerol. The mycelia were frozen in liquid nitrogen for storage at -80^{o}C. Protein purification was initiated by suspending approximately 60 g of frozen mycelia in general enzyme buffer (GEB) and rupturing the cells using a French pressure cell. Nucleic acids were precipitated by streptomycin sulfate addition followed by centrifugation. After removal of the precipitated nucleic acids, crude protein fractionation was performed on the supernatant by

ammonium sulfate precipitation at 30%, 55% and 75% saturation. After centrifugation, each ammonium sulfate pellet was dialyzed overnight against GEB to remove excess salt and the dialysate was tested for activity with L-isoleucine, L-tyrosine and AHEP. Initial testing of the dialysate revealed that most of the activity precipitated in the 55% ammonium sulfate fraction. (Figure III-10)

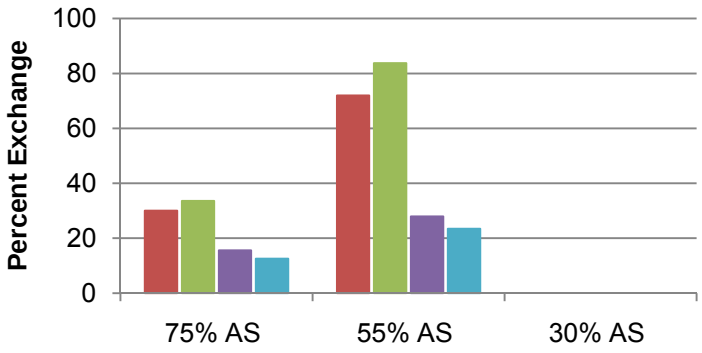


Figure III-10. Ammonium Sulfate Fractionation. Pyrophosphate exchange analysis of ammonium sulfate fractions of the *A. hypotensionis* proteome revealed that the 55% saturation fraction contained most of the proteins responsible for activating the amino acid precursors. Tyr = red, Ile = green, AHEP = purple and no amino acid control = light blue

After ammonium sulfate precipitation, further fractionation of the 55% saturation fraction was performed by anion exchange chromatography using Q-sepharose resin. (Figure III-11) Pyrophosphate activity for tyrosine and isoleucine co-eluted, while no activity was detected for AHEP or the negative control.

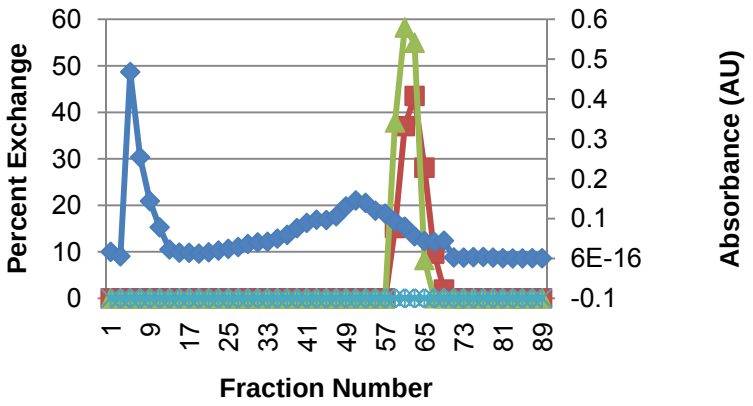


Figure 35. Q-sepharose Fractionation. After ammonium sulfate precipitation, the 55% saturation fraction was further fractionated by Q-sepharose anion exchange. Pyrophosphate activity for tyrosine and isoleucine co-eluted, while no activity was detected for AHEP or the negative control. Tyr = red, Ile= green, AHEP = purple, no amino acid control = light blue, absorbance at 280 = dark blue.

The active fractions from the Q-sepharose column were combined, concentrated and applied to a DEAE sepharose anion exchange column, in an attempt to further purify the proteins responsible for the observed pyrophosphate exchange activity. (Figure III-12)

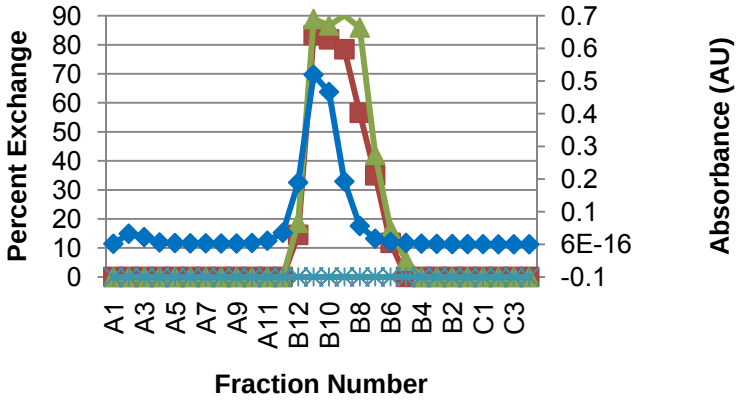


Figure 36. DEAE-sepharose Fractionation. After separation by Q-sepharose resin, the active fractions were further fractionated by DEAE anion exchange. Pyrophosphate activity for tyrosine and isoleucine co-eluted, while no activity was detected for AHEP or the negative control. Tyr = red, Ile= green, AHEP = purple, no amino acid control = light blue, absorbance at 280 = dark blue

The active fractions were combined, concentrated and divided in half to continue purification using two methods: hydrophobic interaction, using Butyl Fast Flow resin, and size exclusion, using Sephacryl S-200 resin. (Figure III-13)

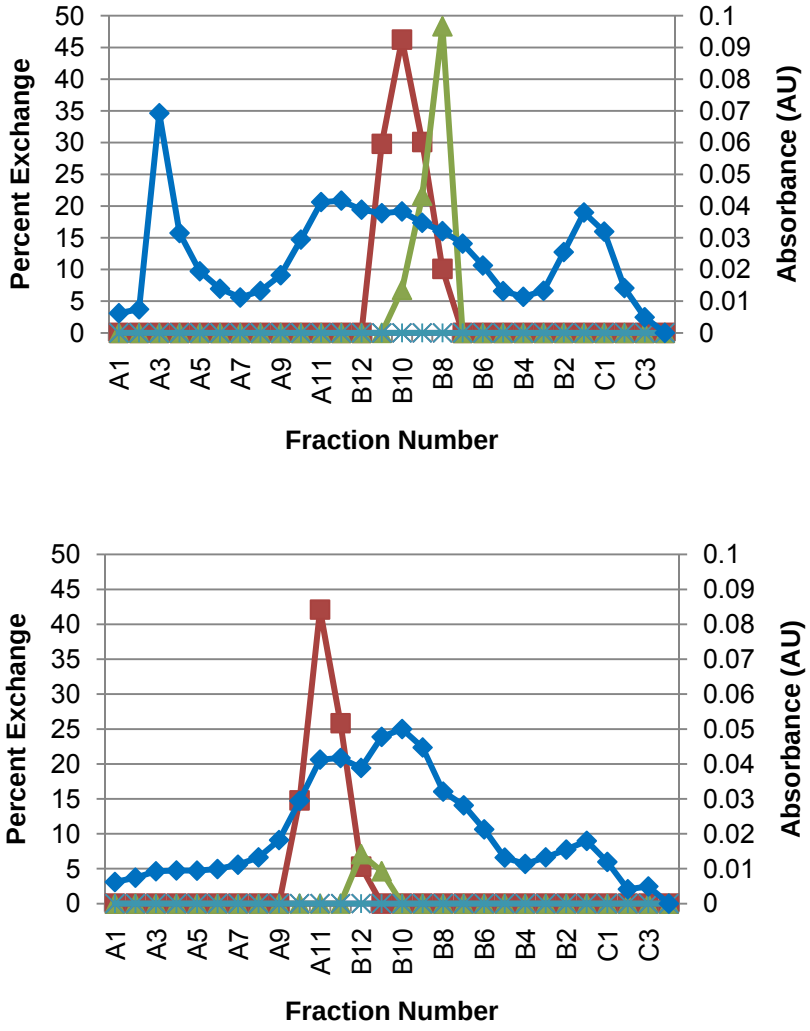


Figure 37. Hydrophobicity and Size Exclusion Fractionation. After DEAE anion exchange fractionation, half of the pooled fractions were applied to either a Butyl Fast Flow column (hydrophobic separation, top graph) or a Sephacryl S-200 column (size exclusion, bottom graph). Some separation of the pyrophosphate exchange activity for tyrosine and isoleucine was observed. Tyr = red, Ile= green, AHEP = purple, no amino acid control = light blue, absorbance at 280 = dark blue

The active fractions were subjected to SDS-PAGE analysis and two fractions were sent for proteomics analysis (Figure III-14): Sephacryl S-200 fraction 11, which showed 42% activity for L-isoleucine and 0% activity for L-tyrosine, and Butyl Fast Flow fraction B9, which showed 30% activity for L-isoleucine and 21% activity for L-tyrosine.

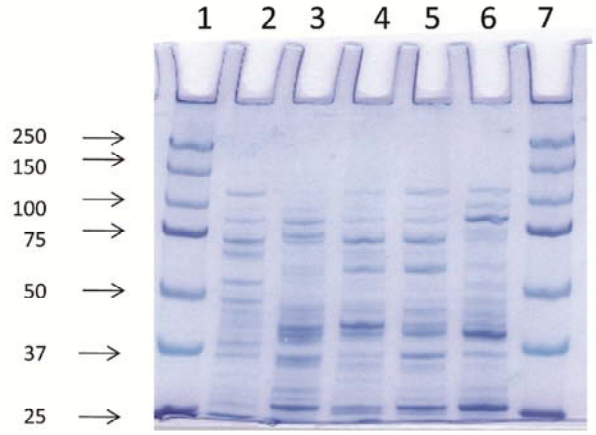


Figure 38. SDS-PAGE of active fractions for isoleucine and/or tyrosine adenylation. Lane 1: Marker, Lane 2: Sephacryl S-200 fraction 11, Lane 3: Sephacryl S-200 fraction 14, Lane 4: ButylFF fraction B8, Lane 5: Butyl FF fraction B9, Lane 6: ButylFF fraction B11, Lane 7: Marker. Sephacryl S-200 fraction 11 and ButylFF fraction B9 were subjected to proteomics analysis.

Proteomics analysis of these two samples revealed 101 proteins, two of which may correspond to adenylation activity: isoleucyl tRNA synthetase and a hypothetical protein containing an AMP-binding domain. To verify the proteomics results and to test whether the tyrosine adenylation activity was also due to a tyrosyl tRNA synthetase, a radioactive tRNA loading assay was performed using active fractions from a duplicate purification where instead of using hydrophobicity or size exclusion, the protein mixture was subjected to anion

exchange fractionation using MonoQ resin.[26] It should be noted that pooled active fractions were tested for pyrophosphate exchange against a panel of amino acids to ensure our isolation of isoleucine and tyrosine activating enzymes. (Figure III-15)

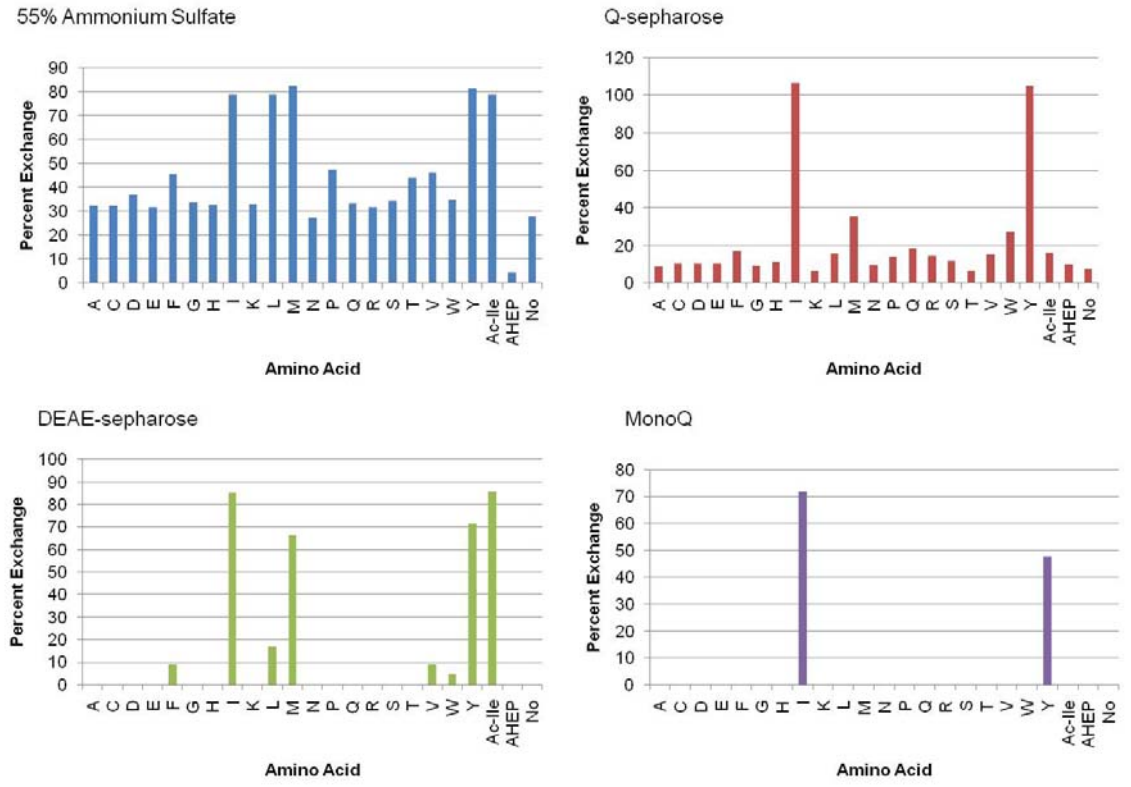


Figure 39. Pyrophosphate Exchange Analysis of Fractions. To verify the isolation of isoleucine and tyrosine activating enzymes, at each step, pooled fractions were tested for pyrophosphate activity against a panel of amino acids.

The MonoQ fraction active for both L-isoleucine and L-tyrosine catalyzed the loading of tritiated L-isoleucine and L-tyrosine onto *E. coli* tRNA, in separate experiments, confirming the isolation of tRNA synthetases from *A. hypotensionis*. (Figure III-16)

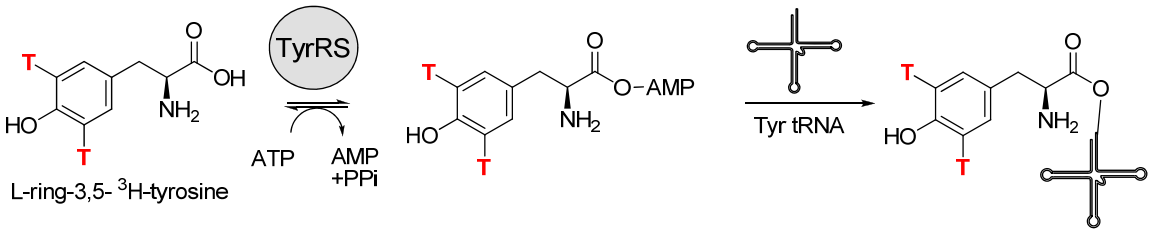


Figure 40. tRNA Charging Assay. To test for the presence of tRNA synthetases, the traditional radioactive tRNA charging assay can be performed. For example, radiolabeled L-tyrosine and tyrosine specific tRNA would be incubated with a protein mixture to test for the presence of TyrRS in a protein mixure. If TyrRS is present, the Tyr tRNA would be loaded with the labeled L-tyrosine.

Additionally from genomic analysis, the hypothetical protein identified via proteomics was part of a small gene cluster containing an *N*-acetyltransferase. To investigate whether this cluster was responsible for K-26 biosynthesis, we heterologously expressed and purified the *N*-acetyltransferase (NAT 97747) and tested for *N*-acetylation of dK-26. (Figure III-17) Neither dK-26 nor Ile-Tyr were acetylated.

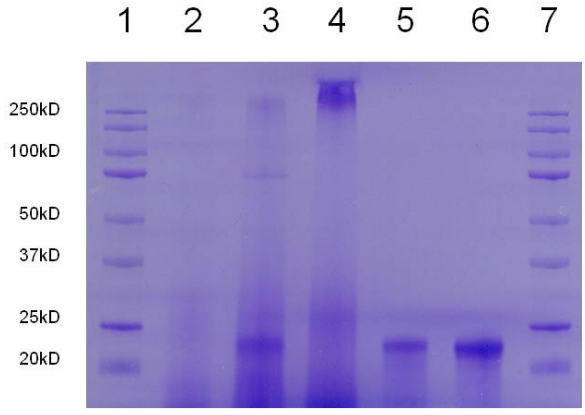


Figure 41. SDS-PAGE of NAT 97747. Lane 1: Marker; Lane 2: Uninduced; Lane 3: Induced; Lane 4 Flow-through; Lane 5: His-Trap Purified; Lane 6: Desalted; Lane 7: Marker.

Discussion

Identification of the biosynthetic genes of K-26 presents a unique challenge. In our hands, K-26 is produced by *A. hypotensionis* at approximately 10 μg/L.[9] At such low levels of production, the only method for verifying K-26 production is through mass spectrometry. The tripeptide nature of K-26 and the unique pharmacophore of AHEP limit the handles available to exploit for biosynthetic investigations. The required genes for K-26 biosynthesis are limited to an *N*-acetyltransferase, peptide bond forming machinery and those required for AHEP biosynthesis. Traditional precursor incorporation experiments with stable-labeled biosynthetic precursors allowed us to propose a NRPS biosynthetic mechanism for K-26 formation and a possible pathway for AHEP formation from tyrosine.[9, 10] Although the genome of *A. hypotensionis* has been sequenced, the biosynthetic cassette for K-26 remains cryptic.

As an *N*-acetyltransferase is formally required for K-26 formation, it was selected as the initial handle to examine K-26 biosynthesis.[20] The results of the precursor incorporation studies suggest that *N*-acetylation occurs after formation of the peptide bond between L-isoleucine and L-tyrosine. Activity guided protein purification and subsequent proteomics analysis identified three *N*-acetyltranferases, which were heterologously expressed in *E. coli*: ORF 1625, which was isolated multiple times; ORF5817, which had high sequence coverage in proteomics analysis and ORF6213, which is situated neighboring pps1, an NRPS tripeptide gene cluster. While heterologously expressed ORF1625 and ORF5817 showed limited turnover of dK-26, ORF6213 showed moderate

turnover for dK-26 and L-Ile- L-Tyr and good turnover for L-Ile-L-Tyr-L-Tyr. The K_M of ORF6213 indicates that ORF6213 may not be dedicated to K-26 biosynthesis; however, it may be the *N*-acetyltransferase responsible for acetylation of dK-26 through off-target activity. Further analysis of the 8- to 10- amino acid specificity code within the A-domains of pps1 revealed that the first and second A-domains within the cluster have identical specificity codes and predict glycine incorporation while the third A-domain shows homology to aspartate activating A-domains making pps1 an unlikely candidate gene cluster for K-26 biosynthesis.

As investigation of *N*-acetyltransferase activity did not reveal the biosynthetic genes of K-26, peptide formation was selected as the second handle to analyze. Typically, small peptides containing non-proteinogenic amino acids are biosynthesized by the NRPS paradigm. To investigate peptide bond formation, we chose to use the mass-based pyrophosphate exchange assay as NRPS A-domains, tRNA synthetases and AMP-binding peptide ligases usually catalyze the formation of an aminoacyl adenylate. Pyrophosphate exchange experiments were carried out with L-isoleucine, L-tyrosine and AHEP. L-isoleucine and L-tyrosine showed pyrophosphate exchange activity, while AHEP did not. The lack of exchange activity was not wholly unexpected as there are some A-domains and tRNA synthetases that do not undergo reversible exchange. By following the L-isoleucine and L-tyrosine pyrophosphate exchange activity, an isoleucyl tRNA synthetase and one hypothetical protein with an AMP-binding domain were identified by proteomics analysis. However, no NRPS or L-tyrosine activating enzymes were identified. To confirm the isolation of

isoleucyl tRNA synthetase and to explore the possibile isolation of tyrosyl tRNA synthetase, we performed the traditional tRNA charging assays for both L-isoleucine and L-tyrosine. Both *E. coli* isoleucyl and tyrosyl tRNAs were charged with the corresponding amino acid indicating the pyrophosphate exchange activity observed with both L-isoleucine and L-tyrosine was caused by tRNA synthetases. To determine whether the protein of unknown function with an AMP-binding motif was involved in K-26 biosynthesis, NAT 97747 within the same gene cluster was heterologously expressed and purified. The purified protein showed no acetylation of dK-26 or L-Ile- L-Tyr.

Unfortunately, we have not been able to unambiguously identify the gene cluster responsible for K-26 biosynthesis. The ability to identify the proteins responsible for K-26 biosynthesis is limited by the production levels of K-26. Correspondingly, the proteins involved in K-26 are presumably also produced at extremely low levels. In order to isolate the proteins responsible for K-26 biosynthesis, it will be necessary to increase the production level of K-26 in *A. hypotensionis* or identify another producer with higher production levels.

We have identified another producer of K-26, and two additional analogues, *Sphaerosporangium rubeum*. However, the production levels of *S. rubeum* are similar to those of *A. hypotensionis*. Briefly, seed culture of *S. rubeum* was grown in K26S media. After seed culture, *S. rubeum* was grown in 6 different media for K-26 analogue production. These cultures were extracted using the methods previously described. AHEP containing peptide natural products characteristically fragment to AHEP when subjected to tandem mass

spectrometry analysis. Using precursor ion scanning, we identified K-26, SF5213B and SF5213C in *S. rubeum* culture broth. (Figure III-18)

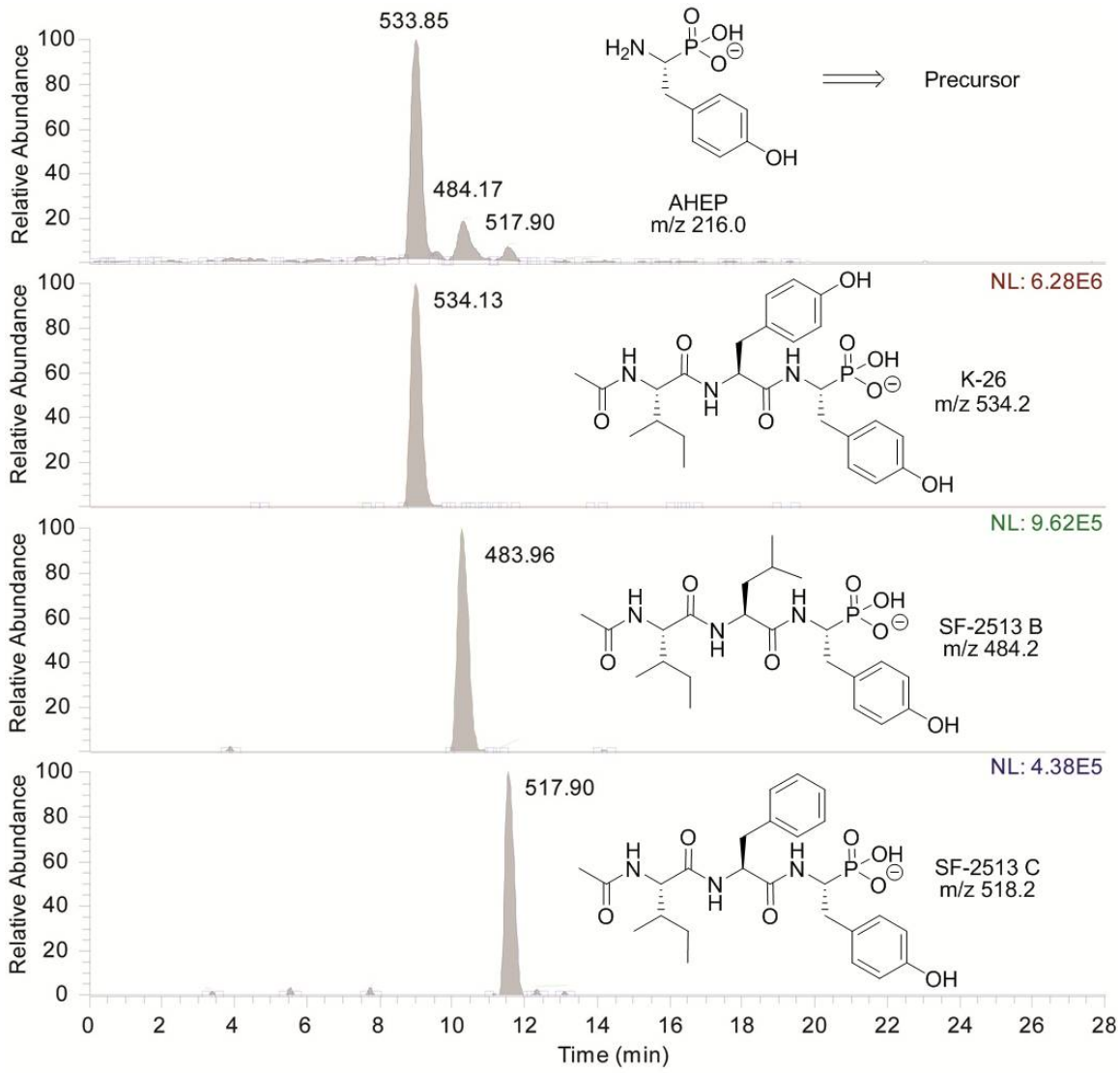


Figure 42. K-26 and Analogue Production by *S. rubeum*. Precursor ion scanning mass spectrometry experiments of *S. rubeum* production cultures identified production of K-26, SF-2513B and SF-22513C.

The results of the aforementioned experiments seem to indicate that the *N*-acetyltransferase responsible for acetylating dK-26 may not be clustered with the rest of the genes responsible for K-26 biosynthesis. Speculatively, if

acetylation is occurring through off-target activity of a variety of *N*-acetyltransferases and peptide bond formation is through a promiscuous ligase, the only genes required for K-26 biosynthesis are those encoding the proteins responsible for AHEP formation. This lack of clustering may limit the handles available for investigation to AHEP formation. Additional difficulties lie in the lack of information about the precursors of AHEP. From the stable-labeled precursor incorporation experiments, we know that AHEP is derived from L-tyrosine and AHEP is a discrete precursor of K-26, however, during the above experiments, at no time was dK-26 or K-26 formation observed when L-isoleucine, L-tyrosine and AHEP were incubated with cell-free extract of *A. hypotensionis*. In order to identify the proteins required for AHEP formation, the phosphate source and additional cofactors will have to be identified. Additionally, as AHEP formation is expected to be the rate limiting step in K-26 biosynthesis, it will be necessary to develop an extremely sensitive assay for AHEP formation.

Complementary to reverse genetic experiments, heterologous expression of A-domains within prospective K-26 gene clusters has been undertaken. While this has been an ongoing route to identify of the K-26 biosynthetic cassette, expression of K-26 A-domains in heterologous hosts has proved to be challenging, limiting the success of this endeavor. Recent reports have indicated that small MbtH-like proteins often found near or within NRPS gene clusters may play a role in increasing the solubility and expression levels of A-domains when co-expressed in heterologous hosts. Future work co-expressing these proteins

may help to identify and/or eliminate prospective gene clusters involved in K-26 biosynthesis.

Materials and Methods

Chemicals and general methods

All non-aqueous reactions were performed under ultra high purity argon in flame-dried glassware. All chemicals were purchased from Sigma-Aldrich at the highest purity available unless otherwise noted. γ-$^{18}O_4$-ATP, [ring-D_4]-L-tyrosine, [1-^{13}C]-acyl chloride and [1-^{13}C]-L-isoleucine were purchased from Cambridge Isotope Labs (Cambridge, MA). Reactions were monitored by thin-layer chromatography (TLC) using E. Merck precoated silica gel 60 F254 plates. Visualization was accomplished by UV light and aqueous stain followed by charring on a hot plate. Flash chromatography was performed using the indicated solvents and silica gel (230-400 mesh). 1H and 13C NMR were recorded on Bruker 300, 400 or 500 MHz spectrometers.

Synthesis of [ring-D_4]-tyramine

[ring-D_4]-L-tyrosine (100 mg, 0.54 mmol) was dissolved in H_2O (100 mL) by microwave heating. A suspension of tyrosine decarboxylase from *Streptococcus faecalis* (2.0 mg, 0.05 unit/mg) and pyridoxal-5-phosphate (7 mg, 0.026 mmol) in acetate buffer (4 mL of 0.1 M, pH 5.5) was added to the cooled tyrosine solution. The solution was incubated at 37^{o}C for 3 hr. The reaction was stopped by heating to boiling. Tyramine was purified using Millipore Ultra 5000 MWCO

centrifugal filter devices (3450g, 30 min). The filtrate was concentrated *in vacuo* and passed through a Dowex 1-x8 anion exchange column. Evaporation of the eluent gave [ring-D_4]-tyamine as a tan powder (98% D, 97% yield). ^{1}H NMR (300 MHz, D_2O): δ2.747 (t, *J* = 7.2Hz, 2H), 3.058 (t, *J* = 7.2Hz, 2H). ^{13}C NMR (100 MHz, D_2O): δ155.33, 129.80 (t, *J*= 19Hz), 127.68, 115.66 (t, *J*= 19Hz), 40.79, 31.89

***Synthesis of* [1-^{13}C]-N-Ac-L-Ile-[ring-D_4]-L-Tyr**

***N*-Boc-[ring-D_4]-L-Tyr.** A solution of 100 mg [ring-D_4]-L-tyrosine (1 eq, 100 mg, 0.54 mmol) in dioxane (7.5 mL), H_2O (7.5 mL) and NaOH (1M, 1 eq, 540 μL, 0.54 mmol) was stirred and cooled in an ice-water bath. Di-*tert*-butyl dicarbonate (1.1 eq, 136 μL, 0.59 mmol) was added and the solution was stirred overnight. The solution was concentrated *in vacuo* and the aqueous solution was acidified to pH 2. The acidified aqueous solution was extracted with EtOAc (3 x 20 mL). The organic layers were pooled, washed with brine (3 x 100 mL), dried ($MgSO_4$) and evaporated to give a white crystalline powder. (0.48 mmol, 138.9 mg, 89.5%) ^{1}H NMR (400 MHz, DMSO): δ9.15 (s, 1H), 3.95 (m, 1H), 2.85 (m, 1H), 2.69 (s, 1H), 1.36 (s, 9H). ^{13}C NMR (100 MHz, DMSO): δ174.13, 155.83, 128.18, 78.38, 55.90, 35.95, 28.55.

***N*-Boc-[ring-D_4]-L-Tyr(OBn)-COOBn.** A solution of *N*-Boc-[ring-D_4]-L-Tyrosine (1 eq, 100 mg, 0.35 mmol), in dry DMF (3 mL) at room temperature was treated with potassium carbonate (2.5 eq, 26.7 mg, 0.70 mmol), benzyl bromide (2.5 eq, 83 μL, 0.70 mmol), and tetrabutylammonium iodide (0.125 eq, 16 mg, 0.044

mmol). The reaction mixture was stirred at room temperature overnight. The aqueous mixture was extracted with ether (3 x 10 mL). The organic layers were combined washed with 1 N HCl (3 x 30 mL) and brine (3 x 30 mL), dried ($MgSO_4$) and concentrated *in vacuo* leaving a yellow solid. (0.24 mmol, 110.8 mg, 68.6%) ^{1}H NMR (500 MHz, DMSO): δ9.19 (s, 1H), 7.43-7.26 (m, 10H), 5.07 (s, 2H), 4.12 (s, 1H), 3.33 (s, 2H), 2.85 (m, 1H), 2.69 (m, 1H), 1.32 (s, 9H). ^{13}C NMR (125 MHz, DMSO): δ172.53, 156.27, 155.80, 137.57, 136.28, 128.79, 128.71, 128.48, 128.34, 128.11, 128.00, 127.61, 117.31, 78.65, 69.52, 66.17, 56.20, 35.93, 28.51.

[ring-D_4]-L-Tyr(OBn)-COOBn. Boc-[ring-D_4]-L-Tyr(OBn)-COOBn (1 eq, 50 mg, 0.11 mmol) was treated with TFA (236eq, 2 mL, 26 mmol) in the presence of anisole (16 eq, 0.2 mL, 1.8 mmol) in an ice-bath for 30 min. TFA was then removed by evaporation to give white crystals. (.095 mmol, 34.5 mg, 85.9%) ^{1}H NMR (500 MHz, DMSO): δ8.446 (s, 2H), 7.441-7.255 (m, 10H), 5.061 (s, 2H), 4.318 (t, *J* = 6.5Hz, 1H), 3.110-3.001 (m, 4H). ^{13}C NMR (125 MHz, DMSO): δ171.32, 157.51, 137.23, 136.14, 131.64, 128.77, 128.56, 128.42, 128.23, 128.09, 127.89, 127.53, 116.27, 77.83, 69.73, 53.78, 38.44.

[1-^{13}C]-*N*-Ac-L-Isoleucine. To a solution of L-Isoleucine (1 eq, 300 mg, 2.3 mmol) in NaOH (4M, 3 eq, 1.75 mL, 7 mmol) at 0^{o}C, [1-^{13}C]-acyl chloride (1.1 eq, 178 μL, 2.5 mmol) was added in 5 portions over 50 min. The mixture was acidified to pH 1.5 with HCl and extracted with EtOAc (3 x 10 mL). The organic layers were combined, dried ($MgSO_4$) and concentrated *in vacuo* to give a white powder. (0.58 mmol, 100 mg, 25%) ^{1}H NMR (500 MHz, DMSO): δ7.98 (t, *J* =

5Hz, 1H), 4.17-4.13 (m, 1H), 1.85 (s, 3H), 1.75 (m, 1H), 1.37 (m, 1H), 1.16 (m, 1H), 0.90-0.81 (m, 5H). ^{13}C NMR (125 MHz, DMSO): δ173.55, 169.76, 56.60, 36.67, 25.10, 22.89, 15.96, 11.67.

[1-^{13}C]-*N*-Ac-L-Ile-[ring-D$_4$]-L-Tyr(OBn)-COOBn. [1-^{13}C]-*N*-Ac-L-Isoleucine (1.05 eq, 18.0 mg, 0.104 mmol), [ring-D$_4$]-L-Tyr(OBn)-COOBn (1 eq, 36.1 mg, 0.099 mmol) and HOBt (1.1 eq, 14.7 mg, 0.109 mmol) were dissolved in DMF (10 mL). The solution was cooled in an ice-bath and treated with EDC (1.1 eq, 20.9 mg, 0.109 mmol) and 2,4,6-Trimethylpyridine (1 eq, 13 μL, 0.099 mmol). The mixture was stirred at 0 ^{o}C for 1 hr, allowed to warm to room temperature and stirred overnight. The reaction was diluted with EtOAc, washed with 1N HCl (3 x 30 mL), 1N $NaHCO_3$ (3 x 30 mL) and brine (3 x 30 mL), dried ($MgSO_4$) and concentrated *in vacuo* to give white crystals. (46.3 mg, 0.089 mmol, 89.6%) 1H NMR (500 MHz, DMSO): δ8.42 (d, J = 7Hz, 1H), 7.80 (dd, *J* = 8.75Hz, 1H), 7.43-7.22 (m, 10H), 5.04-4.88 (m, 4H), 4.44 (q, *J* = 7.5Hz, 1H), 4.21 (t, *J* = 6Hz, 1H), 2.96-2.86 (m, 2H), 1.62 (m, 1H), 1.34 (m, 1H), 0.99 (m, 1H), 0.78-0.71 (m, 5H) ^{13}C NMR (125 MHz, DMSO): δ169.29, 157.42, 149.20, 137.57, 128.80, 128.67, 128.21, 128.16, 127.99, 70.51, 69.51, 66.24, 56.67, 54.26, 36.06, 32.81, 24.59, 23.02, 15.49, 11.29.

[1-^{13}C]-*N*-Ac-L-Ile-[ring-D$_4$]-L-Tyr 1-^{13}C-*N*-Ac-L-Ile-[ring-D$_4$]-L-Tyr(OBn)-COOBn (1 eq, 46.3 mg, 0.089 mmol) was dissolved in ethanol (2 mL) in a degassed flask with 10% wt Pd/C. The suspension was stirred under a H_2 atmosphere at room temperature. The resulting mixture was filtered through a pad of celite and concentrated to produce a white solid. (26.3 mg, 0.078 mmol, 88%) ^{1}H NMR

(500 MHz, D_2O): δ4.71 (q, *J* = 6Hz, 1H), 3.81-3.71 (m, 1H), 2.96-2.61 (m, 2H), 2.06 (m, 1H), 2.04 (s, 3H), 1.70 (m, 1H), 1.29 (m, 1H), 1.08 – 1.02 (m, 5H) ^{13}C NMR (125 MHz, D_2O): δ177.34, 170.92, 170.72, 153.00, 136.77, 58.44, 56.92, 33.78, 32.21, 23.78, 14.58, 10.32.

***Synthesis of* [1-^{13}C]-L-Ile-[ring-D_4]-L-Tyr**

***N*-Cbz-[1-^{13}C]-L-Isoleucine.** [1-^{13}C]-L-isoleucine (1 eq, 32.1 mg, 0.24 mmol) was dissolved in ddH_2O (75 μL) and NaOH (5M, 50 μL). The reaction mixture was cooled to 0^{o}C while stirring. Simultaneously, benzyl chloroformate (1.12 eq, 38.5 μL, 0.27 mmol) and NaOH (2M, 1.12 eq, 135 μL, 0.27 mmol) were added drop-wise under argon. The reaction was allowed to warm to room temperature and stirred overnight. After adjusting the pH of the reaction to 10 (saturated aqueous Na_2CO_3), it was extracted with ether (3 x 10 mL). The layers were separated and the aqueous layer was adjusted to pH 3 and again extracted with ether (3 x 10 mL). The organic layers were combined, dried ($MgSO_4$) and concentrated *in vacuo* to give a clear oil. (50.6 mg, 0.20 mmol, 85%) ^1H NMR (500 MHz, DMSO): δ7.46 (d, *J* = 8.5Hz, 1H), 7.37 (s, 5H), 5.02 (s, 2H), 3.89 (q, *J* = 6Hz, 1H), 1.76 (m, 1H), 1.38 (m, 1H), 1.19 (m, 1H), 0.86-0.82 (m, 5H). ^{13}C NMR (125 MHz, DMSO): δ173.67, 156.69, 137.42, 128.71, 128.41, 128.17, 128.11, 126.99, 126.76, 65.78, 63.26, 36.44, 25.02, 15.99, 11.65.

***N*-Cbz-[1-^{13}C]-L-Ile-[ring-D_4]-L-Tyr(OBn)-COOBn.** N-Cbz-[1-13C]-L-Isoleucine (1.05 eq, 25.8 mg, 1.04 mmol), [ring-D_4]-L-Tyr(OBn)-COOBn (1 eq, 36.1 mg, 0.099 mmol) and HOBt (1.1 eq, 14.7 mg, 0.109 mmol) were dissolved in DMF (10 mL). The solution was cooled in an ice-bath and treated with EDC (1.1 eq,

20.9 mg, 0.109 mmol) and 2,4,6-Trimethylpyridine (1 eq, 13 μL, 0.099 mmol). The mixture was stirred at 0 ^{o}C for 1 hr, allowed to warm to room temperature and stirred overnight. The reaction was diluted with EtOAc, washed with 1N HCl (3 x 30 mL), 1N $NaHCO_3$ (3 x 30 mL) and brine (3 x 30 mL), dried ($MgSO_4$) and concentrated *in vacuo* to give a white powder. (52.8 mg, 0.089 mmol, 89.6%) 1H NMR (500 MHz, DMSO): δ8.15 (d, J = 8.5Hz, 1H), 7.94 (d, J = 8Hz, 1H), 7.70-7.20 (m, 15H), 5.16-4.99 (m, 6H), 4.97 (q, J = 6Hz, 1H), 3.93-3.89 (m, 1H), 2.97-2.86 (m, 2H), 2.06 (m, 1H), 1.63 (m, 1H), 1.39 (m, 1H), 1.09-1.02 (m, 5H) 13C NMR (125 MHz, DMSO): δ170.81, 166.76, 155.93, 136.72, 135.88, 131.13, 128.93, 128.88, 128.73, 128.69, 128.48, 128.13, 70.51, 69.51, 55.95, 54.23, 33.78, 22.21, 15.48, 11.79.

[1-^{13}C]-L-Ile-[ring-D_4]-L-Tyr *N*-Cbz-[1-^{13}C]-L-Ile-[ring-D_4]-L-Tyr(OBn)-COOBn (1 eq, 52.8 mg, 0.089 mmol) was dissolved in ethanol (2 mL) in a degassed flask with 10% wt Pd/C. The suspension was stirred under a H_2 atmosphere at room temperature. The resulting mixture was filtered through a pad of celite and concentrated to produce a white solid. (23.4 mg, 0.078 mmol, 88%) ^{1}H NMR (500 MHz, D_2O): δ4.70 (q, *J* = 6Hz, 1H), 3.81-3.71 (m, 1H), 2.96-2.61 (m, 2H), 2.06 (m, 1H), 1.70 (m, 1H), 1.30 (m, 1H), 1.08 – 1.02 (m, 5H) ^{13}C NMR (125 MHz, D_2O): δ177.34, 170.92, 153.00, 136.77, 58.44, 56.92, 33.78, 32.21, 23.78, 14.58, 10.32.

Fermentation

The K-26 producing strain, *Astrosporangium hypotensionis,* was obtained from the Agricultural Research Service (NRRL 12379). The production protocol of Yamato and coworkers was followed with minor modifications. Seed medium consisted of dextrose, 0.1 g/L; Difco soluble starch, 0.1 g/L; Bacto beef extract, 0.05 g/L; Bacto yeast extract, 0.05 g/L; Bacto tryptone, 0.05 g/L; and $CaCO_3$, 0.02 g/L dissolved in distilled water and was adjusted to pH 7.2 before autoclaving. Fermentation was initiated by aseptically inoculating one loop of mycelia grown on an agar plate into a sterile 50 mL Falcon tube containing 10 mL seed medium (Phase I). The Falcon tube was incubated for 10-12 days at 28 ^{o}C in a shaker incubator. For Phase II, 3 mL of the Phase I seed culture were transferred into a 250 mL flask containing 30 mL seed medium. The Phase II seed culture was incubated for 3-4 days at 28 ^{o}C in a shaker incubator. For Production, 30 mL of Phase II seed culture was inoculated into a 2800 mL Fernbach flask containing 300 mL production medium. The flask was incubated for 5-6 days at 28^{o}C in a shaker incubator. The production medium consisted of Difco soluble starch, 0.4 g/L; soybean meal (Whole Foods), 0.3 g/L; corn steep liquor, 0.05 g/L; K_2HPO_4, 5 mg/L; $MgSO_4$, 2.4 mg/mL; KCl, 3 mg/L and $CaCO_3$, 0.03 g/L dissolved in distilled water and was adjusted to pH 7.8 before autoclaving.

Pulse feeding experiments

1 mM of each precursor (dipeptides and tyramine) was dissolved in 5 mL water and administered separately to Phase III culture through a sterile syringe filter in every 24 hours for 4 days.

K-26 Extraction

The production culture was centrifuged. The cell paste was washed with 20% glycerol and the mycelia were frozen in liquid nitrogen for storage at -80 ^{o}C, while the supernatant was acidified to pH 3.0. To one liter of supernatant was added thirty grams of previously activated Diaion HP-20 resin. The suspension was stirred for 30 minutes. The resin was filtered and washed with water. The washed resin beads were then stirred in 50% methanol-50% water solution for 10 minutes. The suspension was filtered, washed with 50% methanol and discarded. The filtrate was concentrated to 40 mL and neutralized. The filtrate was then additionally concentrated to approximately 1 mL.

K-26 Precursor Incorporation Mass Spectrometry

General

Mass spectrometry was performed using ThermoFinnigan (San Jose, CA) TSQ® Quantum triple quadrupole mass spectrometer equipped with a standard electrospray ionization source outfitted with a 100-µm I.D. deactivated fused Si capillary. Data acquisition and spectral analysis were conducted with Xcalibur™ Software, version 1.3, from ThermoFinnigan (San Jose, CA), on a Dell Optiplex

GX270 computer running the Microsoft® Windows 2000 operating system. The source spray head was oriented at an angle of 90^{o} to the ion transfer tube. Nitrogen was used for both the sheath and auxiliary gas. The sheath and auxiliary gases were set to 33 and 14 (arbitrary units) respectively. Samples were introduced by HPLC. A Surveyor® Autosampler and a Surveyor® MS Pump from ThermoFinnigan (San Jose, CA) were used. The injection volume was 10 µL. K-26 was separated from co-metabolites using a Jupiter™ minibore 5µm C18 column (2.0 mm × 15 cm) with a linear water-acetonitrile gradient (ranging from 95:5 to 5:95 H2O:CH3CN) containing 10 mM ammonium acetate. The flow rate was 0.2 mL/min.

Ring-D_4-tyramine

The mass spectrometer was operated in the negative ion mode and the electrospray needle was maintained at 4200 V. The ion transfer tube was operated at -35V and 350^{o}C. The tube lens voltage was set to -150 V. Source CID (offset voltage between skimmer and the first ion guide, Q00) was used at 15 V. The selected reaction monitoring (SRM) mode was used. Ions were collisionally activated with argon at an indicated pressure of 1.4 mT. The mass spectral resolution was set to a peak width (full width at half maximum, FWHM) of 0.50 u and 0.50 u for precursor and product ions respectively. Mass transitions at the specified collision energy (m/z 534→216; 35 eV), (m/z 538→216; 35 eV) and (m/z 538→220; 35 eV) were monitored for unenriched K-26, enrichment in the central tyrosine and enrichment in AHEP, respectively. The scan width for

product ions was 1.000 u and the cycle time for each ion was 0.25 seconds. The electron multiplier gain was set to 2×10^6. Data were acquired in profile mode. Both a ring-D_4-tyramine enriched sample and an unenriched sample were scanned. The mass spectral data for the two samples was almost identical, indicating no D_4 enrichment of K-26.

[1-^{13}C]-*N*-Ac-L-Ile-[ring-D_4]-L-Tyr and [1-^{13}C]-L-Ile-[ring-D_4]-L-Tyr.

The mass spectrometer was operated in the same manner as for the [ring-D_4]-tyramine experiments. Transitions (m/z 534→216; 35 eV), (m/z 539→216; 35 eV) and (m/z 539→220; 35 eV) were monitored for unenriched K-26, +5 enrichment in *N*-Ac-Ile-Tyr and +1 enrichment in *N*-Ac-Ile-Tyr/+4 enrichment in AHEP. To verify +1 and +4 enrichment, additional transitions were monitored. Transitions (m/z 538→216; 35 eV) and (m/z 538→220; 35 eV) were monitored for +4 enrichment in *N*-Ac-Ile-Tyr and AHEP, respectively, while transitions (m/z 534→379; 35 eV), (m/z 535→379; 35 eV) and (m/z 535→380; 35 eV) were monitored for unenriched K-26, +1 enrichment in *N*-Ac-Ile and +1 enrichment in Tyr-AHEP. An unenriched sample was used to create theoretical curves of ratios of mass isotopomer abundance in K-26 to correct the isotopomer distribution of the enriched sample as described previously.

Cloning, Expression and Purification of N-Acetyltransferases

ORF6213

N-acetyltransferase ORF6213 was cloned from the genomic DNA of *A. hypotensionis* by PCR using the primers 5'-GCATATGCGGGTCTCGTTCATC-3' and 5'-CGTTCGAAGATTAACAAG-3'. The 546 bp PCR product was cloned into the vector pCR®2.1-TOPO® using TOPO TA Cloning® Kit (Invitrogen). The gene was then subcloned into the expression vector pET28a(+) using NdeI/HindIII restriction sites to express N-His_6-tagged protein. The resulting plasmid was introduced into E.coli BL21(DE3) (Novagen) and the protein was expressed as following: in a 2.8-L baffled flask, 2x 500 mL LB medium containing 50 µg/mL Kanamycin was inoculated from 2x 5mL overnight cultures made from a fresh colony of the above strain and grown at 37°C until OD_{600} = ~0.6. IPTG was then added (final concentration of 0.5 mM) and the cultures were grown overnight at 25 °C.

The 20.0kD His_6-tagged protein was purified as following: the cells were pelleted (30 min, 3750 rpm, 4 °C), resuspended in binding buffer (50 mM Tris-HCl pH 8 containing 300 mM NaCl and 10 mM imidazole) . For cell lysis, DNase I (NEB, 0.2 U/mL) was added. The cells were broken using French pressure cell (at 1500 psi). The protein was affinity purified on a HisTrap FF column on an ÄKTA chromatography system (GE Healthcare) using binding buffer with linearly increasing imidazole concentration (10–500 mM). The pure protein was then desalted with a HiTrap Desalting column on the same ÄKTA chromatography

system using the assay buffer (20 mM Tris, pH7.5) and stored in aliquots at -80 °C with 5% glycerol and 1 mM DTT added.

NAT97747

NAT97747 was cloned from the genomic DNA of *A. hypotensionis* by PCR using the primers 5'-CATATGCGTTCTGATGTCACCCTGC-3' and 5'-AAGCTTCAGCCGGCGAGGTCCAC-3'. The 489 bp PCR product was cloned into the vector pCR®2.1-TOPO® using TOPO TA Cloning® Kit (Invitrogen). The gene was then subcloned into the expression vector pET28a(+) using NdeI/HindIII restriction sites to express N-His_6-tagged protein. The resulting plasmid was introduced into *E.coli* BL21(DE3) (Novagen) and the ~20kD protein was expressed and purified similarly to ORF6213.

N-Acetyltransferase Assay

Protein fractions (30μL) were incubated with 1 mM AcCoA and 1 mM dK-26 in 20 mM Tris-HCl pH 8 at a total volume of 90 μL at 30 ^{o}C for 4 hours. After incubation, D_3-K-26 was added at a 10 nM final concentration. For kinetic experiments, dK-26 concentrations ranged from 0.5 mM to 5 mM. The reaction was quenched by addition of an equal volume of methanol and subjected to mass spectrometry analysis.

N-Acetyltransferase Assay Mass Spectrometry

Mass spectrometry was performed using a ThermoFinnigan (San Jose, CA) TSQ Quantum triple quadrupole mass spectrometer equipped with a standard electrospray ionization source outfitted with a 100-μm I.D. deactivated fused Si capillary. The mass spectrometer was operated as above. A Surveyor Autosampler and a Surveyor MS Pump from ThermoFinnigan were used. The injection volume was 10 μL. Peptides were separated using a Luna minibore 3 μm C18(2) column (2.0 mm x 5 cm) with an isocratic mobile phase of 90% Buffer A, 10% Buffer B (Buffer A: 95% water, 5% acetonitrile 10mM ammonium acetate; Buffer B: 5% water, 95% acetonitrile 10 mM ammonium acetate) with a flow rate of 0.2 mL/min. Mass transitions at the specified collision energy (m/z 492→216; 20 eV), (m/z 534→216; 20 eV) and (m/z 537→216; 35 eV) were monitored for dK-26, K-26, and D_3-K-26, respectively.

Adenylating Enzyme Purification

Protein Purification

Purification was initiated by suspending flash frozen mycelia in general enzyme buffer (GEB: 50 mM Tris-HCl pH 7 containing 10% glycerol, 1mM DTT and 0.2 mM PMSF). Cells were ruptured at 0 ^{o}C using a French pressure cell. The ruptured cells were centrifuged (16300 x g, 30 min, 4 ^{o}C). The nucleic acids were precipitated from the supernatant by adding 5% streptomycin sulfate at 1/5 volume, stirring at 4 ^{o}C for 15 min and centrifuging. Ammonium sulfate was added to the supernatant in portions over 10 min to the desired final saturation

and stirred for 1 hr at 4 ^{o}C and centrifuged. The protein pellets were resuspended in a minimum volume of GEB and dialyzed overnight versus 3 L of GEB at 4 ^{o}C. Protein concentrations were measured with Pierce® BCA Protein Assay Kit (Thermo Scientific).

Q-sepharose. The dialysed protein solution was loaded on a Q-sepharose (Sigma-Aldrich) column equilibrated with 50 mM Tris, pH 7.5 at 5.0 mL/min using a peristaltic pump and eluted using a linear salt gradient 0.0 - 0.5 M NaCl in 50 mM Tris at pH 7.5. Active fractions were combined and concentrated using an Amicon stirred cell.

Hydrophobic Interaction chromatography. A high salt protein solution was loaded on a HiTrap ButylFF column (GE) equilibrated with 50 mM Tris, 1M $(NH_4)_2SO_4$, pH 7.5 using an Akta FPLC at a 1 mL/min flow rate. Proteins were eluted using a linear salt gradient 1.0 – 0.0 M $(NH_4)_2SO_4$ in 50 mM Tris at pH 7.5.

Size exclusion chromatography. Protein solutions were loaded on a Sephacryl S-200 (GE) column equilibrated with GEB at 0.5 mL/min and eluted with GEB.

DEAE Sepharose. Desalted protein solutions were loaded on a DEAE sepharose (GE) column equilibrated with 50 mM Tris, pH 7.2 buffer at 1 mL/min using an Akta FPLC. Proteins were eluted using a linear gradient 0.0 - 0.5 M NaCl in 50 mM Tris at pH 7.2.

MonoQ. Desalted protein solutions were loaded on MonoQ (GE) column equilibrated with 50 mM Tris, pH 7.2 buffer at 1 mL/min using an Akta FPLC. Proteins were eluted using a linear gradient 0.0 - 0.5 M NaCl in 50 mM Tris at pH 7.2.

γ-$^{18}O_4$-ATP-PPi Exchange

Assay Conditions

In order to avoid precipitation of magnesium pyrophosphate, assay components were divided into stock solutions comprising 1) 3 mM amino acids containing 15 mM PPi in 20 mM Tris pH 7.5, 2) 3 mM γ–$^{18}O_4$-ATP containing 15 mM $MgCl_2$ in 20 mM Tris pH 7.5 and 3) protein fractions. Exchange reactions containing 2 μL of each component were initiated by the addition of enzyme solution. 6 μL reactions contained final concentrations of 5 mM $MgCl_2$, 5 mM PP_i, 1 mM γ-$^{18}O_4$-ATP, 1mM amino acid and 20 mM Tris-HCl pH 7.5. After incubation at 25 ^oC, the reactions were stopped by the addition of 6 μL 9-aminoacridine in acetone (10 mg/mL) for MALDI-TOF MS analysis or 6 μL acetone for ESI-LC/MS analysis.

γ-$^{18}O_4$-ATP-PPi Exchange Assay Mass Spectrometry

MALDI-TOF MS analyses were performed on a Voyager-DE™ STR (Applied Biosystems, Inc.) using a nitrogen laser (337 nm). Prior to analysis, 1 μL of analyte/matrix mixture was spotted on to a stainless steel MALDI target. External mass calibration was performed in the reflectron mode using a mixture of γ–$^{18}O_4$-ATP and γ–$^{16}O_4$-ATP. Mass spectra were acquired in the negative ion mode over a range of 450 to 1200 m/z. Mass spectra for ATP-PPi exchange analysis were obtained by averaging 100 consecutive laser shots. Data acquisition and quantitative spectral analysis was conducted using Applied Biosystems DataExplorer software, version 4.5.

ESI-LC/MS analyses were performed on a ThermoFinnigan LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, Waltham, MA) equipped with an ESI interface in negative ion mode. Nitrogen was used both for the auxiliary and sheath gas. The auxiliary and sheath gases were set to 20 psi and 36 psi, respectively. The following instrumental parameters were used: capillary temperature 300 °C; source voltage 4.5 kV; source current 100 μA capillary voltage -49.0 V; tube lens -148.30 V; skimmer offset 0.00 V; activation time 50 ms with an isolation width of 1 *m*/*z*. The sensitivity of the mass spectrometer was tuned by infusion of γ-$^{16}O_4$-ATP at a flowrate of 0.1 mL/min. Samples were introduced by a Waters Acquity UPLC system (Waters, Milford, MA) with an injection volume of 5 μL. γ-$^{18}O_4$-ATP was separated from contaminating salts on a 5 μm Hypercarb column (3 x 50mm, ThermoFisher Scientific) with an isocratic method of 82.5% 20 mM ammonium acetate pH 6 containing 0.1% diethylamine and 17.5% acetonitrile over 5 minutes with a flow rate of 0.2 mL/min.

tRNA Charging Assay

Each assay contained 50 mM HEPES-KOH pH 7, 15 mM $MgCl_2$, 15 mM DTT, 5 mM ATP, 2.5 U Roche total tRNA. For TyrRS charging, the reaction mixture contained 50 μM Tyr with 2 μM L-ring-3,5-^{3}H-Tyr (100 Ci/mmol; Amersham). For IleRS charging, the reaction mixture contained 50 μM Ile with 2.4 μM L-[4,5]-^{3}H-Ile (40 Ci/mmol; PerkinElmer). To initiate the reaction, 3 μL of MonoQ purified protein was added to each reaction. Reactions were incubated at 37 ^oC for 25 min. The reactions were placed on ice, 25 μL of a 10 mg/mL BSA solution (New

England Biosciences) was added followed by 1 mL of cold 10% TCA. After a 10 min incubation at 0v^{o}C, the solution was filtered through nitrocellulose filters (Millipore). The filters were washed twice with 10% TCA and dissolved in FiltronX (NationalDiagnostics, Atlanta, GA) for analysis.

K-26 analogue production by Sphaerosporangium rubeum

Culture Conditions

S. rubeum, was obtained from the Agricultural Research Service (NRRL B-2636). The production protocol of K-26 by *A. hypotensionis* was followed. Seed medium consisted of dextrose, 0.1 g/L; Difco soluble starch, 0.1 g/L; Bacto beef extract, 0.05 g/L; Bacto yeast extract, 0.05 g/L; Bacto tryptone, 0.05 g/L; and $CaCO_3$, 0.02 g/L dissolved in distilled water and was adjusted to pH 7.2 before autoclaving. Fermentation was initiated by aseptically inoculating one loop of mycelia grown on an agar plate into a sterile 50 mL Falcon tube containing 10 mL seed medium (Phase I). The Falcon tube was incubated for 10-12 days at 28^{o}C in a shaker incubator. For Phase II, 3 mL of the Phase I seed culture were transferred into a 250 mL flask containing 30 mL seed medium. The Phase II seed culture was incubated for 3-4 days at 28 ^{o}C in a shaker incubator. For Production, 30 mL of Phase II seed culture was inoculated into a 2800 mL Fernbach flask containing 300 mL of production media. Individual production cultures, containing one of six possible production media, were incubated for 5-6 days at 28^{o}C in a shaker incubator. The production media included K26P: Difco soluble starch, 0.4 g/L; soybean meal (Whole Foods), 0.3 g/L; corn steep liquor,

0.05 g/L; K_2HPO_4, 5 mg/L; $MgSO_4$, 2.4 mg/mL; KCl, 3 mg/L and $CaCO_3$, 0.03 g/L dissolved in distilled water and adjusted to pH 7.8 before autoclaving; BA: Soybean powder, 15 g/L; Glucose, 10 g/L; Difco soluble starch, 10 g/L; NaCl, 3 g/L; $MgSO_4 \cdot 7H_2O$, 1 g/L; K_2HPO_4, 1 g/L and 1 mL trace element solution ($FeSO_4 \cdot 7H_2O$, 0.01 g/L; $MnCl_2 \cdot 4H_2O$, 0.08 g/L; $CuSO_4 \cdot 5H_2O$, 0.07 g/L; $ZnSO_4 \cdot 7H_2O$, 0.02 g/L; 1 drop H_2SO_4) dissolved in distilled water and adjusted to pH 7.2 before autoclaving; EA: lactose, 50 g/L; corn steep solids, 5 g/L; glucose, 5 g/L; glycerol, 15 g/L; soybean flour, 10 g/L; Bacto peptone, 5 g/L; $CaCO_3$, 3 g/L; $(NH_4)_2SO_4$, 2 g/L; $FeSO_4 \cdot 7H_2O$, 0.1 g/L; $ZnCl_2$, 0.1 g/L; $MnCl_2 \cdot 4H_2O$, 0.1 g/L and $MgSO_4 \cdot 7H_2O$, 0.5 g/L dissolved in distilled water and adjusted to pH 7.2 before autoclaving; OA: glucose, 10 g/L; glycerol, 5 g/L; corn steep liquor, 3 g/L; beef extract, 3 g/L; malt extract, 3 g/L; yeast extract, 3 g/L and $CaCO_3$, 2 g/L are dissolved in distilled water, adjusted to pH 7.2 and after autoclaving, 0.1 g/L thiamine is added; KA: glucose, 10 g/L; corn steep liquor, 10 g/L; soybean powder, 10 g/L; glycerol, 5 g/L; dry yeast, 5 g/L and NaCl, 5 g/L are dissolved in distilled water, the pH is adjust to 5.7 and 2 g/L $CaCO_3$ is added before autoclaving; QB: soluble starch, 5 g/L; glucose, 6 g/L; corn steep liquor, 2.5 g/L; Proflo powder, 5 g/L and Proflo oil, 2 mL/L are dissolved in distilled water, the pH is adjusted to 7.2 and autoclaved.

K-26 analogue extraction

The production culture was centrifuged. The cell paste was discarded and the supernatant was acidified to pH 3.0. To one liter of supernatant was added

thirty grams of previously activated Diaion HP-20 resin. The suspension was stirred for 30 minutes. The resin was filtered and washed with water. The washed resin beads were then stirred in 50% methanol-50% water solution for 10 minutes. The suspension was filtered, washed with 50% methanol and discarded. The filtrate was concentrated to 40mL and neutralized. The filtrate was then additionally concentrated to approximately 1mL.

K-26 analogue mass spectrometry

Mass spectrometry was performed using FisherScientific (San Jose, CA) TSQ® Quantum Access triple quadrupole mass spectrometer equipped with a standard electrospray ionization source outfitted with a 100-µm I.D. deactivated fused Si capillary. Data acquisition and spectral analysis were conducted with Xcalibur™ Software, The source spray head was oriented at an angle of 90^{o} to the ion transfer tube. Nitrogen was used for both the sheath and auxiliary gas. The sheath and auxiliary gases were set to 33 and 14 (arbitrary units) respectively. Samples were introduced by HPLC. A Surveyor® Autosampler and a Surveyor® MS Pump from ThermoFinnigan (San Jose, CA) were used. The injection volume was 10 µL. K-26 and analogues were separated from co-metabolites using a Jupiter™ minibore 5µm C18 column (2.0 mm × 15 cm) with a linear water-acetonitrile gradient (ranging from 95:5 to 5:95 H2O:CH3CN) containing 10 mM ammonium acetate. The flow rate was 0.2 mL/min.

The mass spectrometer was operated in the negative ion mode and the electrospray needle was maintained at 4200 V. The ion transfer tube was

operated at -35 V and 350 ^{o}C. The tube lens voltage was set to -150 V. Source CID (offset voltage between skimmer and the first ion guide, Q00) was used at 15 V. Precursor ion scanning was used. Ions were collisionally activated with argon at an indicated pressure of 1.4 mT and a product ion of m/z 216 was chosen. The mass spectral resolution was set to a peak width (full width at half maximum, FWHM) of 0.50 u and 0.50 u for precursor and product ions respectively. The scan width for product ions was 1.000 u and the cycle time for each ion was 0.25 seconds. The electron multiplier gain was set to 2×10^{6}. Data were acquired in profile mode. Three molecules were identified corresponding to K-26, SF2513B and SF2513C in BA and EA media, while only K-26 was identified in OA and KA. Production cultures in QB and K26P media did not produce any K-26 or analogues.

REFERENCES

1. Yamato, M., Koguchi, T., Okachi, R., Yamada, K., Nakayama, K., Kase, H., Karasawa, A. and Shuto, K. (1986) K-26, a novel inhibitor of angiotensin I converting enzyme produced by an actinomycete K-26. *Journal of Antibiotics (Tokyo)*, **39**, 44-52.

2. Ntai, I. and Bachmann, B. O. (2008) Identification of ACE pharmacophore in the phosphonopeptide metabolite K-26. *Bioorganic & Medicinal Chemistry Letters*, **18**, 3068-3071.

3. Koguchi, T., Yamada, K., Yamato, M., Okachi, R., Nakayama, K. and Kase, H. (1986) K-4, a novel inhibitor of angiotensin I converting enzyme produced by Actinomadura spiculosospora. *Journal of Antibiotics (Tokyo)*, **39**, 364-371.

4. Liu, S. J., Lu, Z. B., Jia, Y., Dunaway-Mariano, D. and Herzberg, O. (2002) Dissociative phosphoryl transfer in PEP mutase catalysis: Structure of the enzyme/sulfopyruvate complex and kinetic properties of mutants. *Biochemistry*, **41**, 10270-10276.

5. Mcqueney, M. S., Lee, S. L., Swartz, W. H., Ammon, H. L., Mariano, P. S. and Dunaway-mariano, D. (1991) Evidence for an Intramolecular, Stepwise Reaction Pathway for Pep Phosphomutase Catalyzed P-C Bond Formation. *Journal of Organic Chemistry*, **56**, 7121-7130.

6. Metcalf, W. W. and van der Donk, W. A. (2009) Biosynthesis of Phosphonic and Phosphinic Acid Natural Products. *Annual Review of Biochemistry*, **78**, 65-94.

7. Bowman, E., Mcqueney, M., Barry, R. J. and Dunaway-mariano, D. (1988) Catalysis and Thermodynamics of the Phosphoenolpyruvate Phosphonopyruvate Rearrangement - Entry into the Phosphonate Class of Naturally-Occurring Organo-Phosphorus Compounds. *Journal of the American Chemical Society*, **110**, 5575-5576.

8. Bowman, E. D., McQueney, M. S., Scholten, J. D. and Dunaway-Mariano, D. (1990) Purification and characterization of the Tetrahymena pyriformis P-C bond forming enzyme phosphoenolpyruvate phosphomutase. *Biochemistry*, **29**, 7059-7063.

9. Ntai, I., Manier, M. L., Hachey, D. L. and Bachmann, B. O. (2005) Biosynthetic origins of C-P bond containing tripeptide K-26. *Organic Letters*, **7**, 2763-2765.

10. Ntai, I., Phelan, V. V. and Bachmann, B. O. (2006) Phosphonopeptide K-26 biosynthetic intermediates in Astrosporangium hypotensionis. *Chemical Communications*, 4518-4520.

11. Boger, D. L. and Yohannes, D. (1987) Selectively Protected L-Dopa Derivatives - Application of the Benzylic Hydroperoxide Rearrangement. *Journal of Organic Chemistry*, **52**, 5283-5286.

12. Carpino, L. A., Elfaham, A. and Albericio, F. (1995) Efficiency in Peptide Coupling - 1-Hydroxy-7-Azabenzotriazole Vs 3,4-Dihydro-3-Hydroxy-4-Oxo-1,2,3-Benzotriazine. *Journal of Organic Chemistry*, **60**, 3561-3564.

13. Chen, C., Zhu, Y. F. and Wilcoxen, K. (2000) An improved synthesis of selectively protected L-Dopa derivatives from L-tyrosine. *Journal of Organic Chemistry*, **65**, 2574-2576.

14. Chenault, H. K., Dahmer, J. and Whitesides, G. M. (1989) Kinetic Resolution of Unnatural and Rarely Occurring Amino-Acids - Enantioselective Hydrolysis of N-Acyl Amino-Acids Catalyzed by Acylase-I. *Journal of the American Chemical Society*, **111**, 6354-6364.

15. Ecija, M., Diez, A., Rubiralta, M., Casamitjana, N., Kogan, M. J. and Giralt, E. (2003) Synthesis of 3-aminolactams as X-Gly constrained pseudodipeptides and conformational study of a Trp-Gly surrogate. *Journal of Organic Chemistry*, **68**, 9541-9553.

16. Walker, D. M., Mcdonald, J. F., Franz, J. E. and Logusch, E. W. (1990) Design and Synthesis of Gamma-Oxygenated Phosphinothricins as Inhibitors of Glutamine-Synthetase. *Journal of the Chemical Society-Perkin Transactions 1*, 659-666.

17. Finking, R. and Marahiel, M. A. (2004) Biosynthesis of nonribosomal peptides. *Annual Review of Microbiology*, **58**, 453-488.

18. Challis, G. L., Ravel, J. and Townsend, C. A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry & Biology*, **7**, 211-224.

19. Stachelhaus, T., Mootz, H. D. and Marahiel, M. A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology*, **6**, 493-505.

20. Ntai, I., Vanderbilt University, Nashville, 2007.

21. Fujikawa, K., Sakamoto, Y. and Kurahash.K (1971) Biosynthesis of Tyrocidine by a Cell-Free Enzyme System of Bacillus-Brevis Atcc 8185 .3. Further Purification of Components I and Ii and Their Functions in Tyrocidine Synthesis. *Journal of Biochemistry*, **69**, 869-&.

22. Keller, U. and Schlumbohm, W. (1992) Purification and Characterization of Actinomycin Synthetase-I, a 4-Methyl-3-Hydroxyanthranilic Acid-Amp Ligase from Streptomyces-Chrysomallus. *Journal of Biological Chemistry*, **267**, 11745-11752.

23. Lee, S. G. and Lipmann, F. (1975) Tyrocidine synthetase system. *Methods in Enzymology*, **43**, 585-602.

24. Schwecke, T., Aharonowitz, Y., Palissa, H., Vondohren, H., Kleinkauf, H. and Vanliempt, H. (1992) Enzymatic Characterization of the Multifunctional Enzyme Delta-(L-Alpha-Aminoadipyl)-L-Cysteinyl-D-Valine Synthetase from Streptomyces-Clavuligerus. *European Journal of Biochemistry*, **205**, 687-694.

25. Janc, J. W., Egan, L. A. and Townsend, C. A. (1995) Purification and characterization of clavaminate synthase from Streptomyces antibioticus. A multifunctional enzyme of clavam biosynthesis. *The Journal of Biological Chemistry*, **270**, 5399-5404.

26. Forster, A. C. (2009) Low modularity of aminoacyl-tRNA substrates in polymerization by the ribosome. *Nucleic Acids Research*, **37**, 3747-3755.

CHAPTER IV

BIOSYNTHETIC INVESTIGATIONS OF ANTHRAMYCIN

First described in the early 1940s and structurally characterized in 1963, anthramycin is an antitumor and antibiotic peptide natural product.[1] Produced by the thermophilic actinomycete *Streptomyces refuineus* sbsp *thermotolerans*, anthramycin is a member of the pyrrolobenzodiazepine (PBD) class of natural products[2]. (Figure IV-1) These natural products are distinguished by a tricyclic ring system formed by an anthranilate (A), a diazepine (B) and a hydropyrrole (C) moiety. Chemical diversity between members of this class is established by the varied substitution patterns of the A- and C- rings. Additional diversity can be found within the C-ring including unsaturation at the C2-C3 bond and/or exocyclic unsaturated moiety at the C2 position. The chemical diversity decorating the pyrrolobenzodiazepine core modulates the biological potency of these compounds, where non-covalent interactions with TA-rich regions in the minor groove of double stranded DNA and the irreversible covalent modification of the N^2-amino group of a neighboring guanine provide the general mechanism for both the antitumor and antibacterial activities.[3-12]

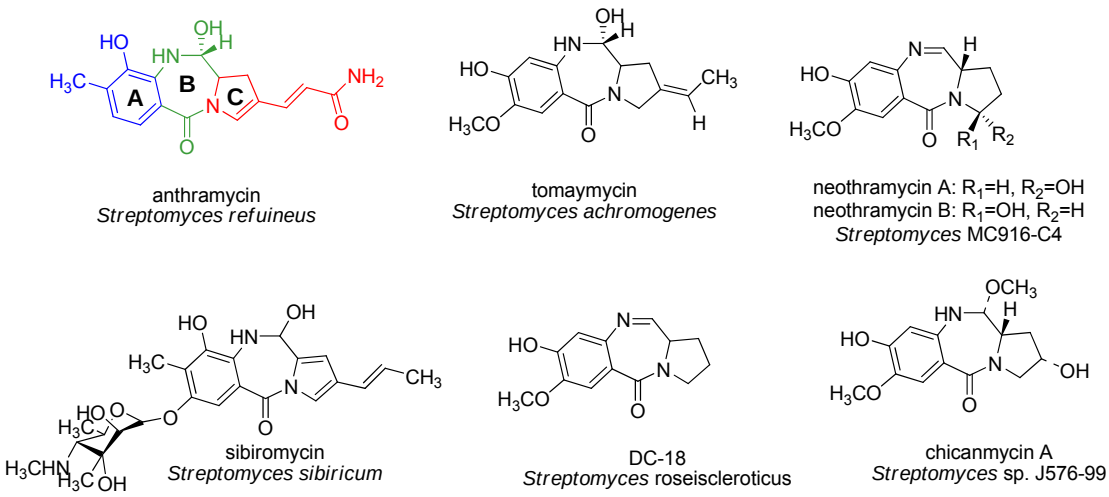


Figure 43. Representative Examples of Bacterial PBDs. These natural products are distinguished by a tricyclic ring system formed by an anthranilate (A), a diazepine (B) and a hydropyrrole (C) moiety. Chemical diversity between members of this class is established by the varied substitution patterns of the A- and C-rings.

Before our investigations into the biosynthesis of anthramycin, little was known about the biosynthesis of PBD natural products. Previous investigations by Hurley and coworkers established the amino acid precursors to both the anthranilate and hydropyrrole hemispheres of anthramycin, sibiromycin and tomaymycin.[2, 13-16] (Figure IV-2) Both hemispheres of anthramycin are derived from oxidative ring opening of amino acids: the 3-hydroxyanthranilic acid (HA) portion of 4-methyl-3-hydroxyanthranilic acid (MHA) hemisphere is derived from tryptophan while the dehydropyrrole hemisphere is derived from tyrosine. Methylation of HA and the hydropyrrole moiety are both through methionine.

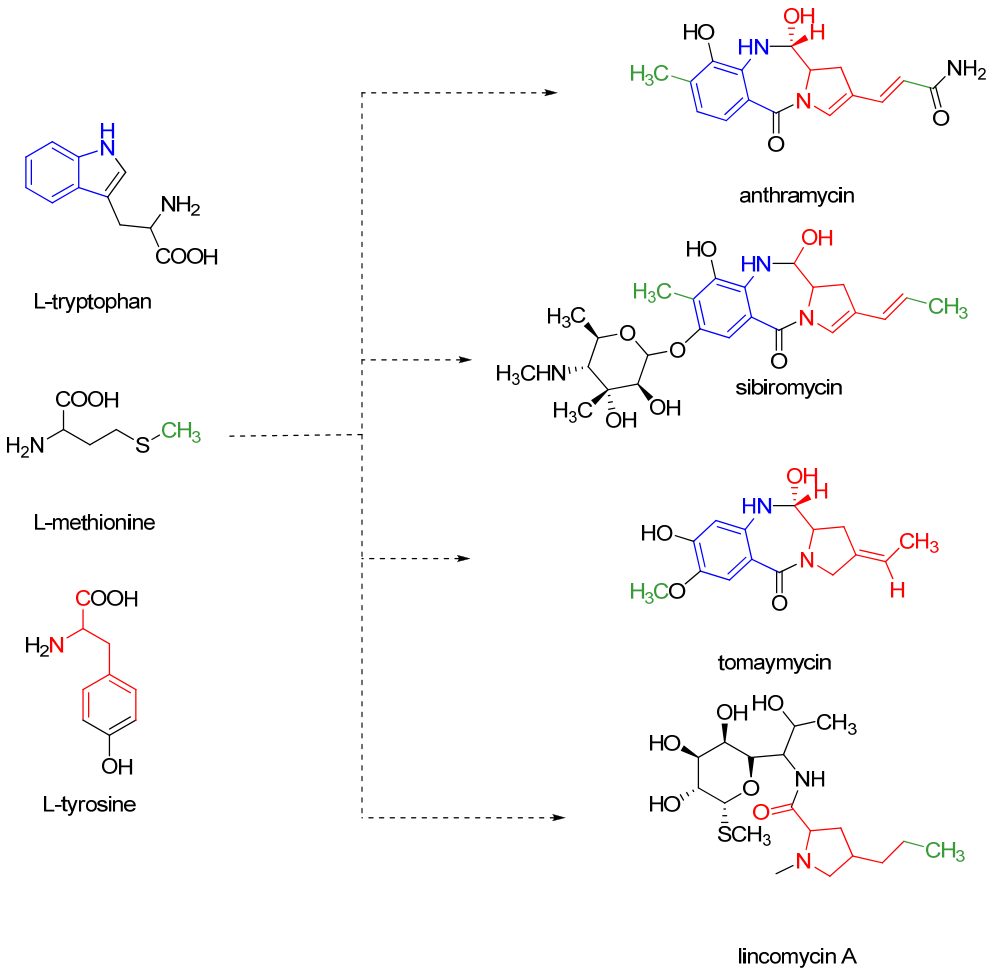


Figure IV-2. Precursor Incorporation Studies of PBDs. The amino acid precursors of the PBDs have been established. The A-ring of the PBDs is derived from tryptophan, while the C-ring is derived from tyrosine. Methylation of the core structure is methionine dependent.

From these results, Hurley proposed that the MHA portion of anthramycin is derived from tryptophan via the kynurenine primary metabolic pathway to give HA followed by *S*-adenosylmethionine (SAM) dependent aromatic C-methylation. Additionally, it was proposed that the dehydropyrrole hemisphere is derived from tyrosine via a biosynthetic pathway analogous to the biosynthesis of the antibiotic lincomycin, which contains a 4-propyl-4,5-dehydroproline moiety, through oxidative ring opening of L-DOPA. Incorporation studies with CH_3-labeled

methionine into the C-terminal carboxamide of anthramycin and the terminal methyl group of the propyl side chain of lincomycin provide additional evidence of parallel biosynthesis of anthramycin and lincomycin.

Results

Biosynthetic hypothesis

Based on the non-proteinogenic nature of the two amino acids proposed in the formation anthramycin, evidenced by the aforementioned isotopic incorporation experiments, we proposed an NRPS based biosynthetic hypothesis with a proposed domain string of A-T-C-A-T-Re. The first A-domain of the NRPS would activate either HA or MHA and the second A-domain would activate a dehydroproline acrylamide moiety for loading on to their respective T-domains. A specialized reductase domain would catalyze the final cyclization to form the benzodiazepine ring of anthramycin. Genes homologous to those found with the kynurenine primary metabolic pathway will provide the biosynthetic route for HA biosynthesis, while those homologous to the biosynthesis 4-propyl-4,5-dehydroproline moiety of lincomycin will synthesize the dehydroproline moiety found in anthramycin.

Genetic Identification of Anthramycin Gene Cluster

The complete gene cluster for anthramycin biosynthesis was identified by our collaborators, Ecopia Biosciences (Thallion Pharmaceuticals), by rapid genome-scanning.[17] Briefly, a two-tiered genomic DNA library of 1.5-3 kb and 30-

50 kb fragments was constructed from genomic DNA isolated from *S. refuineus* sbsp *thermotolerans.* The short insert library was cloned to give a number of genomic sequence tags by sequencing random clones. These clones were analyzed for sequence similarity to verified secondary metabolism gene products in the National Center for Biotechnology Information (NCBI) nonredundant protein database. A single genome sequence tag was identified containing sequence similarity to the predicted requisite NRPS and used in colony hybridization, identifying two cosmids, 024CA and 024CO, generated from the large fragment DNA library. These two overlapping cosmids were sequenced providing a putative gene cluster for anthramycin biosynthesis harboring 25 genes including those consistent with HA biosynthesis, dehydropyrrole moiety biosynthesis, a bimodular NRPS terminating with a reductase domain and gene candidates consistent with regulation, resistance and transport. (Figure IV-3)

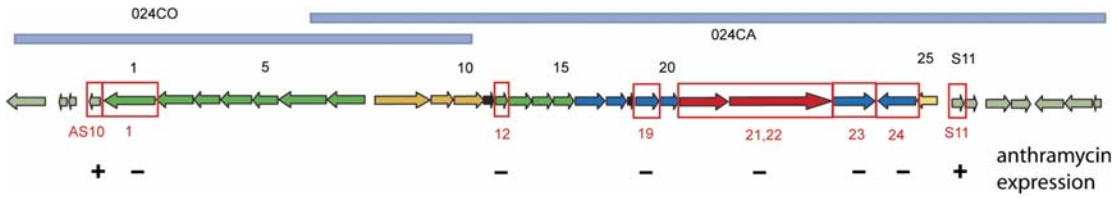


Figure IV-344. The Anthramycin Gene Cluster. The scientists at Ecopia Biosciences identified the gene cluster of anthramycin on two cosmids. This gene cluster contains an NRPS (red), genes for MHA biosynthesis (blue), the dehydroproline acrylamide moiety (green) and regulatory genes (yellow). Gene disruptions of the genes indicated by red boxes were generated. Anthramycin production was not observed in production cultures of the strains with gene disruptions within the cluster. Used with permission.[17]

To establish the boundaries of this gene cluster and to validate its role in anthramycin biosynthesis, a series of PCR-targeted gene replacement experiments were performed by Hu *et al*.[17, 18] Developing and using a variation

of the λ-RED recombination system, a series of mutants was created including disruptions within *orfAS10, orf1, orf12, orf19, orf21/orf22, orf23, orf24* and *orfS11.* (Figure IV-3) *Orf1* and *orf12* are proposed to be involved in dehydropyrrole moiety biosynthesis, while *orf19, orf23* and *orf24* are proposed to be involved in MHA biosynthesis. (Table IV-1) *Orf21* and *orf22* encode the NRPS modules of the biosynthetic pathway with a domain string of A-T (orf21) C-A-T-Re (orf22) and *orfAS10* and *orfS11* are genes flanking the anthramycin gene cluster. (Figure IV-3) Production of anthramycin via TLC-bioautography (anti-*Bacillus*) and LC/MS was observed in replacement mutants of *orfAS10* and *orfS11*, the genes flanking either side of the anthramycin gene cluster, while anthramycin production was not observed in mutants within the proposed gene cluster. This verified the identification and isolation of the anthramycin biosynthetic gene cluster.

Table IV-1. ORFs Found Within the Anthramycin Gene Cluster

ORF	#AA	GenBank Homology	% Identity/Similarity	Proposed Function
AS10	138	AAU19321.1	46/68	Unknown, hypothetical protein
1	624	BAB12569.1	57/68	Amidotransferase
2	500	CAD30313.1		Aldehyde dehydrogenase
3	354	EAO60654.1	50/65	Alcohol dehydrogenase
4	410	CAJ23858.1	49/63	Cytochrome P-450 hydroxylase
5	352	487713	79/87	ImbA, methyltransferase
6	621	CAA55746.1	73/80	ImbW, unknown function
7	487	ABF90321.1	38/54	FAD oxidoreductase
8	764	AAL06654.1	59/75	Drug-resistance pump
9	256	CAB55527.1	27/40	Putative hydrolase/glyoxylase
10	377	EAL16816.1	41/66	Transporter
11	89	-	-	None
12	169	CAA55747.1	48/63	ImbB1, L-DOPA 2,3-dioxygenase
13	302	CAA55748.1	42/54	ImbB2, L-tyrosine 3-hydroxylase
14	297	CAA55772.1	50/63	ImbY, unknown
15	276	CAA55771.1	34/41	ImbX, unknown
16	413	42543461	37/57	Kynureninase
17	261	4753870	37/52	Tryptophan 2,3-dioxygenase
18	58	-	-	None
19	348	37542638	43/63	Aromatic C-methyltransferase
20	296	EAU10758.1	36/49	Aryl formamidase
21	600	CAD92850.1	35/51	NRPS
22	1146	ABF90459.1	32/46	NRPS
23	500	ABF39686.1	36/50	Kynurenine 3-monooxygenase
24	475	ABF87356.1	36/50	Flavin-containing oxidoreductase
25	273	BAC79018.1	46/64	Repressor-response regulator
S11	174	BAC69182.1	63/68	Zn-dependent hydrolase
S12	199	EAU12469.1	31/53	Unknown, hypothetical protein

Based on genomic analysis, the NRPS within this cluster consists of a domain string of A-T-C-A-T-Re with an AT didomain on one module (Orf21) and C-A-T-Re (Orf22) on a second module. (Figure IV-4) To begin studying the biochemical mechanisms within anthramycin biosynthesis, we endeavored to identify the amino acid substrates of the A-domains within the NRPS. Based on the 8- to 10- amino acid specificity code involved in amino acid discrimination within the binding pocket of A-domains, we established no similarity between the binding pockets of the A-domains in anthramycin biosynthesis and the specificity

code of A-domains with known substrate binding, including that of actinomycin, an MHA containing natural product.[19]

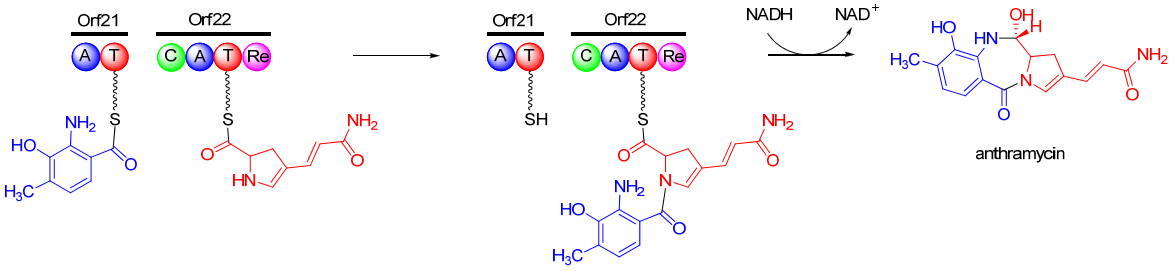
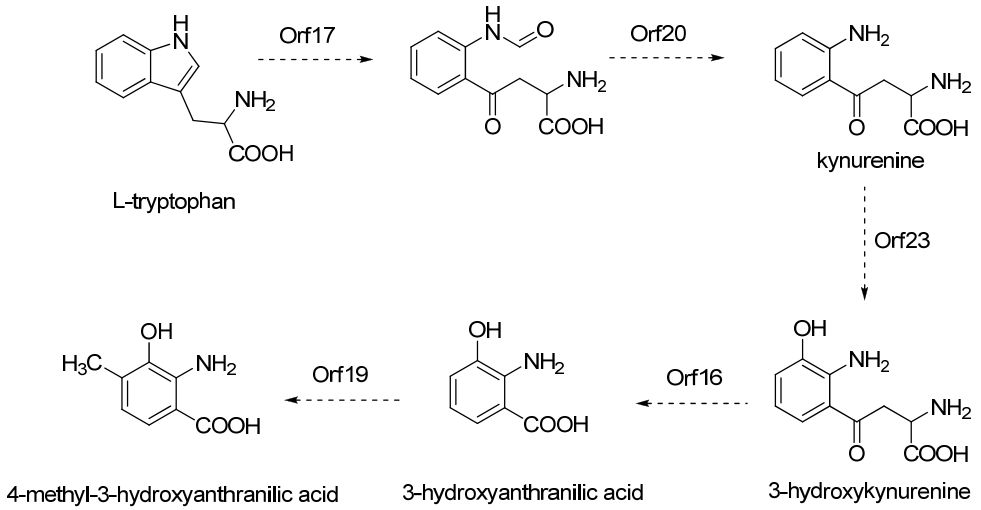


Figure IV-4. Proposed Peptide Biosynthesis in Anthramycin. Based on the structure of anthramycin and the identification of a 2 module, 6 domain NRPS, we propose that anthramycin is condensed from two hemispheres: MHA and a dehydroproline acrylamide moiety. An NADH-dependent reductase forms the benzodiaxepine ring and releases the peptide from the NRPS.

Proposed MHA biosynthesis

Based on homology of primary metabolic genes involved in HA biosynthesis, several genes within the anthramycin gene cluster were putatively assigned roles in MHA biosynthesis. Four genes, *orf16, orf17, orf20* and *orf23*, show high similarity to genes within the biosynthetic gene cluster of the MHA containing natural product actinomycin. The proposed biosynthetic pathway for HA is through oxidative cleavage of L-tryptophan through the kynurenine pathway. (Scheme IV-1) *Orf17* encodes for a tryptophan-2,3-dioxygenase, proposed to catalyse the ring opening of L-tryptophan to give *N*-formylkynurenine. De-formylation of *N*-formylkynurenine by orf20, an aryl formamidase, provides L-kynurenine which is then hydroxylated at the 3 position by orf23, a kynurenine-3-monooxygenase, to give 3-hydroxykynurenine. Orf16 catalyzes the loss of alanine from 3-hydroxykynurenine to yield HA. The

presence of a *C*-methyl substituent at the 4 position of the anthranilate ring, shown to be derived from methionine, indicates that a methionine-dependent *C*-methyltransferase should be present within the anthramycin biosynthetic gene cluster. Two genes within the gene cluster have been putatively assigned as methyltransferases: *orf5* and *orf19*. Orf5 shows high sequence similarity to a *C*-methyltransferase within the lincomycin biosynthetic pathway, which has a similar dehydroproline subunit, indicating that orf5 is probably involved in the biosynthesis of the dehydroproline acrylamide moiety in anthramycin. Based on deductive reasoning, we speculate that Orf19 is the *C*-methyltransferase responsible for appending the methyl group to the 4 position of the anthranilate ring. An additional gene, *orf24*, is clustered with *orf23* and proposed to encode a flavin-dependent or amine oxidase, however, the role of Orf24 remains speculative.

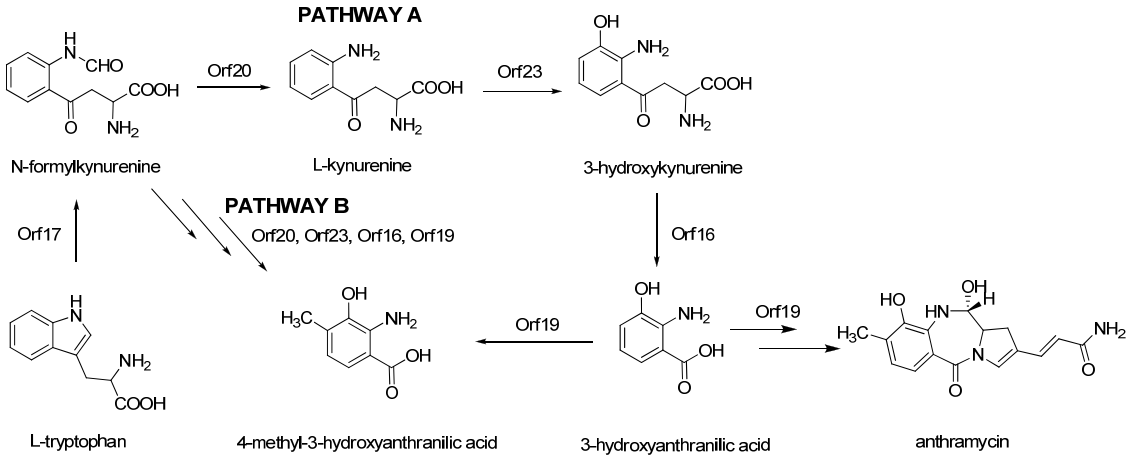


Scheme IV-1. Possible Biosynthetic Pathway for MHA. Previous precursor incorporation studies coupled with the sequence homology of proteins within the anthramycin biosynthetic cassette to proteins involved in primary metabolic kynurenine pathway suggest that MHA may be derived from tryptophan through kynurenine.

Chemical complementation studies to elucidate MHA biosynthesis

While primary metabolic formation of HA as a precursor to nicotinate-derived cofactors is through the linear kynurenine pathway from tryptophan, the order and substrate specificity of the homologous genes within the anthramycin gene cluster may deviate from this paradigm. In order to establish the timing of each step of MHA biosynthesis a series of chemical complementation studies was undertaken. Of particular interest is the timing of *C*-methylation at the 4 position. Isotopic incorporation studies by Hurley and coworkers demonstrated incorporation of HA into anthramycin, while MHA was not incorporated.[2, 13, 14] This may be due to lack of transport of MHA across the cell membrane of *S. refuineus*. However, this hypothesis has not been substantiated experimentally.[20]

There are two general biosynthetic possibilities for the formation of MHA. (Scheme IV-2) In pathway A, *C*-methylation is proposed to occur subsequent to HA formation, via the primary metabolic paradigm. In this case, *C*-methylation can occur directly after HA formation to give MHA before incorporation into the dipeptide or, alternatively, the aryl *C*-methyl can be inserted during or after dipeptide assembly. In pathway B, methylation occurs at an unknown position in the primary anthranilic acid pathway resulting in enzyme substrate divergence downstream.



Scheme IV-2. Possible Biosynthetic Pathways for MHA. Formation of MHA may follow the primary metabolic kynurenine pathway to HA followed by C-methylation either before or after incorporation (Pathway A). Alternatively, the proteins involved in MHA may function in a sequence divergent from primary metabolism if C-methylation occurs before HA formation (Pathway B).

Using *S. refuineus* strains harboring gene disruptions in *orf19*, *orf23 and orf24,* genes homologous to those of HA biosynthesis, precursor incorporation studies were performed.[17] Four possible biosynthetic intermediates: L-kynurenine, 3-hydroxy- L-kynurenine, HA and MHA were added to growing cultures of the *S. refuineus* mutants to test for their ability to complement the three gene-replacement mutants to produce anthramycin. Butanolic extracts prepared from production cultures of the gene-replacement mutants grown were subjected to TLC-bioautography and LC/MS analysis. Both 3-hydroxy- L-kynurenine and HA were shown to restore benzodiazepine production when exogenously supplementing all three mutants, while neither L-kynurenine nor MHA were able to complement any of the mutants to produce anthramycin. (Table IV-2)

Table IV-2. Results from Chemical Complementation Experiments. Anthramycin = m/z 316; desmethylanthramycin = m/z 302

	L-kynurenine		3-hydroxykynurenine		3-hydroxyanthranilic acid		4-methyl-3-hydroxyanthranilic acid	
Orf	**Th**	**Obs**	**Th**	**Obs**	**Th**	**Obs**	**Th**	**Obs**
19	-	-	302	-	302	302	316	-
23	-	-	302 316	302 316	302	302	316	-
24	302 316	- -	302 316	- 316	302	302	316	-

For pathway A to be operative and encompass the experimental data, transport of L-kynurenine must be unfavorable under our fermentation conditions or aryl oxidation may occur earlier in the biosynthesis of HA, slightly diverging from primary metabolism. Additionally, for pathway B to be operative and agree with our data, MHA transport must be unfavorable under our fermentation conditions, as our data suggests that MHA is not a discrete precursor of anthramycin. As the intermediacy of 3-hydroxy-L-kynurenine is established based on its ability to complement anthramycin biosynthesis, we favor pathway A. Chemical complementation of 3-hydroxy-L-kynurenine and HA with the *orf23* and *orf24* mutants produced anthramycin. However, chemical complementation of the *orf19* mutant with these proposed intermediates led to the accumulation of a major shunt metabolite, desmethylanthramycin. (Table IV-2) This data indicates Orf19 is the C-methyltransferase responsible for insertion of the *C*-methyl at the 4 position of MHA, but the aromatic substrate of this enzyme remains elusive. These experiments reveal two possible substrates for incorporation into anthramycin: HA and MHA.

Orf21 amino acid activation

As discussed above, an NRPS was identified with in the anthramycin gene cluster. This dimodular NRPS consists of two genes containing six catalytic domains in co-linear order on two separate modules orf21 (A-T) and orf22 (C-A-T-Re). Based on the linear order of the domains within these modules, we hypothesized that the first A-domain residing within Orf21 was responsible for activating either MHA or HA for incorporation into anthramycin. Sequence analysis of the putative A-domain peptide sequence indicates that it is highly divergent from previously studied A-domains.[19] (Table IV-3)

Table IV-3. Specificity Code of the A-domain of Orf21

		Residue Position Based on GrsA Numbering									
Protein	Activated AA	235	236	239	278	299	301	322	330	331	517
Orf21	HA/MHA	A	A	T	N	I	S	A	A	L	K
AcmsA	MHA	N	M	M	Y	V	G	V	L	I	K
GrsA	Phe	D	A	W	T	I	A	A	I	C	K
AcvA	Aad	E	P	R	N	I	V	E	F	V	K
DhbE	Dhb/Sal	P	L	P	A	Q	G	V	V	N	K
VibE	Dhb/Sal	P	L	P	A	Q	G	V	V	N	K
EntE	Dhb/Sal	A	M	P	A	Q	G	V	V	N	K
MbtA	Dhb/Sal	P	L	P	A	Q	G	V	L	N	K
PchD	Dhb/Sal	T	L	P	A	Q	G	V	I	C	K
YbtE	Dhb/Sal	P	L	P	A	Q	G	V	L	C	K
FxbB	β-ala	D	I	N	Y	W	G	G	I	G	K
BlmIV	β-ala	V	D	W	V	I	S	L	A	D	K
CmdC	β-ala	D	G	V	Q	M	A	G	V	G	K
MdpC1	β-tyr	D	P	C	Q	V	M	V	I	A	K
SgcC1	β-tyr	D	P	A	Q	L	M	L	I	A	K

Analysis of the residues in the substrate binding region indicate Asp-235 (PheA numbering), essential for binding α-amino functionality, is substituted by alanine in Orf21. Furthermore, the 8- to 10- amino acid selectivity conferring code bears no similarity to previously described A-domains including, notably,

actinomycin synthetase ACMS I, which has been reported to activate the MHA analog *p*-toluic acid in the MHA-containing peptide actinomycin.

To provide direct biochemical evidence for substrate activation of the A-domain of ORF21, the encoding gene was cloned via polymerase chain reaction and ligated into pETDEST-42 for overproduction as a C-terminal His_6-tagged protein and purified using Ni^{+2}-affinity chromatography. Purified ORF21 activates only MHA and HA, as determined by the mass-based pyrophosphate exchange assay, stimulating a 3-fold higher rate of exchange for MHA when the reaction mixture is incubated for 30 min at 47°C, the optimal temperature for anthramycin production.[19] For further characterization of Orf21, apparent kinetic parameters were calculated for both MHA and HA. ESI-LC MS analysis yielded apparent K_M of 33 ± 3 mM and apparent k_{cat} of 130 ± 7 min^{-1} for MHA and K_M of 154 ± 9 mM and apparent k_{cat} of 99 ± 2 min^{-1} for HA. (Figure IV-5)

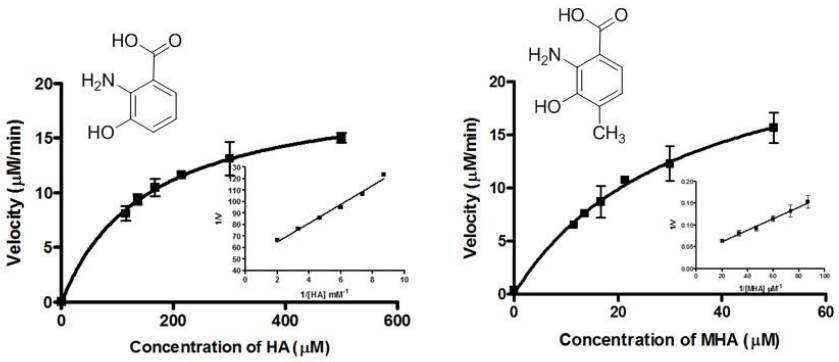


Figure IV-545. Amino Acid Concentration Dependence of Orf21. In order to establish which amino acid was the substrate of the A-domain of Orf21, substrate dependent pyrophosphate exchange activity was monitored. MHA stimulated a 3-fold higher exchange rate than HA.

Additionally, as an additional layer of substrate discrimination is involved in peptide bond formation through the binding pocket of the C-domain of an NRPS, MHA loading onto holo-Orf21 was performed.[21-24] Using a modified version of the T-loading assay, apo-Orf21 was phosphopantetheinylated by Sfp, a promiscuous phosphopantetheinyl transferase, and incubated with excess MHA. MHA was successfully loaded onto the T-domain of Orf21. This indicates that MHA is most likely a discrete precursor to anthramycin. (Figure IV-6)

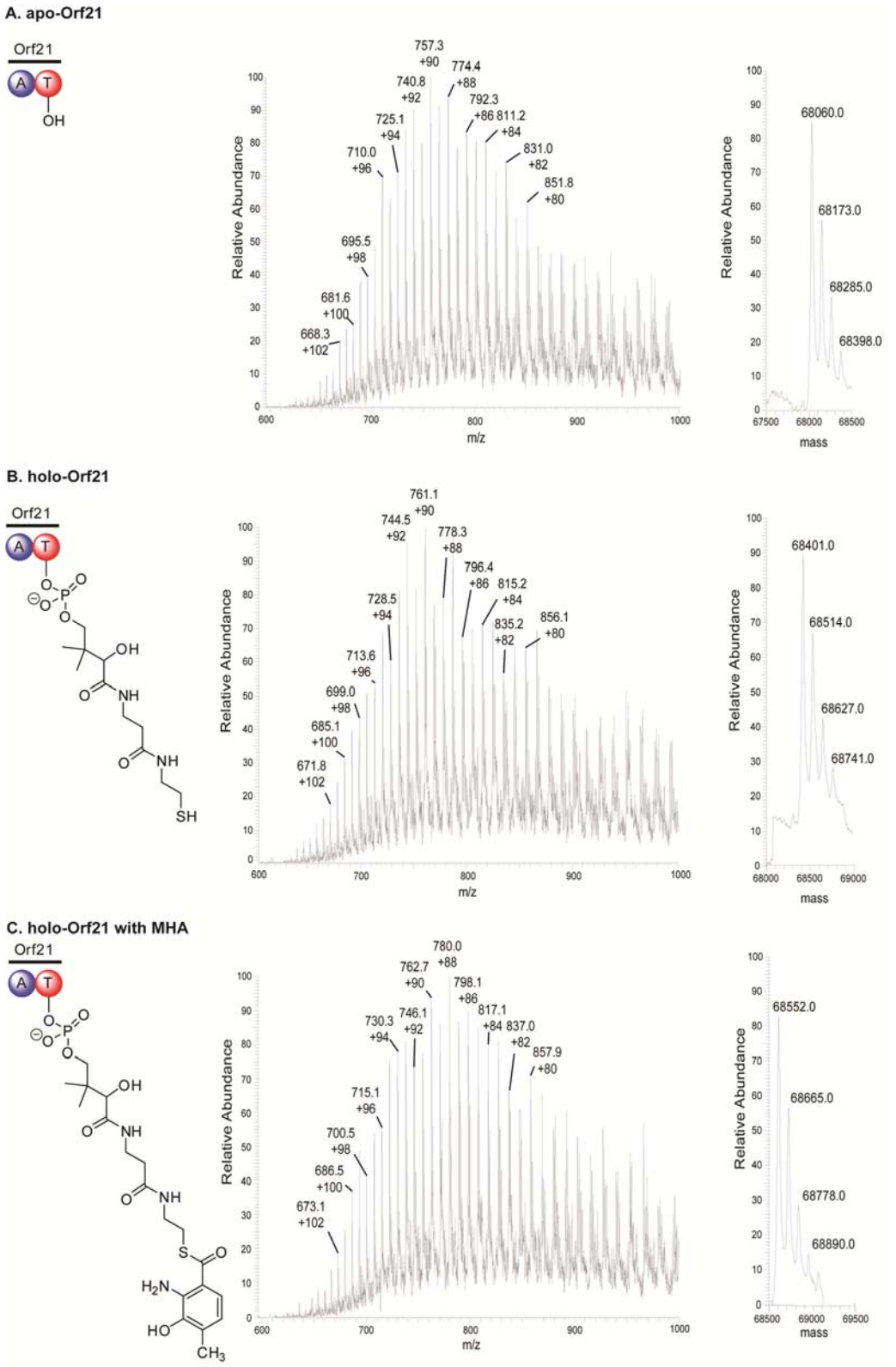


Figure IV-6. Orf21 T-loading Assay with MHA. To verify that MHA is the immediate precursor to anthramycin, the T-loading assay was performed. As expected, changes in m/z correlate to the addition of the phosphopantetheinyl group (B) and MHA (C).

While the apparent kinetic parameters indicate that MHA is a better substrate for the A-domain of Orf21, decoupled A-domains are not formally catalytic. Therefore, concentration-response curves of pyrophosphate exchange assays cannot be strictly interpreted in terms of Michaelis-Menten steady-state kinetics. In order to fully biochemically characterize Orf21 with MHA, we comprehensively derived kinetic constants using exchange kinetics based on previously described methods. [25-31]

Like tRNA synthetases, ATP-PPi exchange catalyzed by A-domains proceeds through an enzyme-bound aminoacyl adenylate intermediate and the liberated pyrophosphate product and a modified enzyme form are generated in equimolar amounts.[32] The binding of amino acid and ATP may proceed by an ordered mechanism (mechanism I and II) or by a random mechanism (mechanism III) as shown in Figure IV-7.[30] Extensive derivations and interpretations of these mechanisms have been done by Cleland, Cole and Schimmel.[27, 28, 30] Briefly, in mechanism I, a product of the enzymatic reaction is released before all reactants bind producing two or more stable enzyme forms. However, in mechanisms II and III, all reactants bind, the enzymatic reaction occurs and all products are released producing only one stable enzyme form. In order to distinguished between these mechanisms, initial velocity is measured as a function of the concentration of the variable substrate, while the secondary substrate remains constant.[28] It is possible to distinguish between possible mechanisms when several experiments at different fixed concentrations for the secondary substrate are plotted on the same graph.[30] For mechanism I, a

double-reciprocal plot will yield parallel lines, while mechanisms II and III will yield intersecting lines. For mechanism II, the intersection will occur left of the vertical axis when amino acid concentration is varied and on the vertical axis when ATP concentration is varied. For mechanism III, the lines will intersect left of the vertical axis when both amino acid and ATP concentration are varied

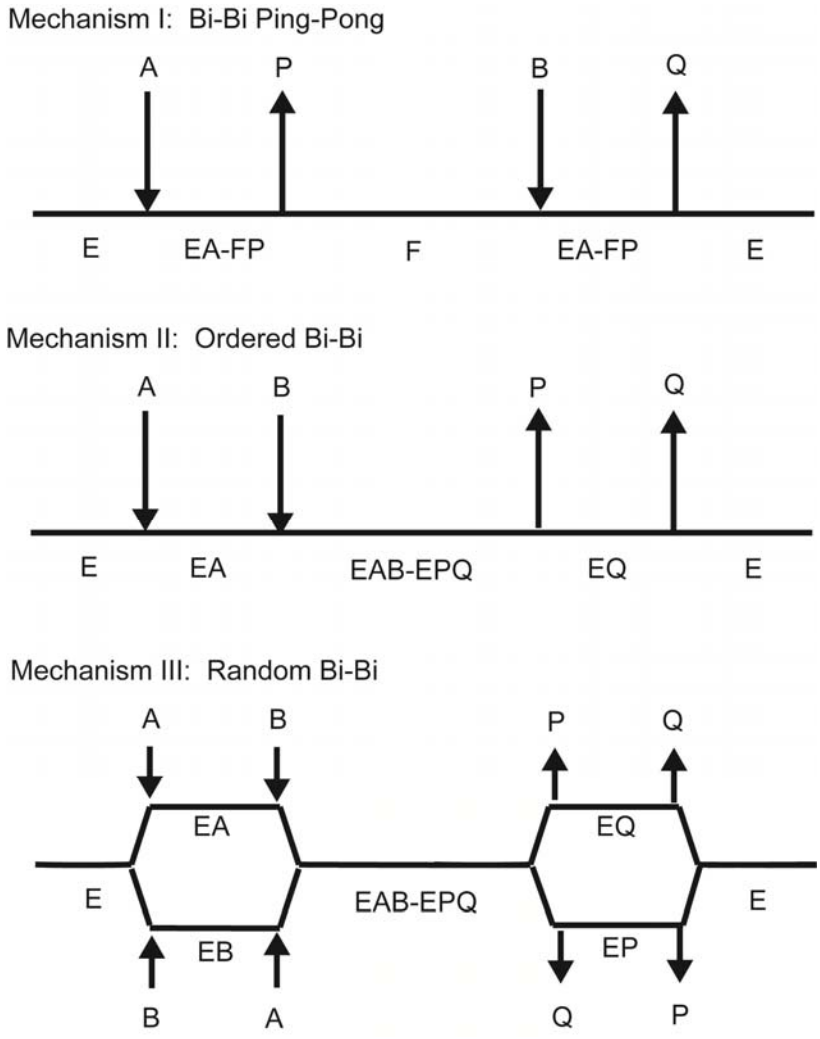


Figure IV-7. Possible Biochemical Mechanisms for A-domains. A bimolecular enzymatic reaction has three possible mechanisms: Bi-Bi Ping-Pong, where substrate A binds, releasing product P to create a bound intermediate F, followed by addition of substrate B and subsequent release of product Q to give the free enzyme; Ordered Bi-Bi, where substrate A binds, followed by substrate B subsequently, the products P and Q are released in a defined order; and Random Bi-Bi where substrates A and B bind in no defined order and products P and Q are released at random.

The mass based pyrophosphate exchange assay was used following the guidelines suggested by Cleland.[30] For these experiments, the concentration dependence of the exchange rate was investigated by varying the concentrations of ATP, pyrophosphate and MHA. The range of concentrations investigated were 0.25 mM - 4 mM ATP, 0.25 mM-4 mM PPi and 10-100 μM MHA. ATP-PPi exchange rates were calculated by standard equations derived by Cole and Schimmel[28]:

$$\frac{[E_o]}{V} = \frac{1}{k_{app}}\left(\frac{1}{[PP]} + \frac{\varphi_1}{[AA]} + \frac{\varphi_2}{[ATP]} + \frac{\varphi_3}{[AA][ATP]} + \frac{\varphi_4[PP]}{[AA][ATP]} + \varphi_5\right)$$

Where:

$$\varphi_1 = \frac{K_{eq}}{K_{ATP}};\ \varphi_2 = \frac{K_{eq}}{K_{AA}};\ \varphi_3 = K_{eq};\ \varphi_4 = \frac{K_{eq}}{K_{PP}};\ \varphi_5 = \frac{1}{K_x}$$

Data was analyzed by plotting V^{-1} versus $[MHA]^{-1}$ at different ATP and pyrophosphate concentrations and V^{-1} versus $[ATP]^{-1}$ at different MHA and pyrophosphate concentrations. (Figure IV-8)

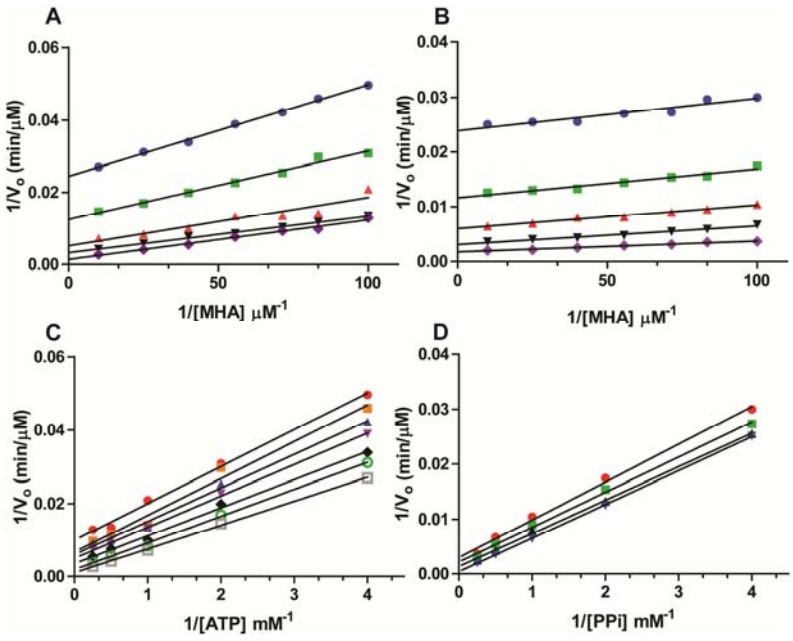


Figure IV-8. Equilibrium Kinetics of Orf21 with MHA. In order to extract equilibrium kinetics, pyrophosphate exchange reactions were performed varying ATP and MHA concentrations and PPi and MHA concentration. Plots of this data yield mechanistic information, as described in the text. A: 1/V versus 1/[MHA] for five [ATP] (0.25, 0.50, 1.0, 2.0 and 4.0 mM from top to bottom). B: 1/V versus 1/[MHA] for five [PP] (0.25, 0.50, 1.0, 2.0 and 4.0 mM from top to bottom). C: 1/V versus 1/[ATP] for seven [MHA] (10, 25, 40, 56, 71, 83, 100 μM from bottom to top). C: 1/V versus 1/[PP] for seven [MHA] (10, 25, 40, 56, 71, 83, 100 μM from bottom to top – only 10, 40, 71 and 100 μM are shown for clarity).

Extrapolations of the primary plots varying both MHA and ATP intersected left of the Y-axis. Based on the work of Cleland and coworkers, this intersection pattern is consistent with a random bi-bi mechanism (mechanism III) and is distinguished for mechanism I, a non-rapid equilibrium mechanism by lack of substrate inhibition by high levels MHA.[30] (Figure IV-7)

To calculate the equilibrium constants, secondary plots were constructed plotting the slopes and y-intercepts of the lines versus the changing variable.[27] For instance, in A of Figure IV-8, five different ATP concentrations were used over a range of MHA concentrations to yield five lines. The y-intercepts and

slopes of these lines were then plotted against 1/[ATP]. (Figure IV-9) The same holds true for the pyrophosphate concentration plots.

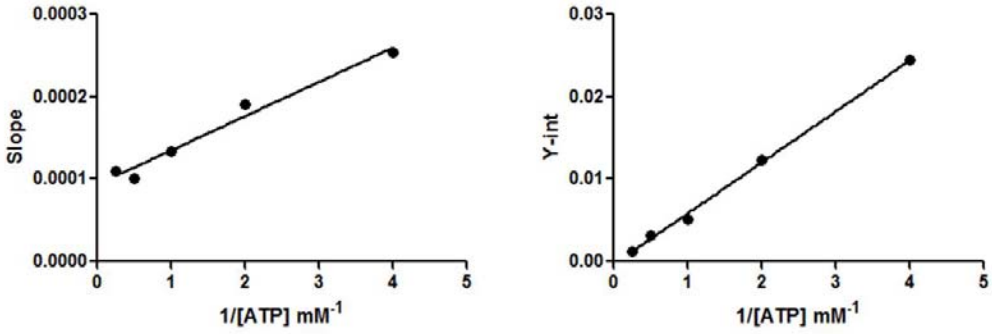


Figure IV-9. Secondary Plots of Orf21 Equilibrium Kinetics. The y-intercept of the slope replot versus 1/[ATP] (Left) gives the K_{MHA} constant, while the x-intercept of the y-intercept replot versus 1/[ATP] (Right) give the K_{ATP} constant.

The y-intercept of the slope plot versus 1/B, where B is [ATP] or [PP] and the slope is equal to (K_a/V) where K_a is equal to $K_{AA,}$ while the x-intercept of the y-intercept replot is equal to -1/$K_{b.}$ This yields K_{AA}, K_{ATP} *and* K_{PP}. To calculate K_{eq} and k_{app}, double reciprocal plots of 1/V and 1/[AA] are created to yield apparent kinetic parameters. These values can then be applied to the equation derived by Cole and Schimmel to give K_x, where x is all enzyme forms.[28] (Table IV-4)

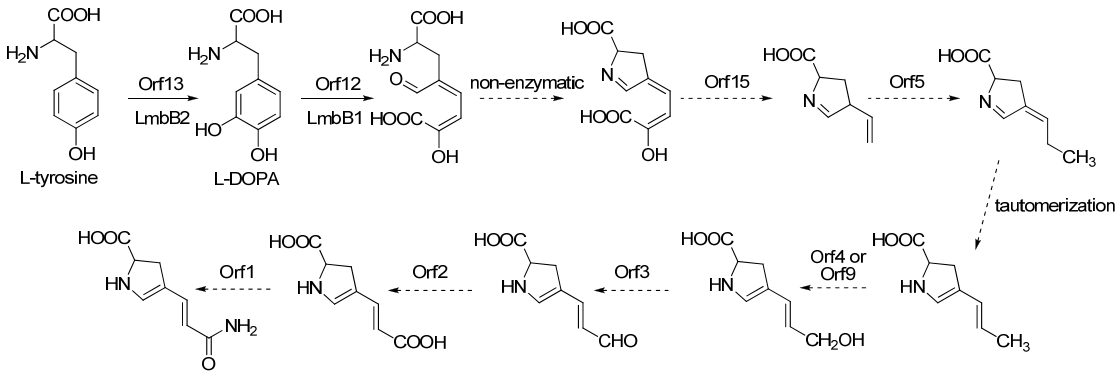
Table IV-4. Equilibrium and Kinetic Parameters of Orf21

K_{MHA} (M)	3.35×10^{-7}
K_{ATP} (M)	1.99×10^{-5}
K_{PP} (M)	5.00×10^{-5}
K_{eq} (M)	1.86×10^{-5}
K_X (M)	1.9×10^{-5}
k_{app} ($M^{-1}min^{-1}$)	6.7×10^{6}

Proposed dehydroproline acrylamide biosynthesis

The biosynthetic pathway for the dehydroproline acrylamide hemisphere of anthramycin is evidenced by high sequence similarity of a group of genes found in the anthramycin gene cluster to those involved in 4-propyl-4,5-dehydroproline biosynthesis in lincomycin.[33] Precursor experiments identified tyrosine as the precursor to the dehydropyrrole portion of both natural products while methionine dependent methylation is required for formation of the acrylamide carbonyl and propyl methyl group of anthramycin and lincomycin, respectfully.[34-37] While the precise order of biotransformations leading to the dehydroproline hemisphere is mostly speculation, the high similarity of *orf12* and o*rf13*, of the anthramycin gene cluster, to *lmbB1* and *lmbB2,* of the lincomycin gene cluster, suggest their involvement in biosynthesis of this moiety. (Table IV-1) *LmbB1* and *lmbB2* have been demonstrated to convert L-tyrosine to L-DOPA and catalyze 2,3-extradiol cleavage via an Fe^{2+}-dependent mechanism.[38-40] (Scheme IV-3) Based on labeled precursor experiments with lincomycin, after oxidative ring opening, glycoxylate (or a chemical equivalent) is lost.[34] We propose Orf5, a putative methyl transferase with 79% sequence identity to lincomycin methyltransferase LmbA, methylates the vinylogous dehydroproline intermediate providing a possible branch point for anthramycin and lincomycin biosynthesis. Tautomerization of the imine provides a possible substrate for hydroxylation by *Orf4* which shows high similarity to cytochrome P-450 hydroxylases. Subsequent transformations by *Orf3,* a putative alcohol dehydrogenase; *orf2,* a putative aldehyde dehydrogenase and *orf1,* a putative

amidotransferase provide a possible route for biosynthesis of the dehydroproline acrylamide hemisphere of anthramycin.



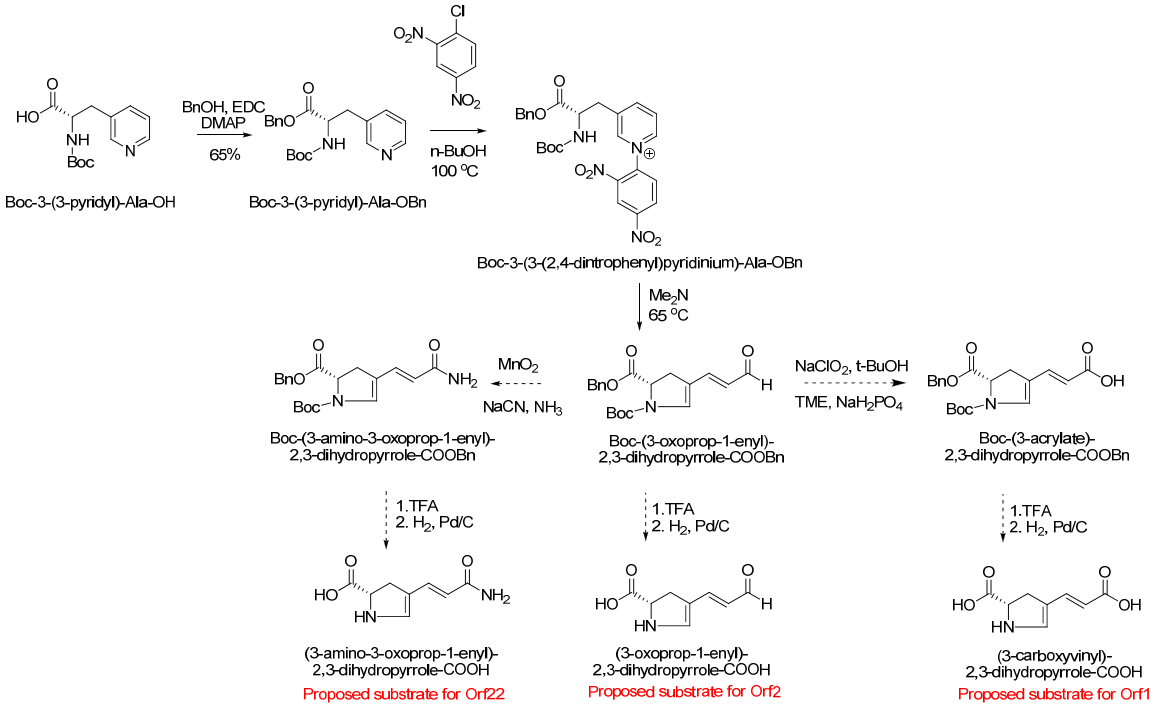
Scheme IV-3. Proposed Biosynthetic Pathway for Dehydroproline Acrylamide Hemisphere. Based on the precursor incorporation studies of Hurley *et al* and the similarity of genes within the anthramycin biosynthetic cassette to lincomycin biosynthesis, oxidative ring opening of L-DOPA followed by tailoring steps will yield the dehydroproline acrylamide moiety found within anthramycin.

Synthesis of dehydroproline acrylamide Precursor

While the biosynthetic precursors of the MHA hemisphere of anthramycin were readily available from commercial sources or through simple chemical reactions, the biosynthetic precursors of the dehydroproline acrylamide hemisphere have to be synthesized. Recently, Vanderwal and coworkers developed a new synthetic route for the synthesis of porothramycins A and B, close PBD relatives of anthramycin, through a Zincke pyridinium ring-opening/ring-closing cascade.[41, 42] Modification of this synthetic route could possibly provide us with three of the proposed intermediates in the biosynthesis of the dehydroproline acrylamide hemisphere of anthramycin.

In order to synthesize the proposed dehydroproline acrylamide substrate of Orf22, our synthetic route began with Boc-3-(3-pyridyl)-Ala-OH. (Scheme IV-4)

After benzyl protection of the carboxylic acid to give Boc-3-(3-pyridyl)-Ala-OBn, reflux with 1-chloro-2,4-dinitrobenzene in 1-butanol gave the dinitrobenzyl pyridium compound. Reaction of the dinitrocompound with dimethylamine yields the protected aldehyde. Deprotection of the aldehyde would provide the substrate of orf2. From the protected aldehyde, oxidation to the carboxylic acid and subsequent deprotection provides the substrate for orf1. Reaction of the protected aldehyde with manganese oxide in the presence sodium cyanide and ammonia provides the protected amide. Subsequent deprotection provides the free amide, yielding the proposed substrate for Orf22 activation.



Scheme IV-4. Proposed Synthesis of Dehydroproline Acrylamide and Intermediates.

Discussion

Unlike the biosynthetic pathway of K-26, the gene cluster for anthramycin was readily identified and verified. Our collaborators at Ecopia Biosciences (now Thallion Pharmaceuticals) were able to isolate the cosmids containing the entire biosynthetic cassette for anthramycin biosynthesis by using a two-tiered genetic approach. Genetic experiments by Yunfeng Hu in our lab generated a series of mutant strains unable to produce anthramycin establishing the role of this gene cluster in the biosynthesis. While the role of a variety of gene products within this cluster can be inferred through comparative genetics to similar biosynthetic pathways, the biochemical roles of these proteins remained completely uninvestigated. Based on the chemical and genetic similarity of the MHA hemisphere of anthramycin to primary metabolic production of HA and the dehydroproline acrylamide to the dehydroproline moiety in lincomycin, we were able to propose a possible biosynthetic route for both halves.

For MHA biosynthesis, chemical complementation studies provided us with two possible substrates for the NRPS: HA and MHA. Using the mass-based pyrophosphate exchange assay, we were able to show that MHA is most likely the immediate precursor to anthramycin through both activation by the A-domain of Orf21 and the subsequent loading of the T-domain on the same module. Additionally, we determined that the mechanism of an isolated A-domain under reversible conditions is random bi-bi. While we were completing this work, Aldrich and coworkers established using a terminal assay that adenylate forming

enzymes proceed through a bi-uni-uni-bi ping-pong kinetic mechanism when an activated T-domain is present.[43] (Figure IV-10) This corresponds well with the mechanism of tRNA synthetases with (bi-uni-uni-bi) and without (bi-bi) tRNA.[25]

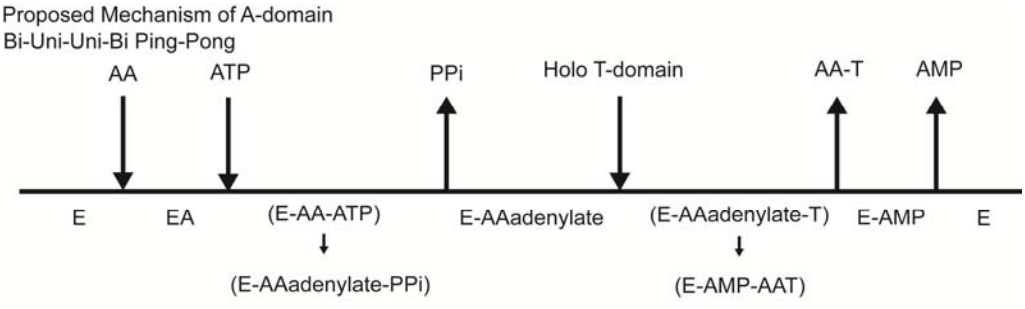


Figure IV-10. Proposed Mechanism of A-domain. Recent literature suggests that when a T-domain is present, the mechanism of an A-domain is a Bi-Uni-Uni-Bi Ping-Pong mechanism where the amino acid binds first, followed by ATP. The aminoacyl adenylate is formed releasing pyrophosphate. Subsequently, the T-domain binds and is loaded with the amino acid. The primed T-domain is then released. Lastly, AMP is released from the A-domain.

With the substrate of the MHA hemisphere of anthramycin in hand, we turned our investigations to the identification of the substrate of Orf22. Unlike the possible intermediates of the MHA biosynthetic pathway, the substrates necessary to investigate the biosynthesis of the dehydroproline acrylamide hemisphere were not readily available. A short synthetic route provides us with the amide. Currently, *orf22* is being synthesized by GenScript as cloning of the approximately 4kb was unsuccessful. Once gene synthesis is complete, Orf22 will be tested for activity with the dehydroproline acrylamide substrate with both pyrophosphate exchange and the T-loading assay. Additionally, investigations of the biosynthesis of the amide can be investigated using gene knockouts and chemical complementation. Through slight adjustments of the synthetic pathway

for the amide, proposed substrates for *orf2* (aldehyde) and *orf1* (carboxylic acid), can be synthesized.

In addition to assay based biochemical investigations, the small dimodular nature of the NRPS within the anthramycin gene cluster provides the perfect system for protein crystallography of an entire NRPS. The large size of the total NRPS systems limits the ability to crystallize the protein complex successfully.[44-50] Co-expression of Orf21 and Orf22 and *in vitro* production of anthramycin will provide the groundwork for crystallography work. Future crystallization of the NRPS of anthramycin may provide mechanistic insight into the protein-protein interactions within these complexes. Additionally, as more and more genome sequences for PBD producers are becoming accessible, it may be possible to genetically modify the A-domains to bind non-native substrates to produce novel molecules with superior or novel biological activities.[51]

After identification and functional determination of the anthramycin gene cluster, the bacterial gene clusters of sibiromycin, tomaymycin and diazepinomicin/ECO-4601, a novel farnesylated dibenzodiazepinone, were identified.[52-54] Comparative genomics of the sibiromycin, tomaymycin, anthramycin and lincomycin producers provided additional evidence, through high similarity and homology of analogous proposed proteins within the cluster, that our gene assignments were correct. Additionally, the 8- to 10- amino acid code conferring A-domain specificity of the A-domains activation HA derivatives in sibiromycin, tomaymycin, anthramycin and a series of fungal gene clusters were extracted. This information revealed that bacterial substituted anthranilic

acid activating A-domains cluster separately from fungal anthranilic acid activating A-domains.[55] Additionally, all of these β-amino acid have a hydrophobic residue, alanine for bacterial A-domains and glycine for fungal A-domains, at position 235 (GrsA numbering), replacing aspartate. (Table IV-4).

Table IV-4. Specificity Code of Orf21, SibE and TomA

		Residue Position Based on GrsA Numbering									
Protein	Activated AA	235	236	239	278	299	301	322	330	331	517
Orf21	HA/MHA	A	A	T	N	I	S	A	A	L	K
SibE	MHA	A	A	T	N	I	S	A	A	L	K
TomA	DHA	A	A	I	S	L	S	G	S	I	K
AcmsA	MHA	N	M	M	Y	V	G	V	L	I	K
GrsA	Phe	D	A	W	T	I	A	A	I	C	K
DhbE	Dhb/Sal	P	L	P	A	Q	G	V	V	N	K
VibE	Dhb/Sal	P	L	P	A	Q	G	V	V	N	K
MbtA	Dhb/Sal	P	L	P	A	Q	G	V	L	N	K
YbtE	Dhb/Sal	P	L	P	A	Q	G	V	L	C	K
AnaPS_A1	AA	G	A	L	F	F	A	A	G	V	K

Materials and Methods

Bacterial Strains and Culture Conditions

Streptomyces refuineus var. *thermotolerans* (NRRL 3143) and its derivatives were maintained and grown on either ISP4 medium or TSB medium with appropriate antibiotics at 37°C. For anthramycin production, *S. refuineus* and its derivatives were cultured in production medium (1% corn starch, 2% peptonized milk, and 0.3% yeast extract at pH 7.0) at 47 °C. *Bacillus* sp. TA (NRRL B-3167) was cultured in nutrient agar and used as a test organism for the antibacterial activity of *S. refuineus* and its derivatives. DH10B (Invitrogen) served as host for *S. refuineus* genomic library construction. *E. coli* BW25113 containing plasmid pIJ790 was used for targeted gene disruption in *S. refuineus*. *E. coli* ET12567

containing the RP4 derivative pUZ8002 was used for intergenetic conjugation between *E. coli* and *S. refuineus*. *E. coli* strains were grown in LB medium supplemented with appropriate antibiotics for selection of plasmids.

Preparation of Biosynthetic Intermediates of Anthramycin

L-kynurenine, 3-hydroxy-L-kynurenine, hydroxyanthranilic acid were purchased directly from Sigma Corporation. MHA was prepared by catalytic hydrogenation of 4-methyl-3-hydroxy-2-nitrobenzoic acid (Aldrich, Inc.) with Pd/C and H_2 at 1 atm in ethanol.

Chemical Complementation of Anthramycin

For chemical complementation, corresponding compounds were added directly to the production medium at the beginning of fermentation for a final concentration of 2 mM.

Production and Detection of Anthramycin

S. refuineus and its mutant derivatives were cultured in 50 ml seed medium at 47°C for 24 hr. A 5% inoculum was then added to 50 ml production medium and cultured at 47°C for 24 hr. Anthramycin was extracted from the production medium with 50 mL butanol. Butanol extracts were concentrated *in vacuo* and redissolved in MeOH. Antibacterial activity of anthramycin was detected by thin-layer chromatography bioautography. Anthramycin butanolic extracts (dissolved in MeOH) were run on 25DC-Alufolien Kieselgel plates with solvent MeOH:$CHCl_3$ (1:9); LB agar containing indicator strain Bacillus sp. TA was overlaid on TLC

plates and cultured at 37 °C for 20 hr to detect anti-*Bacillus* activity (evidenced by growth inhibition zones) of anthramycin.

Detection of Anthramycin by LC/MS

Anthramycin production was further confirmed by HPLC/MS. Mass spectrometry was performed by using ThermoFinnigan (San Jose, CA) TSQ Quantum triple quadrupole mass spectrometer equipped with a standard electrospray ionization source outfitted with a 100 mm I.D. deactivated fused Si capillary. The injection volume was 10 ml. Anthramycin was separated from co-metabolites by using a Jupiter minibore 5 μm C18 column (2.0 mm x 15 cm) with a linear gradient (0-100% Buffer B; Buffer A 95% water 5% acetonitrile 10mM ammonium acetate; Buffer B 5% water 95% acetonitrile 10 mM ammonium acetate). The flow rate was 0.2 ml/min. The mass spectrometer was operated in the positive (or negative) ion mode, and the electrospray needle was maintained at 4,200 V. The ion transfer tube was operated at 35 V and 342 °C (~35 V and 300 °C for negative). The tube lens voltage was set to 85 V (~220 V for negative). Source CID (offset voltage between skimmer and the first ion guide, Q00) was used at 15 V. The mass spectrometer was operated in full scan mode with Quad 1. The mass spectral resolution was set to a peak width of 0.70 u (full width at half maximum, FWHM). Full scan spectra were acquired from m/z 150.0 to 700.0 (m/z 150.0 to 1200.0 for negative) over 1.0 s. Data were acquired in profile mode. The electron multiplier gain was set to 3×10^5.

Cloning and Expression of ORF21

ORF21 was cloned from the *Streptomyces refuineus* genomic DNA via PCR using the primers 5'-CACCATGACAGTACGCAGCACCGCC-3' and 5'-GCCTCGGGAACGCTTGGTG-3' (synthesized by Sigma-Aldrich). The 1.8 kb PCR product was cloned into the vector pENTR^{TM}/SD/D-TOPO® using the pENTR™ directional TOPO cloning kit (Invitrogen). The gene was subcloned into an expression vector pET-DEST42 via LR-recombination using Gateway LR Clonase enzyme mix (Invitrogen). The resulting plasmid was introduced into the ROSETTA *E. coli* strain (Novagen). The culture was inoculated (1:100) with an overnight culture made from a fresh colony of the above strain and grown at 37 °C in a 2.8 L baffled flask containing 500 mL LB medium with 100 μg/mL OD_{600} = ~ 0.6. IPTG was then added (final concentration of 0.5 mM) and the culture grown at °C. The cells were pelleted (30 min, 3750 rpm, 4°C) and resuspended in binding buffer (500 mM NaCl, 20 mM NaH_2PO_4, 20 mM imidazole, pH 7.4). For cell lysis, DNase I (NEB, 0.2 U/mL) was added, the cells were disrupted using a French pressure cell and filtered through a 0.45-μm filter. The 63 kD protein was purified on a HisTrap FF column (GE Healthcare) on an ÄKTA chromatography system (GE Healthcare) using binding buffer with linearly increasing imidazole concentration (20–500 mM). The pure protein was then desalted with a HiTrap Desalting column using 20 mM Tris, pH 7.5 and stored in aliquots at –80°C in storage buffer containing 5% glycerol and 1 mM DTT.

ATP-PPi Exchange Assay Conditions

In order to avoid precipitation of magnesium pyrophosphate, assay components were divided into stock solutions comprising 1) 3 mM amino acids containing 15 mM PPi in 20 mM Tris pH 7.5, 2) 3 mM γ–$^{18}O_4$-ATP containing 15mM $MgCl_2$ in 20 mM Tris pH 7.5 and 3) 600 nM enzyme in 20 mM Tris pH 7.5 containing 5% glycerol and 1mM DTT. Exchange reactions containing 2 μL of each component were initiated by the addition of enzyme solution. 6 μL reactions therefore contained final concentrations of 5 mM $MgCl_2$, 5 mM PP_i, 1 mM γ-$^{18}O_4$-ATP, 1mM amino acid and 20 mM Tris-HCl pH 7.5. After an incubation period (30 min at 47 ^{o}C), the reactions were stopped by the addition of 6 μL acetone for ESI-LC/MS analysis.

Orf21 Michaelis-Menten Kinetic Parameters

For Orf21 kinetic measurements, the reactions were performed as described with the following modifications: % exchange was measured over a range of MHA concentrations (11.5 – 50.0 μM) and HA concentrations (0.115-0.5 mM) and rates were measured as single time points at 10 minutes and 15 minutes, respectively, corresponding to a maximal % exchange of 15%. The 'apparent' substrate dependence parameters, K_M and k_{cat}, were calculated from initial velocity time curves with varied MHA or HA concentrations using a non-linear least square fit to the Michaelis-Menten equation.

Orf21 Equilibrium Kinetic Parameters

For Orf21 kinetic measurements, the reactions were performed as described with the following modifications: % exchange was measured over a range of MHA concentrations (10 – 100 μM), ATP (0.25 – 4mM) and pyrophosphate (0.25-4mM) and rates were measured as single time points at 10 minutes. The equilibrium constants were derived from double reciprocal plots and secondary plots of slopes and intercepts versus substrate concentration. Slope and intercept were determined by applying linear regressions using GraphPad Prism.

Mass Spectrometry

ESI-LC/MS analyses were performed on a ThermoFinnigan LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, Waltham, MA) equipped with an ESI interface in negative ion mode. Nitrogen was used both for the auxiliary and sheath gas. The auxiliary and sheath gases were set to 20 psi and 36 psi, respectively. The following instrumental parameters were used: capillary temperature 300 °C; source voltage 4.5 kV; source current 100 μA capillary voltage -49.0 V; tube lens -148.30 V; skimmer offset 0.00 V; activation time 50 ms with an isolation width of 1 *m*/*z*. The sensitivity of the mass spectrometer was tuned by infusion of γ-$^{16}O_4$-ATP at a flowrate of 0.01 mL/min. Samples were introduced by a Waters Acquity UPLC system (Waters, Milford, MA) with an injection volume of 5 μL. γ-$^{18}O_4$-ATP was separated from contaminating salts on a 5 μm Hypercarb column (3 x 50 mm, ThermoFisher Scientific) with a isocratic method of 82.5% 20 mM ammonium acetate pH 6 containing 0.1% diethylamine

and 17.5% acetonitrile over 5 minutes with a flow rate of 0.2 mL/min. A divert valve was used for the first two minutes to avoid introducing salts into the source and the retention time of ATP species under these conditions was 3 minutes. Data acquisition and quantitative spectral analysis was conducted using the Thermo-Finnigan Xcaliber software, version 2.0 Sur 1.

Data Analysis

Monoisotopic peak area was determined using the respective software for each method. For MALDI-TOFMS analysis, the ratio of the area of γ–$^{16}O_4$-ATP (m/z 506) to the area of total ATP including unlabeled, partially labeled, fully labeled and monosodium-coordinated ions (m/z 506, 508, 510, 512, 514, 528, 530, 532, 534, 536) was calculated using Microsoft excel and adjusted to reflect actual incorporation based on maximum theoretical incorporation to yield percent turnover. For ESI-LC/MS analysis, the ratio of the area of γ–$^{16}O_4$-ATP to the area of total ATP including unlabeled, partially labeled, fully labeled, monosodium-coordinated and sodium acetate adduct ions was calculated using Microsoft excel and adjusted to reflect actual incorporation based on maximum theoretical incorporation to yield percent turnover.

Sfp cloning and expression

Sfp was cloned as previously described. The resulting plasmid was introduced into the ROSETTA *E. coli* strain (Novagen). The culture was inoculated (1:100) with an overnight culture made from a fresh colony of the above strain and grown at 37°C in a 2.8 L baffled flask containing 500 mL LB

medium with 50 μg/mL kanamycin until OD_{600} = ~ 0.6 was reached. IPTG was then added (final concentration of 1 mM) and the culture grown overnight at 15°C. The cells were pelleted (30 min, 3750 rpm, 4°C) and resuspended in binding buffer (500 mM NaCl, 20 mM NaH_2PO_4, 20 mM imidazole, pH 7.4). For cell lysis, DNase I (NEB, 0.2 U/mL) was added, the cells were disrupted using a French pressure cell and filtered through a 0.45-μm filter. The 26 kD His-tagged protein protein was purified on a HisTrap FF column (GE Healthcare) on an ÄKTA chromatography system (GE Healthcare) using binding buffer with linearly increasing imidazole concentration (20–500 mM). The pure protein was then desalted with a HiTrap Desalting column using 20 mM Tris, pH 7.5, flash frozen in liquid nitrogen and stored in aliquots at –80°C in storage buffer containing 10% glycerol and 1 mM EDTA.

T-loading assay

Holo-Orf21 was generated *in situ* by incubating 25 μM apo-Orf21 with 4 μM Sfp in the presence of 10 mM $MgCl_2$ and 250 μM Coenzyme-A. After 1hr, the reaction mix was supplemented with 5 mM ATP and 1 mM MHA. The reaction was allowed to incubate for 30 minutes and extraneous salts were removed by filtration by centrifugation (10 kD MW cutoff).

Mass Spectrometry for T-domain Loading

ESI-LC/MS analyses were performed on a ThermoFinnigan LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific, Waltham, MA) equipped with an

ESI interface in positive ion mode. Nitrogen was used both for the auxiliary and sheath gas. The auxiliary and sheath gases were set to 20 psi and 36 psi, respectively. The following instrumental parameters were used: capillary temperature 300 °C; source voltage 4.5 kV; source current 100 μA capillary voltage -49.0 V; tube lens -148.30 V; skimmer offset 0.00 V; activation time 50 ms with an isolation width of 1 *m*/*z*. Samples were introduced by a Waters Acquity UPLC system (Waters, Milford, MA) with an injection volume of 10 μL. Orf21 was separated from contaminating co-purified proteins on a Waters Acquity UPLC HEB C8 column (1.0 x 150 mm, 1.7 μm), with a gradient method of 30% B to 100% B over 10 minutes (Buffer A: 80% water, 10% acetonitrile, 10% isopropanol, 2% acetic acid, 0.2% trifluoroacetic acid; Buffer B: 20% water, 70% acetonitrile, 10% isopropanol, 2% acetic acid, 0.2% trifluoroacetic acid). Data acquisition and spectral analysis was conducted using the BioMass software.

Synthesis of dehydroproline acrylamide derivatives

Boc-3-(3-pyridyl)-Ala-OBn

To a stirred solution of Boc-3-(3-pyridyl)-Ala-OH (1 eq, 300 mg, 1.13 mmol) in anhydrous CH_2Cl_2 (10 mL) was added DMAP (0.1 eq, 13.8 mg, .113 mmol) and benzyl alcohol (4 eq, 468 μL, 4.52 mmol). The reaction was cooled to 0^{o}C and EDC (3 eq, 238 mg, 3.39 mmol) was added.. After stirring for 5 minutes, the reaction was allowed to warm to room temperature and stirred overnight. The reaction mixture was washed with HCl (0.5N, 2 x 30 mL) followed by saturated

$NaHCO_3$ (3 x 30 mL) and dried ($MgSO_4$). The solvent was removed *in vacuo* to yield the crude ether. The ether was purified by column chromotography (100% DCM followed by 4% MeOH in DCM) to give the pure compound (230 mg, 0.64 mmol, 57%) ^{1}H NMR (400 MHz, $CDCl_3$): δ8.49 (s, 1H), 8.36 (s, 1H), 7.58-7.33 (m, 4H), 7.16 (s, 1H), 5.07 (d, 2H), 4.98 (s, 1H), 4.55 (s, 1H), 3.19 (m, 1H), 3.06 (m, 1H), 1.44 (s, 9H). ^{13}C NMR (100 MHz, $CDCl_3$): δ171.10, 150.52, 148.37, 136.70, 134.86, 131.52, 128.60, 123.26, 67.33, 54.03, 35.48, 28.18.

Boc-3-(3-(2,4-dintrophenyl)pyridinium)-Ala-OBn

To a solution of Boc-3-(3-pyridyl)-Ala-OBn (1 eq, 50 mg, 0.14 mmol) in 1-butanol (0.1 mL) was added 1-chloro-2,4-dinitrobenzene (1 eq, 28.4 mg, 0.14 mmol). The resulting solution was heated to 100^{o}C and stirred for 3 hours. After cooling to room temperature, the solvent was removed *in vacuo*. The material was purified by silica gel chromatography (20% MeOH in DCM). (36.8 mg, 0.066 mmol ,47%) ^{1}H NMR (400 MHz, DMSO): δ9.400 (s, 1H), 9.322 (m, 1H), 9.129 (s, 1H), 8.989 (m, 1H), 8.840 (m, 1H), 8.380 (m, 2H), 7.368 (s, 5H), 5.168 (s, 1H), 4.540 (broad s, 1H), 3.223 (m, 2H) 1.326 (s, 9H) ^{13}C NMR (100 MHz, DMSO): δ162.70, 155.85, 149.57, 143.29, 138.95, 136.11, 132.26, 130.50, 128.79, 128.52, 128.24, 121.92, 79.10, 66.72, 53.99, 36.17, 33.42, 31.15, 28.42.

Boc-(3-oxoprop-1-enyl)-2,3-dihydropyrrole-COOBn

To a solution of Boc-3-(3-(2,4-dinitrophenyl)pyridinium)-Ala-OBn (1 eq, 10 mg, 0.018mmol) in ethanol (100 proof, 471μL) at 50^{o}C, under anhydrous condtions,

was added dimethylamine (2.0 M in MeOH, 5 eq, 45μL, 0.09 mmol) and the reaction vessel was opened to atmosphere. The solution was heated to 65°C and stirred until the solution turned from a deep maroon to red. The mixture was allowed to cool to room temperature and concentrated *in vacuo*. Water was added and the aqueous solution was extracted with DCM (3 x 10 mL). The organic extracts were combined, washed with brine (3 x 30 mL), dried ($MgSO_4$), filtered and concentrated *in vacuo*. The aldehyde was purified via silica gel chromatography (1:1 EtOAc:hexanes) followed by phenylhexyl LC/MS. (2.95 mg, 0.008 mmol, 46%).

REFERENCES

1. Tendler, M. D. and Korman, S. (1963) 'Refuin': A Non-Cytotoxic Carcinostatic Compound Proliferated by a Thermophilic Actinomycete. *Nature*, **199**, 501.

2. Hurley, L. H., Zmijewski, M. and Chang, C. J. (1975) Biosynthesis of Anthramycin - Determination of Labeling Pattern by Use of Radioactive and Stable Isotope Techniques. *Journal of the American Chemical Society*, **97**, 4372-4378.

3. Hurley, L. H. and Petrusek, R. (1979) Proposed Structure of the Anthramycin-DNA Adduct. *Nature*, **282**, 529-531.

4. Kizu, R., Draves, P. H. and Hurley, L. H. (1993) Correlation of DNA-Sequence Specificity of Anthramycin and Tomaymycin with Reaction-Kinetics and Bending of DNA. *Biochemistry*, **32**, 8712-8722.

5. Kohn, K. W., Bono, V. H. and Kann, H. E. (1968) Anthramycin a New Type of DNA-Inhibiting Antibiotic - Reaction with DNA and Effect on Nucleic Acid Synthesis in Mouse Leukemia Cells. *Biochimica Et Biophysica Acta*, **155**, 121-&.

6. Kohn, K. W., Glaubige.D and Spears, C. L. (1974) Reaction of Anthramycin with DNA .2. Studies of Kinetics and Mechanism. *Biochimica Et Biophysica Acta*, **361**, 288-302.

7. Kohn, K. W. and Spears, C. L. (1968) Reaction of Anthramycin with DNA - Kinetics and Specificity. *Federation Proceedings*, **27**, 800-&.

8. Kohn, K. W. and Spears, C. L. (1970) Reaction of Anthramycin with Deoxyribonucleic Acid. *Journal of Molecular Biology*, **51**, 551-&.

9. Kopka, M. L., Goodsell, D. S., Baikalov, I., Grzeskowiak, K., Cascio, D. and Dickerson, R. E. (1994) Crystal structure of a covalent DNA-drug adduct: anthramycin bound to C-C-A-A-C-G-T-T-G-G and a molecular explanation of specificity. *Biochemistry*, **33**, 13593-13610.

10. Krugh, T. R., Graves, D. E. and Stone, M. P. (1989) 2-Dimensional Nmr-Studies on the Anthramycin-D(Atgcat)2 Adduct. *Biochemistry*, **28**, 9988-9994.

11. Kumar, R. and Lown, J. W. (2003) Recent developments in novel pyrrolo[2,1-c][1,4]benzodiazepine conjugates: synthesis and biological evaluation. *Mini Reviews in Medicinal Chemistry*, **3**, 323-339.

12. Pepper, C. J., Hambly, R. M., Fegan, C. D., Delavault, P. and Thurston, D. E. (2004) The novel sequence-specific DNA cross-linking agent SJG-136 (NSC 694501) has potent and selective in vitro cytotoxicity in human B-cell chronic lymphocytic leukemia cells with evidence of a p53-independent mechanism of cell kill. *Cancer Research*, **64**, 6750-6755.

13. Hurley, L. H. (1980) Elucidation and Formulation of Novel Biosynthetic Pathways Leading to the Pyrrolo[1,4]Benzodiazepine Antibiotics Anthramycin, Tomaymycin, and Sibiromycin. *Accounts of Chemical Research*, **13**, 263-269.

14. Hurley, L. H. and Gairola, C. (1979) Pyrrolo (1,4) benzodiazepine antitumor antibiotics: Biosynthetic studies on the conversion of tryptophan to the anthranilic acid moieties of sibiromycin and tomaymycin. *Antimicrobial Agents and Chemotherapy*, **15**, 42-45.

15. Hurley, L. H., Lasswell, W. L., Malhotra, R. K. and Das, N. V. (1979) Pyrrolo[1,4]benzodiazepine antibiotics. Biosynthesis of the antitumor antibiotic sibiromycin by Streptosporangium sibiricum. *Biochemistry*, **18**, 4225-4229.

16. Hurley, L. H., Zmijewski, M. and Chang, C. J. (1975) Biosynthesis of anthramycin. Determination of the labeling pattern by the use of radioactive and stable isotope techniques. *Journal of the American Chemical Society*, **97**, 4372-4378.

17. Hu, Y., Phelan, V., Ntai, I., Farnet, C. M., Zazopoulos, E. and Bachmann, B. O. (2007) Benzodiazepine biosynthesis in Streptomyces refuineus. *Chemistry & Biology*, **14**, 691-701.

18. Hu, Y., Phelan, V. V., Farnet, C. M., Zazopoulos, E. and Bachmann, B. O. (2008) Reassembly of anthramycin biosynthetic gene cluster by using recombinogenic cassettes. *Chembiochem*, **9**, 1603-1608.

19. Phelan, V. V., Du, Y., McLean, J. A. and Bachmann, B. O. (2009) Adenylation enzyme characterization using gamma -(18)O(4)-ATP pyrophosphate exchange. *Chemistry & Biology*, **16**, 473-478.

20. Rokem, J. S. and Hurley, L. H. (1981) Sensitivity and Permeability of the Anthramycin Producing Organism Streptomyces-Refuineus to Anthramycin and Structurally Related Antibiotics. *Journal of Antibiotics*, **34**, 1171-1174.

21. Dorrestein, P. C. and Kelleher, N. L. (2006) Dissecting non-ribosomal and polyketide biosynthetic machineries using electrospray ionization Fourier-Transform mass spectrometry. *Natural Products Reports*, **23**, 893-918.

22. Dorrestein, P. C., Bumpus, S. B., Calderone, C. T., Garneau-Tsodikova, S., Aron, Z. D., Straight, P. D., Kolter, R., Walsh, C. T. and Kelleher, N. L. (2006) Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. *Biochemistry*, **45**, 12756-12766.

23. Hicks, L. M., Mazur, M. T., Miller, L. M., Dorrestein, P. C., Schnarr, N. A., Khosla, C. and Kelleher, N. L. (2006) Investigating nonribosomal peptide and polyketide biosynthesis by direct detection of intermediates on >70 kDa polypeptides by using Fourier-transform mass spectrometry. *Chembiochem*, **7**, 904-907.

24. Dorrestein, P. C., Blackhall, J., Straight, P. D., Fischbach, M. A., Garneau-Tsodikova, S., Edwards, D. J., McLaughlin, S., Lin, M., Gerwick, W. H., Kolter, R., Walsh, C. T. and Kelleher, N. L. (2006) Activity screening of carrier domains within nonribosomal peptide synthetases using complex substrate mixtures and large molecule mass spectrometry. *Biochemistry*, **45**, 1537-1546.

25. Freist, W., Sternbach, H. and Cramer, F. (1982) Isoleucyl-tRNA synthetase from Escherichia coli MRE 600. Different pathways of the aminoacylation reaction depending on presence of pyrophosphatase, order of substrate addition in the pyrophosphate exchange, and substrate specificity with regard to ATP analogs. *European Journal of Biochemistry*, **128**, 315-329.

26. Boyer, P. D. (1959) Uses and limitations of measurements of rates of isotopic exchange and incorporation in catalyzed reactions. *Archives of Biochemistry and Biophysics*, **82**, 387-410.

27. Cleland, W. W. (1967) Enzyme kinetics. *Annual Review of Biochemistry*, **36**, 77-112.

28. Cole, F. X. and Schimmel, P. R. (1970) On the rate law and mechanism of the adenosine triphosphate--pyrophosphate isotope exchange reaction of amino acyl transfer ribonucleic acid synthetases. *Biochemistry*, **9**, 480-489.

29. Midelfort, C. F. and Mehler, A. H. (1974) Applications of kinetic methods to aminoacyl-tRNA synthetases. *Methods in Enzymology*, **29**, 627-642.

30. Santi, D. V., Webster, R. W., Jr. and Cleland, W. W. (1974) Kinetics of aminoacyl-tRNA synthetases catalyzed ATP-PPi exchange. *Methods in Enzymology*, **29**, 620-627.

31. Yagil, G. and Hoberman, H. D. (1969) Rate of isotope exchange in enzyme-catalyzed reactions. *Biochemistry*, **8**, 352-360.

32. Finking, R. and Marahiel, M. A. (2004) Biosynthesis of nonribosomal peptides. *Annual Review of Microbiology*, **58**, 453-488.

33. Peschke, U., Schmidt, H., Zhang, H. Z. and Piepersberg, W. (1995) Molecular Characterization of the Lincomycin-Production Gene-Cluster of Streptomyces Lincolnensis-78-11. *Molecular Microbiology*, **16**, 1137-1156.

34. Brahme, N. M., Gonzalez, J. E., Rolls, J. P., Hessler, E. J., Mizsak, S. and Hurley, L. H. (1984) Biosynthesis of the Lincomycins .1. Studies Using Stable Isotopes on the Biosynthesis of the Propyl-L-Hygric and Ethyl-L-Hygric Acid Moieties of Lincomycin-a and Lincomycin-B. *Journal of the American Chemical Society*, **106**, 7873-7878.

35. Hurley, L. H., Lasswell, W. L., Ostrander, J. M. and Parry, R. (1979) Pyrrolo[1,4]benzodiazepine antibiotics. Biosynthetic conversion of tyrosine to the C2- and C3-proline moieties of anthramycin, tomaymycin, and sibiromycin. *Biochemistry*, **18**, 4230-4237.

36. Ostrander, J. M., Hurley, L. H., McInnes, A. G., Smith, D. G., Walter, J. A. and Wright, J. L. (1980) Proof for the biosynthetic conversion of L-[indole-15N]tryptophan to [10-15N]anthramycin using (13C, 15N) labelling in conjunction with 13C-NMR and mass spectral analysis. *Journal of Antibiotics (Tokyo)*, **33**, 1167-1171.

37. Witz, D. F., Hessler, E. J. and Miller, T. L. (1971) Bioconversion of Tyrosine into Propylhygric Acid Moiety of Lincomycin. *Biochemistry*, **10**, 1128-&.

38. Colabroy, K. L., Hackett, W. T., Markham, A. J., Rosenberg, J., Cohen, D. E. and Jacobson, A. (2008) Biochemical characterization of L-DOPA 2,3-dioxygenase, a single-domain type I extradiol dioxygenase from lincomycin biosynthesis. *Archives of Biochemistry and Biophysics*, **479**, 131-138.

39. Harayama, S. and Rekik, M. (1989) Bacterial Aromatic Ring-Cleavage Enzymes Are Classified into 2 Different Gene Families. *Journal of Biological Chemistry*, **264**, 15328-15333.

40. Novotna, J., Honzatko, A., Bednar, P., Kopecky, J., Janata, J. and Spizek, J. (2004) L-3,4-Dihydroxyphenyl alanine-extradiol cleavage is followed by intramolecular cyclization in lincomycin biosynthesis. *European Journal of Biochemistry*, **271**, 3678-3683.

41. Kearney, A. M. and Vanderwal, C. D. (2006) Synthesis of nitrogen heterocycles by the ring opening of pyridinium salts. *Angewandte Chemie-International Edition*, **45**, 7803-7806.

42. Michels, T. D., Kier, M. J., Kearney, A. M. and Vanderwal, C. D. (2010) Concise Formal Synthesis of Porothramycins A and B via Zincke Pyridinium Ring-Opening/Ring-Closing Cascade. *Organic Letters*, **12**, 3093-3095.

43. Neres, J., Wilson, D. J., Celia, L., Beck, B. J. and Aldrich, C. C. (2008) Aryl acid adenylating enzymes involved in siderophore biosynthesis: fluorescence polarization assay, ligand specificity, and discovery of non-nucleoside inhibitors via high-throughput screening. *Biochemistry*, **47**, 11735-11749.

44. Drake, E. J., Cao, J., Qu, J., Shah, M. B., Straubinger, R. M. and Gulick, A. M. (2007) The 1.8 A crystal structure of PA2412, an MbtH-like protein from the pyoverdine cluster of Pseudomonas aeruginosa. *The Journal of Biological Chemistry*, **282**, 20425-20434.

45. Reger, A. S., Carney, J. M. and Gulick, A. M. (2007) Biochemical and crystallographic analysis of substrate binding and conformational changes in acetyl-CoA synthetase. *Biochemistry*, **46**, 6536-6546.

46. Samel, S. A., Wagner, B., Marahiel, M. A. and Essen, L. O. (2006) The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide. *Journal of Molecular Biology*, **359**, 876-889.

47. Drake, E. J., Nicolai, D. A. and Gulick, A. M. (2006) Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain. *Chemistry & Biology*, **13**, 409-419.

48. May, J. J., Kessler, N., Marahiel, M. A. and Stubbs, M. T. (2002) Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 12120-12125.

49. Keating, T. A., Marshall, C. G., Walsh, C. T. and Keating, A. E. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nature Structural Biology*, **9**, 522-526.

50. Bruner, S. D., Weber, T., Kohli, R. M., Schwarzer, D., Marahiel, M. A., Walsh, C. T. and Stubbs, M. T. (2002) Structural basis for the cyclization

of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure*, **10**, 301-310.

51. Keasling, J. D. (2010) Manufacturing Molecules Through Metabolic Engineering. *Science*, **330**, 1355-1358.

52. Li, W., Chou, S., Khullar, A. and Gerratana, B. (2009) Cloning and characterization of the biosynthetic gene cluster for tomaymycin, an SJG-136 monomeric analog. *Applied and Environmental Microbiology*, **75**, 2958-2963.

53. Li, W., Khullar, A., Chou, S., Sacramo, A. and Gerratana, B. (2009) Biosynthesis of sibiromycin, a potent antitumor antibiotic. *Applied and Environmental Microbiology*, **75**, 2869-2878.

54. McAlpine, J. B., Banskota, A. H., Charan, R. D., Schlingmann, G., Zazopoulos, E., Piraee, M., Janso, J., Bernan, V. S., Aouidate, M., Farnet, C. M., Feng, X., Zhao, Z. and Carter, G. T. (2008) Biosynthesis of diazepinomicin/ECO-4601, a Micromonospora secondary metabolite with a novel ring system. *Journal of Natural Products*, **71**, 1585-1590.

55. Ames, B. D. and Walsh, C. T. (2010) Anthranilate-activating modules from fungal nonribosomal peptide assembly lines. *Biochemistry*, **49**, 3351-3365.

CHAPTER V

FUTURE DIRECTIONS

The research described within the previous chapters has been the initial biosynthetic investigations of both K-26 and anthramycin. As these are initial and preliminary investigations more experiments will be required to fully investigate the biosynthesis of these molecules. Herein, I will discuss possible paths of investigations.

K-26

Unfortunately, we have not been able to unambiguously identify the gene cluster responsible for K-26 biosynthesis by genomic analysis. The ability to identify the proteins responsible for K-26 biosynthesis is limited by the production levels of K-26. Correspondingly, the proteins involved in K-26 are presumably also produced at extremely low levels. In order to isolate the proteins responsible for K-26 biosynthesis, it will be necessary to increase the production level of K-26 in *A. hypotensionis* or identify another producer with higher production levels. To increase production levels of K-26 in *A. hypotensionis* or another producer, mutagenesis experiments were begun by Dr. Jaeheon Lee using UV-radiation.

The results of the aforementioned experiments seem to indicate that the *N*-acetyltransferase responsible for acetylating dK-26 may not be clustered with the rest of the genes responsible for K-26 biosynthesis. Speculatively, if

acetylation is occurring through off-target activity of a variety of *N*-acetyltransferases and peptide bond formation is through a promiscuous ligase, the only genes required for K-26 biosynthesis are those encoding the proteins responsible for AHEP formation. This lack of clustering may limit the handles available for investigation to AHEP formation. Additional difficulties lie in the lack of information about the precursors of AHEP. From the stable-labeled precursor incorporation experiments, we know that AHEP is derived from L-tyrosine and AHEP is a discrete precursor of K-26, however, during the previously described experiments, at no time was dK-26 or K-26 formation observed when L-isoleucine, L-tyrosine and AHEP were incubated with cell-free extract of *A. hypotensionis*. In order to identify the proteins required for AHEP formation, the phosphate source and additional cofactors will have to be identified. Additionally, as AHEP formation is expected to be the rate limiting step in K-26 biosynthesis, it will be necessary to develop an extremely sensitive assay for AHEP formation.

In order to do this, the best course of action would require the ability to detect AHEP in cell-free extract. AHEP is easily separated from contaminating salts by using a Hypercarb column. Initial experiments with a full gradient from 0% acetonitrile (100% 20 mM ammonium acetate pH 6 with 0.1% TEA) to 100 % acetonitrile will provide a good starting point for method development for LC/MS methodology. After development of the LC/MS method, limit of detection experiments must be undertaken. In order to detect AHEP formation, it will be necessary to detect 0.01% AHEP in cell-free extract. To do this, L-tyrosine, a

phosphate mixture and AHEP of known concentrations are added to cell-free extract. The cell-free extract is heated to 95 ^{o}C, to denature the proteins and centrifuged to remove the precipitated proteins. The supernatant would then be subjected to LC/MS analysis. After limits of detection experiments in both full scan and SRM (AHEP readily fragments, verify by product ion scanning), initial cell-free experiments can be undertaken. If AHEP still cannot be detected, it may be necessary to further concentrate AHEP from the cell-free extract. To do this, I would recommend using either a gallium or zinc affinity column, similar to nickel affinity used for protein purification. This would necessitate identifying another compound capable of eluting AHEP from the column.

Once the detection method is developed and AHEP can be detected in cell-free extraction, identification of the phosphate source can be undertaken. As mentioned in Chapter III, L-tyrosine is an immediate precursor of AHEP, however, the phosphate source and any required cofactors remain unidentified. Under cell-free extract conditions, no supplementation with phosphate or cofactors should be necessary to observe conversion of tyrosine to AHEP. To identify the phosphate source, incubation of a co-factor mix containing all possible co-factors with a mix of possible phosphate sources should increase the conversion of tyrosine to AHEP in cell-free extract. After formation of AHEP can be detected, process of elimination can first identify the phosphate source and then identify the required co-factors.

Subsequent to the identification of the phosphate source and the required co-factors, protein fractionation can be performed using traditional bioactivity

guided fractionation similar to that discussed in Chapter III. Proteomics analysis of active protein fractions will identify prospective proteins. Hopefully, possible candidates can be eliminated by correlating identified prospective proteins with a gene cluster with either an NRPS or genes for isoleucine or tyrosine biosynthesis.

Complementary to reverse genetic experiments, heterologous expression of A-domains within prospective K-26 gene clusters has been undertaken. While this has been an ongoing route to identify of the K-26 biosynthetic cassette, expression of K-26 A-domains in heterologous hosts has proved to be challenging, limiting the success of this endeavor. Recent reports have indicated that small MbtH-like proteins often found near or within NRPS gene clusters may play a role in increasing the solubility and expression levels of A-domains when co-expressed in heterologous hosts. Future work co-expressing these proteins may help to identify and/or eliminate prospective gene clusters involved in K-26 biosynthesis.

Anthramycin

Unlike the possible intermediates of the MHA biosynthetic pathway, the substrates necessary to investigate the biosynthesis of the dehydroproline acrylamide hemisphere are not readily available. A short synthetic route will provide the amide. Our synthetic route also yields the aldehyde and acid proposed biosynthetic precursors of the amide (See Chapter IV). *Orf22* was successfully synthesized by GenScript and is currently subcloned into the

pET28(a) vector. Successful expression and purification of Orf22 will allow for biochemical characterization of the enzyme. Orf22 will be tested for activity with the dehydroproline acrylamide substrate with both pyrophosphate exchange and the T-loading assay. Additionally, investigations of the biosynthesis of the amide can be investigated using gene knockouts and chemical complementation. Through slight adjustments of the synthetic pathway for the amide, proposed substrates for *orf2* (aldehyde) and *orf1* (carboxylic acid), can be synthesized.

In addition to assay based biochemical investigations, the small dimodular nature of the NRPS within the anthramycin gene cluster provides the perfect system for protein crystallography of an entire NRPS. The large size of the total NRPS systems limits the ability to crystallize the protein complex successfully.[1-7] Co-expression of Orf21 and Orf22 and *in vitro* production of anthramycin will provide the groundwork for crystallography work. Future crystallization of the NRPS of anthramycin may provide mechanistic insight into the protein-protein interactions within these complexes. Additionally, as more and more genome sequences for PBD producers are becoming accessible, it may be possible to genetically modify the A-domains to bind non-native substrates to produce novel molecules with superior or novel biological activities.[8]

REFERENCES

1. Drake, E. J., Cao, J., Qu, J., Shah, M. B., Straubinger, R. M. and Gulick, A. M. (2007) The 1.8 A crystal structure of PA2412, an MbtH-like protein from the pyoverdine cluster of Pseudomonas aeruginosa. *The Journal of Biological Chemistry*, **282**, 20425-20434.

2. Reger, A. S., Carney, J. M. and Gulick, A. M. (2007) Biochemical and crystallographic analysis of substrate binding and conformational changes in acetyl-CoA synthetase. *Biochemistry*, **46**, 6536-6546.

3. Samel, S. A., Wagner, B., Marahiel, M. A. and Essen, L. O. (2006) The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide. *Journal of Molecular Biology*, **359**, 876-889.

4. Drake, E. J., Nicolai, D. A. and Gulick, A. M. (2006) Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain. *Chemistry & Biology*, **13**, 409-419.

5. May, J. J., Kessler, N., Marahiel, M. A. and Stubbs, M. T. (2002) Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 12120-12125.

6. Keating, T. A., Marshall, C. G., Walsh, C. T. and Keating, A. E. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nature Structural Biology*, **9**, 522-526.

7. Bruner, S. D., Weber, T., Kohli, R. M., Schwarzer, D., Marahiel, M. A., Walsh, C. T. and Stubbs, M. T. (2002) Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure*, **10**, 301-310.

8. Keasling, J. D. (2010) Manufacturing Molecules Through Metabolic Engineering. *Science*, **330**, 1355-1358.