Studies of Key Enzymes Involved in the Biosynthesis of the Enediyne Antitumor Antibiotics Neocarzinostatin and C-1027

Author: Heather A. Cooke

Boston College

The Graduate School of Arts and Sciences

Department of Chemistry

STUDIES OF KEY ENZYMES INVOLVED IN THE BIOSYNTHESIS OF THE ENEDIYNE ANTITUMOR ANTIBIOTICS NEOCARZINOSTATIN AND C-1027

A dissertation

by

HEATHER ANGELINA COOKE

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

December 2009

STUDIES OF KEY ENZYMES INVOLVED IN THE BIOSYNTHESIS OF THE ENEDIYNE ANTITUMOR ANTIBIOTICS NEOCARZINOSTATIN AND C-1027

Heather A. Cooke and Professor Steven D. Bruner (Research Advisor)

Abstract

The enediyne antitumor antibiotics are produced by complex biosynthetic machinery in acetomycetes. This dissertation will focus on the study of three enzymes involved in key steps in the biosynthesis of two enediynes, neocarzinostatin and C-1027.

Neocarzinostatin is biosynthesized by a number of enzymes that synthesize and decorate the enediyne core and the peripheral moieties. NcsB1 is one enzyme involved in functionalizing the naphthoic acid portion of neocarzinostatin, a key group involved in binding to target DNA duplexes. The enzyme has been shown to be a promiscuous (*S)*-adenosylmethionine-dependent *O*-methyltransferase responsible for methylating a variety of hydroxynaphthoic acids. Multiple crystal structures of NcsB1 cocomplexed to substrate and/or cofactor have been solved. These structures revealed a displacement of the C-terminal domain when not bound to substrate, a movement that likely opens up the active site for naphthoate binding. Additionally, the ternary complex structure of 1,4-dihydroxynaphthoic acid, (*S)*-adenosylhomocysteine, and NcsB1 was solved and showed a rotation of this alternate substrate in the binding pocket, allowing for methylation. These results led us to probe NcsB1 activity using active site mutants, demonstrating altered substrate specificity and revealing key residues in substrate binding.

The final step of neocarzinostatin biosynthesis involves multiple enzymes that convergently assemble the multiple biosynthetic intermediates to form the chromophore.

NcsB2, originally proposed to catalyze the attachment of the naphthoic acid moiety to the enediyne core, has been characterized *in vitro*. Studies into its substrate specificity as an adenylation domain led to a revised biosynthetic pathway of 2-hydroxy-7-methoxy-5-methyl naphthoic acid. Instead of catalyzing the attachment of an enzyme bound naphthoic acid to the enediyne core, NcsB2 was found to act as a CoA-ligase, activating a variety of naphthoic acids and forming a naphthoyl-CoA intermediate. The results of these studies present an outstanding opportunity to produce novel analogs of neocarzinostatin by manipulating its biosynthesis.

C-1027 is an architecturally similar enediyne that is also biosynthesized in a convergent route. C-1027 is a member of a class of enediynes that contains a functionalized β-tyrosine derived from L-tyrosine. The first catalytic step towards this β-tyrosine moiety is achieved by *Sg*TAM, a tyrosine aminomutase that catalyzes a 2,3-amino shift on L-α-tyrosine to form (*S*)-β-tyrosine. The first X-ray crystal structure of *Sg*TAM was recently solved by our group, revealing structural homology to ammonia lyases. Through site-directed mutagenesis, X-ray crystallography, and biochemical analysis, residues that influence the mechanism by which *Sg*TAM catalyzes this difficult transformation were explored. From these studies, the enzymatic base and other pertinent residues involved in catalysis have been identified. In addition, residues that close the tunnel leading to the active site, thought to play a key role in mutase activity, were probed. Further study of rational mutants of *Sg*TAM will allow us to engineer its activity to alter its substrate specificity and the type of product it produces.

Acknowledgements

There are many people I must thank for contributing to my life over the past five years. First, I must thank my advisor, Prof. Steven D. Bruner, who took on this novice biochemist and trusted that I could *learn*. He has convinced me that the best way to understand something is by doing it, trial and error. It takes more time but the lessons are that much more valuable. He always has been more than willing to listen and offer advice when needed. I am thankful he let me be a part of "The Bruner Lab."

Of course, I am forever indebted to my labmates. Elisha Fielding has been my baymate (the bay-of-rage) for nearly the entire time. It will be strange not to be sitting back-to-back soon. Nearly every idea I had was discussed with her prior to execution. Thanks to all the other current and past lab members, including Dr. Carl Christianson and Tim Montavon, who both worked on the *Sg*TAM project with me, and the undergraduates who contributed greatly to the NCS projects—Elizabeth Guenther, Meg Griffin, Ed Feris, and Dan Chung. Also, thanks to Eric Dimise, Elisha, Tim, and Heather Condurso for critical reading of this document. I am also grateful to the members of the Boston College Chemistry Department. It is a competitive environment but in the most constructive of ways. Every lab is more than generous with chemicals, supplies, and knowledge.

There are multiple chemistry professors who have had an enormous influence on me. My professors at Trinity College put me on the path I am on now. Profs. Sharon Stickley and Lisa Nestor convinced me that I *could* do chemistry and that the field *needed* me. There is not much more that needs to be said to an unsure undergraduate. Also, Prof. Timothy Curran, my professor and undergraduate research

advisor, stood by my side as I ran my first columns. He encouraged me to always keep moving forward.

I grew up among professors of organizational behavior, each of whom has helped me in countless ways. My mother, Prof. Denise Rousseau, is a master of her field. She has given me endless support throughout my life, especially during graduate school. My father, Prof. Robert Cooke, likely gave me the genes to become a chemist even though he would probably deny it and has been so supportive over the past few years. My stepfather, Prof. Paul Goodman, is the true professor. His expectations for me to make an impact on people's lives will continue to shape my choices. Throughout my PhD he has lent me an ear daily and always checked to make sure that my protein was "expressing itself." My stepmother, Dr. Janet Szumal, taught me how to cook. You absolutely cannot trust a chemist who cannot cook. I cannot begin to thank my parents enough. I am so fortunate to have each of them in my life.

I have also had the support of many siblings, especially my twin sister Jessica. She is my constant sidekick in life and our daily phone conversations kept me feeling like we lived five blocks away instead of five states. I must offer endless thanks and gratitude to my Trinity girls, especially the ones who live nearby and whom I am now leaving behind. The first five years with them made me who I am. These last five years with them kept me sane. And finally thanks to Dr. Brian D. Steinberg, my best friend and the person I want by my side in every battle of life. His high expectations and unending support have made me a better chemist and the last five years would have been far less fun without him.

Table of Contents

Chapter 1: History, Activity, and Biosynthesis of Enediyne Antitumor Antibiotics

1.1 Overview 2

1.2 Discovery of enediynes 3

1.3 Apoprotein and peripheral moieties 5

1.4 Mechanism of action 6

1.5 Therapeutic applications 9

1.6 Biosynthesis 10

1.7 Future outlook 14

1.8 References 16

Chapter 2: Molecular Basis of Substrate Promiscuity for the SAM-Dependent *O*-Methyltransferase NcsB1 Involved in the Biosynthesis of the Enediyne Antitumor Antibiotic Neocarzinostatin

2.1 Introduction 22

2.2 Crystallization and x-ray structure determination 26

2.3 Overview of the structures 26

2.4 SAM/SAH binding pocket 32

2.5 Naphthoic acid binding pocket 33

2.6 NcsB1 residues involved in methyltransferase chemistry 34

2.7 Binding of 1,4-dihydroxy-2-naphthoic acid in active site of NcsB1 35

2.8 Probing the determinants of the substrate specificity for NcsB1 36

2.9 Conclusion 39

2.10 Experimentals 40

2.11 References 50

Chapter 3: Characterization of NcsB2 as a Promiscuous Naphthoic Acid/Coenzyme A Ligase Integral to the Biosynthesis of the Enediyne Antitumor Antibiotic Neocarzinostatin

3.1 Introduction 54

3.2 Initial studies 56

3.3 Characterization of NcsB2 as a CoA ligase 58

3.4 Revision of biosynthetic pathway 60

3.5 Substrate specificity of NcsB2 60

3.6 Crystallization attempts 62

3.7 Truncated constructs of NcsB2 65

3.8 Conclusion 67

3.9 Experimentals 68

3.10 References 96

Chapter 4: Biochemical and Structural Investigations into the Mechanism of the Tryosine Aminomutase SgTAM

4.1 Introduction 99

4.2 Activity assays of mutants 107

4.3 Selectivity switch, His93 107

4.4 A potential MIO modulator, Tyr308 108

4.5 A possible mutase/lyase switch, Tyr415 111

4.6 The predicted enzymatic base, Tyr63 112

4.7 Tunnel mutants 114

4.8 Binding assays of inactive enzymes 118

4.9 Folding as determined by circular dichroism 121

4.10 Conclusion 122

4.11 Experimentals 123

4.10 References 131

Appendix 1: Discussion and Experimentals for the NcsEs and for the Synthesis of the Benzoxazolinate Moiety of C-1027

A1. 1 Biosynthesis of the neocarzinostatin enediyne core 135

A1.2 Experimentals for NcsEs 136

A1.3 Benzoxazolinate moiety 149

A1.4 References 156

Appendix 2: Structure and Chemistry of 4-Methylideneimidazole-5-one Containing Enzymes

A2.1 Summary of recent advances 158

A2.2 Introduction 158

A2.3 MIO-based ammonia lyases 160

A2.4 MIO-based aminomutases 161

A2.5 Structural characterization of MIO-based ammonia lyases 163

A2.6 Structural basis for MIO aminomutase activity 165

A2.7 Small molecule probes into the mechanism of MIO-enzymes 166

A2.8 Unified mechanism for MIO-based enzymes 169

A2.9 Utilization of MIO-aminomutases in chemoenzymatic synthesis applications 171

A2.10 Rational engineering of lyase and aminomutase activity 172

A2.11 Conclusions 173

A2.12 References 174

Abbreviations

A domain	adenylation domain
AIP	2-aminoindan-2-phosphonate
AMP	adenosine monophosphate
ATP	adenosine triphosphate
BSA	bovine serum albumin
CD	circular dichroism
CoA	coenzyme A
DAS	Dense Alignment Surface Method
DBU	1,8-diazabicycloundec-7-ene
DCC	*N*,*N*'-dicyclohexylcarbodiimide
DCM	dichloromethane
DHN	1,4-dihydroxy naphthoic acid
DM	double mutant
DMAP	4-dimethylaminopyridine
DMF	dimethylformamide
DMSO	dimethylsulfoxide
dNTPs	dinuleotide triphosphates
DTT	dithiothreitol
ESI	electrospray ionization
HAL	histidine ammonia lyase
HCA	*p*-hydroxy cinammic acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

His_6	hexahistidine
HOBt	hydroxybenzotriazole
HPLC	high pressure liquid chromatography
HRMS	high resolution mass spectrometry
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
LB	Luria Bertani
LC-MS	liquid chromatography mass spectrometry
LDA	lithium diisopropyl amine
LRMS	low resolution mass spectrometry
mAb	monocolnal antibody
MBP	maltose binding protein
MIO	4-methylideneimidazole-5-one
MNA	5-methyl-2-hydroxy naphthoic acid
MTase	methyltransferase
NCS	neocarzinostatin
NHS	*N*-hydroxysuccinimide
Ni-NTA	nickel-nitrilotriacetic acid
NMP	*N*-methyl-2-pyrrolidone
NMR	nuclear magnetic resonance
NRPS	nonribosomal peptide synthetase
OD	optical density
ORF	open reading frame

PAL	phenylalanine ammonia lyase
PAM	phenylalanine aminomutase
PCP	peptidyl-carrier protein
PDB	protein database
PEG	polyethylene glycol
PKS	polyketide synthase
PMSF	phenylmethanesulphonylfluoride
PPi	pyrophosphate
RMSD	root mean squared difference
SAH	(*S*)-adenosyl-L-homocysteine
SAM	(*S*)-adenosyl-L-methionine
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SMANCS	poly(styrene-co-maleic anhydride)-conjugated form of NCS
TAL	tyrosine ammonia lyase
TAM	tyrosine aminomutase
TEA	triethylamine
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TMAO	trimethylamine *N*-oxide
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
UV	ultraviolet
WT	wild-type

Chapter 1

History, Activity, and Biosynthesis of Enediyne Antitumor Antibiotics

1.1 Overview

Enediyne antitumor antibiotics are a class of highly cytotoxic secondary metabolites. They are characterized by a core motif consisting of two acetylene units conjugated to a *Z*-olefin within a larger nine- or ten-membered ring. Additionally, many enediynes have a DNA recognition unit and a nucleophilic component that triggers the formation of the active species. The active form of an enediyne is a diradical species, which acts as a "warhead" causing single and/or double stranded DNA breakage, leading to apoptosis. Enediynes have gained popularity because of their potent antitumor activity. Their biosynthetic machinery can likely be cloned and reproduced using *E. coli* as a heterologous host, adding to their value as drugs (*1*). Still, the extreme cytotoxicity of these structurally complex necessitates control of their reactivity for use as therapeutics (*2*). The nine-membered subfamily of enediynes are composed of a small molecule chromophore bound to an apoprotein, which sequesters and stabilizes the reactive enediyne and aids in delivery to target cells (*3*). The biosynthetic gene clusters of multiple enediynes from actinomycetes have recently been sequenced and annotated, allowing detailed investigation into their complex biosynthesis (*4-9*). The pathways utilize diverse enzymology to construct the natural products, including the use of polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) machinery. Alteration of the peripheral moieties surrounding the enediyne core may be a way to mediate the reactivity and specificity of these therapeutically useful compounds (*2*). This dissertation will focus on structural and mechanistic studies into the biosynthesis of two enediyne antitumor antibiotics, neocarzinostatin (NCS) and C-1027 (Figure 1.1).

1.2 Discovery of enediynes

In 1965, Ishida et al reported evidence for the first enediyne from extracts of *Streptomyces carzinostaticus* (*10*). The two-component complex showed substantial activity against a variety of gram-positive bacteria. Additionally, the complex was effective against a strain of leukemia cells but was found to spare non-cancerous cells. The remarkable ability of this compound to target cancer cells that were typically impenetrable by the current therapeutics of the time made this natural product a compound of interest. Unfortunately because of its instability, it was not until 20 years later that the chemical structure of the NCS chromophore was elucidated (Figure 1.1). It was found to contain a novel bicyclo[7.3.0]dodecadiynene ring system (*11*), which was decorated with a deoxyamino sugar and a naphthalene moiety, making it the archetypal member of the soon to be growing class of nine-membered endiynes. The absolute stereochemistry of NCS was determined in subsequent studies (*12*). The discovery of two additional classes of enediynes, the calicheamicins from *Micromonospora echinospora* ssp. calichensis (*13, 14*) and the esperamicins from *Actinomadura verrucosospora (15, 16)* (Figure 1.1), sparked further interest into this now growing class of natural products. The calicheamicins and esperamicins differ significantly from NCS in that they both possess a bicyclo[7.3.1]tridecadiynene system, and constituted a new class of antitumor antibiotics: the ten-membered enediynes. Both are decorated with deoxyamino sugars and are functionalized with a trisulfide responsible for converting the enediyne core to its active species. A fourth structural was isolated in 1989; dynemycin A (*17*) was the third ten-membered enediyne to be added to this class. Since then,

additional members of the nine-membered enediyne class have been discovered, including C-1027 (*18, 19*), kedarcidin (*20*), and maduropeptin (*21*) (Figure 1.1).

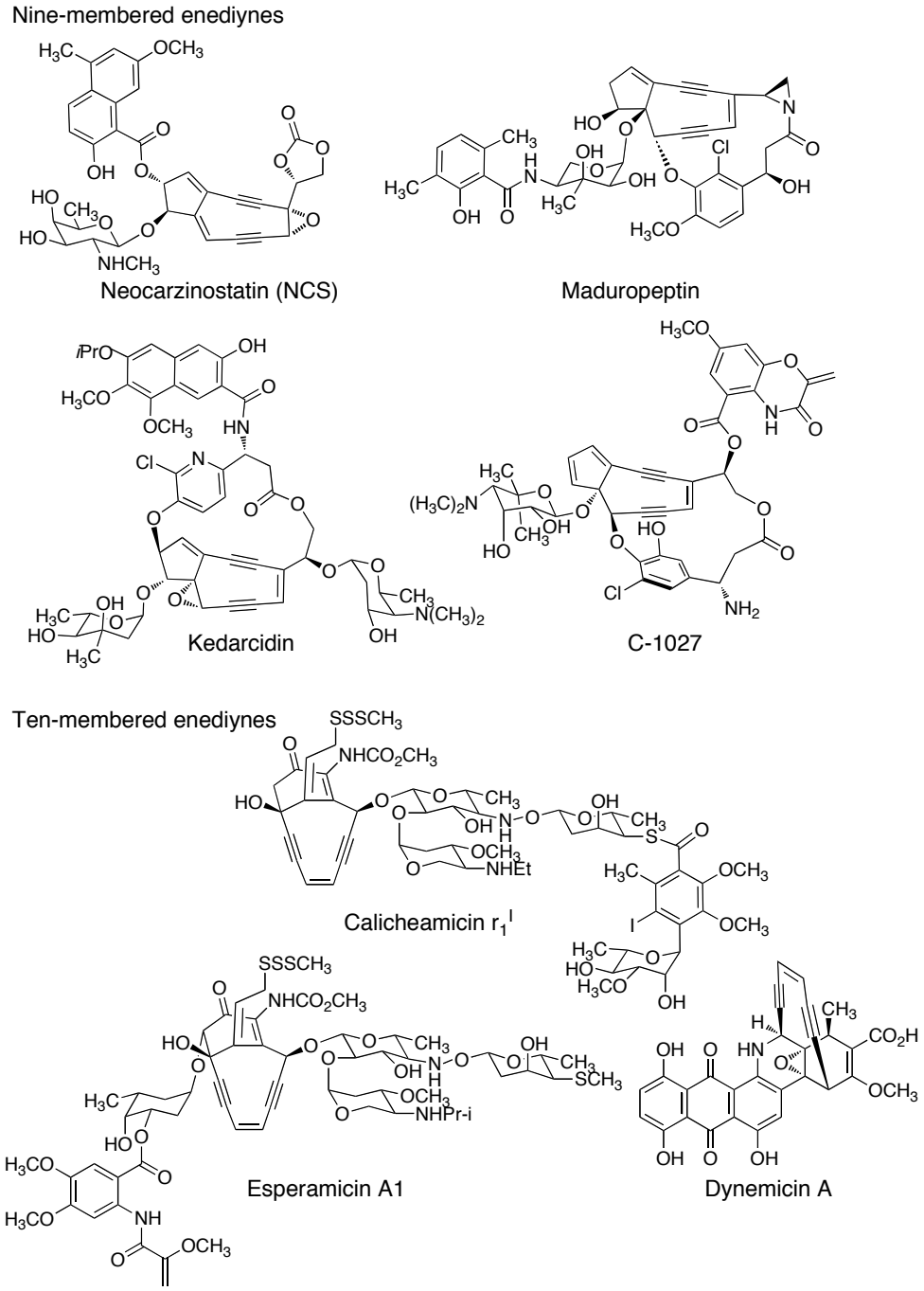


Figure 1.1: Nine- and ten-membered enediyne chromophores.

1.3 Apoprotein and peripheral moieties

Unlike ten-membered enediynes, nine-membered enediynes readily cycloaromatize to their active, diradical species and therefore require an outside source of stabilization so that they are not toxic to the producing organism. NCS naturally occurs in complex with a 147 amino acid protein, called the NCS apoprotein (NcsA). Early on, it was realized that the biological activity of NCS is imparted by the chromophore and the role of the apoprotein is to modulate activity by stabilizing, sequestering, and transporting the chromophore. Studies towards exploitation of the apoprotein to improve the stability and delivery of NCS have been reportd (*22*).

C-1027, isolated from *Streptomyces globisporus* in 1988 *(23-25)* and initially structurally characterized in 1993 (*18, 19*), also belongs to the chromoprotein class of antitumor antibiotics. Like NCS, the enediyne core of C-1027 is also decorated with peripheral moieties, including a benzoxazolinate, a functionalized β-tyrosine, and a deoxyamino sugar and is bound to the 119 amino acid apoprotein CagA. CagA and NcsA share 40% sequence identity.

Both the benzoxazolinate from C-1027 and the naphthalene unit from NCS play a role in binding to the apoprotein (*26-30*). From structural studies of both apoproteins, the enediynes rely on crucial π-stacking interactions, which preclude the enediynes from cyclizing (*30-32*). Upon release from the apoprotein and following binding to DNA, the chromophores convert to their corresponding active diradical species. The apoprotein of NCS also acts to sequester the chromophore from endogenous thiols, which trigger the electronic rearrangement required to form the reactive diradical species.

1.4 Mechanism of action

The reactive agent of all enediynes is a carbon-centered diradical species. The general mode of action to form this species begins with the enediyne binding to DNA. Diradical formation occurs automatically or following an electronic rearrangement, which then causes DNA lesions via single-stranded and double-stranded cleavage. The mechanism by which individual enediynes reach this species varies slightly and mechanisms for C-1027 and NCS will be discussed here.

C-1027 is released from the apoprotein and intercalates DNA via its benzoxazolinate moiety. Because the apoprotein is no longer stabilizing the enediyne core through π-stacking interactions, there is sufficient orbital overlap for a Bergman cyclization to occur, forming the active benzenoid diradical species, which in turn abstracts hydrogens from target DNA (Figure 1.2).

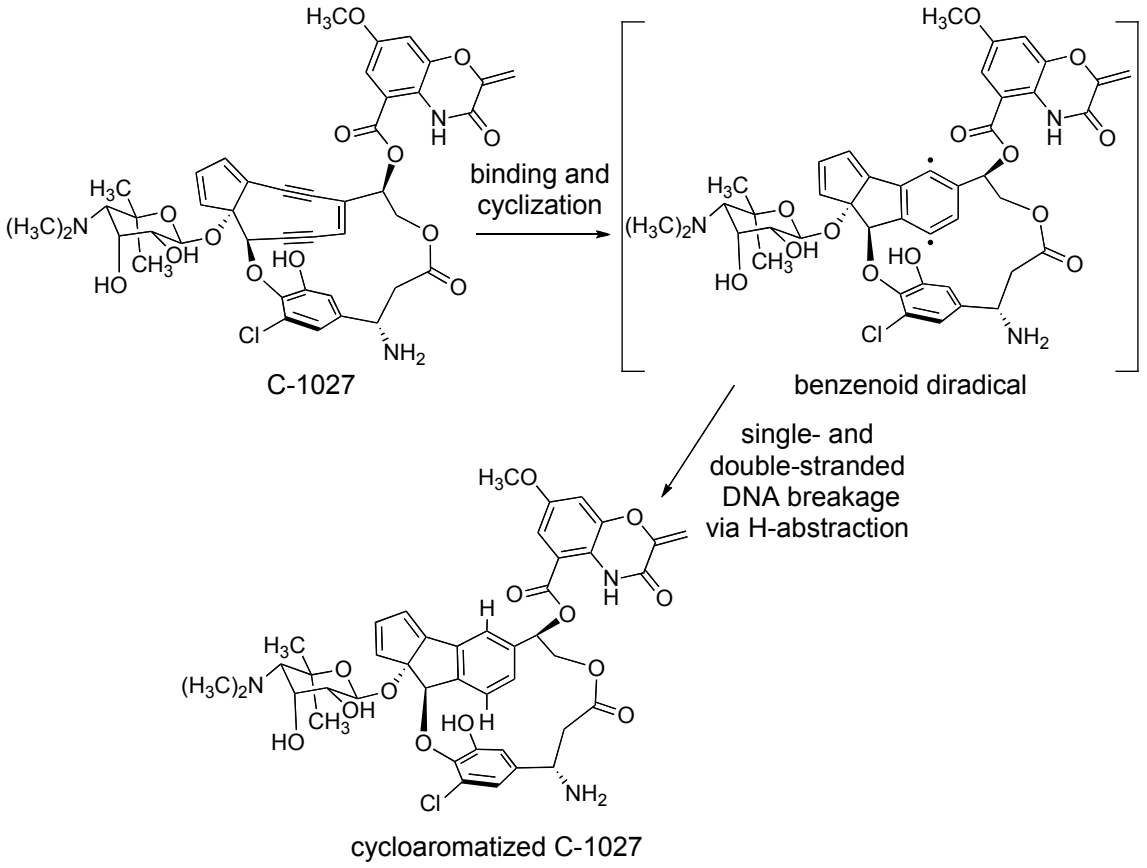


Figure 1.2: Mechanism for C-1027 activity. Upon release from the apoprotein, the enediyne intercalates DNA. Next, the enediyne rapidly undergoes a Bergman cyclization, forming a benzenoid diradical species, which is responsible for hydrogen abstraction from the deoxyribose sugar of DNA.

Contrarily, NCS requires activation via an endogenous thiol to affect a Myers-Saito rearrangement (Figure 1.3). The NCS chromophore intercalates into the minor groove of DNA via its naphthalene unit, positioning the enediyne core for hydrogen abstraction from the deoxyribose backbone (*33*). Upon nitrogen-assisted (from the deoxyamino-sugar moiety) nucleophilic attack by a thiol at C12 (*30*), the enediyne core rearranges, resulting in opening of the epoxide and cumulene formation. At this point, the diradical species is formed by rapid cycloaromatization, followed by hydrogen abstraction from the DNA backbone.

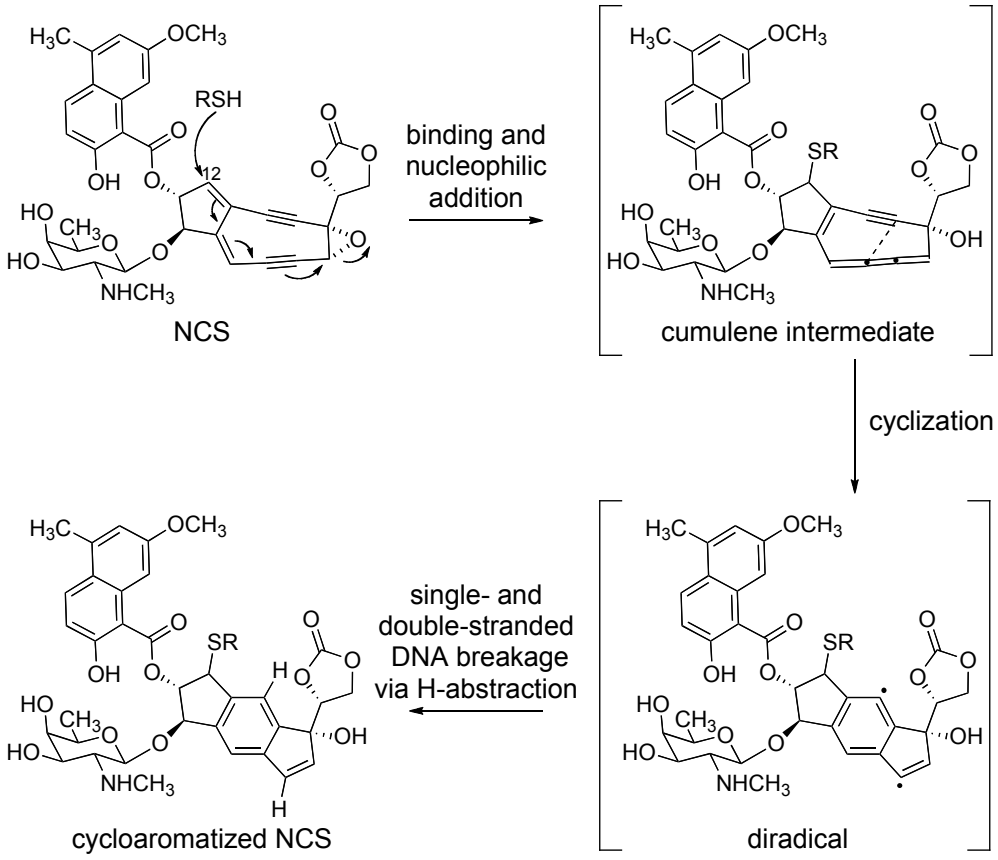


Figure 1.3: Mechanism for NCS activity. Release from the apoprotein and intercalation of DNA leads to a Myers-Saito cyclization triggered by nucleophilic addition of a thiol into C12. Electronic rearrangement of the enediyne core results in formation of a cumulene intermediate, which rapid collapses to the active diradical species.

In both cases, hydrogen abstraction occurs at the deoxyribose sugar of the DNA strand, resulting in the formation of deoxyribose carbon-centered radicals. Rapid reaction with molecular oxygen results in the production of abasic sites and strand breakage, producing DNA fragments (Figure 1.4) (*34*). Double-stranded DNA breaks in

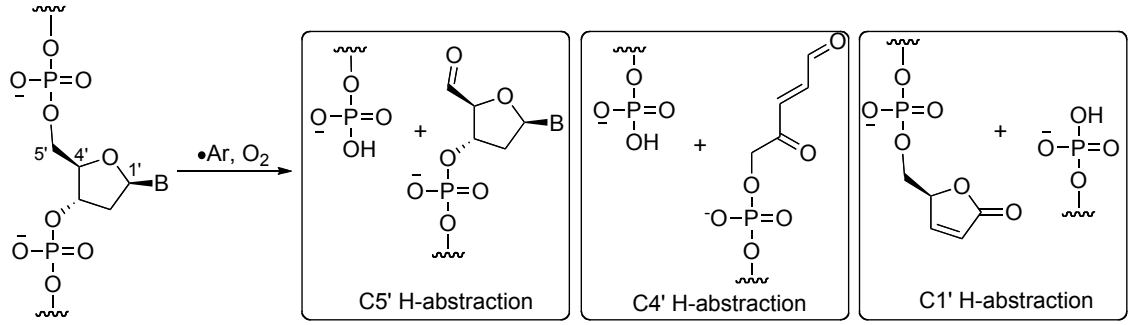


Figure 1.4: DNA cleavage products. Hydrogen abstraction occurs primarily at C5' and sometimes at C4' and C1'.

particular are difficult to repair by the host cell and therefore are highly detrimental, causing apoptosis. NCS cleavage occurs primarily at ACT/AGT trinucleotide sites, with 80% of hydrogen abstraction occurring at the C5' of A and T residues (*35-37*). Less than 20% of DNA strand breaks are a result of abstraction from C4' and C1'. C-1027 usually causes breakage at the penta-nucleotide site GTTAT/ATAAC (*38, 39*). Hydrogen abstraction typically occurs at the C4' and C5' hydrogen atoms of the A from GTTAT and ATAAC, respectively.

1.5 Therapeutic applications

While enediynes are highly toxic, they still can be of therapeutic benefit. The producing organisms are able to control their toxicity by coproducing them with an apoprotein and by decorating the highly reactive enediyne core with peripheral moieties

that can modulate their reactivity. Also, studies of their DNA cleaving activity demonstrate that they do not act randomly and that DNA binding instigates diradical formation. This controlled reactivity is very important for the development of drugs, but the ability to target cancer cells is far more crucial. Enediynes have been successfully used as anticancer drugs in covalent conjugation to a stabilizing agent or a drug carrier. NCS was the first enediyne to be manipulated in such a manner. The poly(styrene-co-maleic anhydride)-conjugated form of NCS (SMANCS) was approved for use in Japan in 1993 for treatment of liver cancer (*40-42*). Due to their higher stability, ten-membered enediynes have also found utility as chemotherapeutics. Calicheamicin conjugated to a CD33 monoclonal antibody (mAB) is now approved for use in the United States under the name Mylotarg (*43*). CD33 is expressed in most leukemia cells and therefore the calicheamicin conjugate targets these cells and is substantially less toxic to normal cells. C-1027 also showed some early success an as antihepatomic agent as a mAB-C-1027 conjugate (*44*). More recently C-1027 has been shown to be an effective radiomimetic that can specifically cause DNA strand breaks by targeting hypoxic cancer cells (*45*).

1.6 Biosynthesis

Despite efforts in developing total synthetic routes to these compounds, nine-membered enediynes are particularly difficult to access through standard organic methods due to their structural complexity and chemical instability when not in complex with an apoprotein. The biosynthetic gene clusters of C-1027 and NCS have been sequenced and annotated (*5, 7*) and both have been found to utilize interesting chemistry in order to access these complex natural products (Figure 1.5). The C-1027 gene cluster contains

sgcE, which encodes for SgcE, a novel type I polyketide synthase (PKS) (*7*). A homologous gene is found in the NCS gene cluster, *ncsE*. Both genes encode for enzymes that are responsible for producing the unsaturated polyene that makes up the carbon scaffold of the respective enediyne cores of the two natural products *(46, 47)*. In the C-1027 gene cluster, a variety of open reading frames (ORFs), *sgcE1* to *sgcE11,* were identified and share sequence homology to genes within the NCS biosynthetic gene cluster (*ncsE1 to ncsE11)* but were not homologous to any proteins of known function (*5*). These enzymes, along with the epoxide hydrolase SgcF (and the corresponding enzymes for NCS, NcsF1 and NcsF2), are likely responsible for producing an enediyne intermediate of C-1027 (Figure 1.5A). PCR amplification of PKSs from known gene clusters has resulted in a high-throughput method to access minimal PKS gene cassettes from organisms previously not known to produce enediynes (*4*).

In addition to genes responsible for encoding proteins involved in production of the enediyne core, for C-1027 six enzymes were assigned to produce the D-glucose derived deoxyamino sugar (SgcA to SgcA5), five enzymes for the functionalized β-tyrosine (SgcC to SgcC4), and six enzymes for the benzoxazolinate (SgcD to SgcD5). Additionally, each peripheral moiety was assigned a gene responsible for encoding an enzyme that would catalyze its attachment to the C-1027 enediyne core (SgcA6, SgcC5, and SgcD6) (Figure 1.5A).

Similarly for NCS, six enzymes were assigned to produce the deoxyamino sugar (NcsC to NcsC5), three enzymes for the functionalized naphthoic acid (NcsB, NcsB1, and NcsB3), followed by two additional enzymes (NcsB2 and NcsC6) that

catalyze the addition of the two moieties to the enediyne core (Figure 1.5B).

The goal of this dissertation is to study specific enzymes involved in the biosynthesis of these two enediynes. Their convergent biosynthesis could be exploited to produce enediyne analogs with altered antitumor properties by engineering one or more

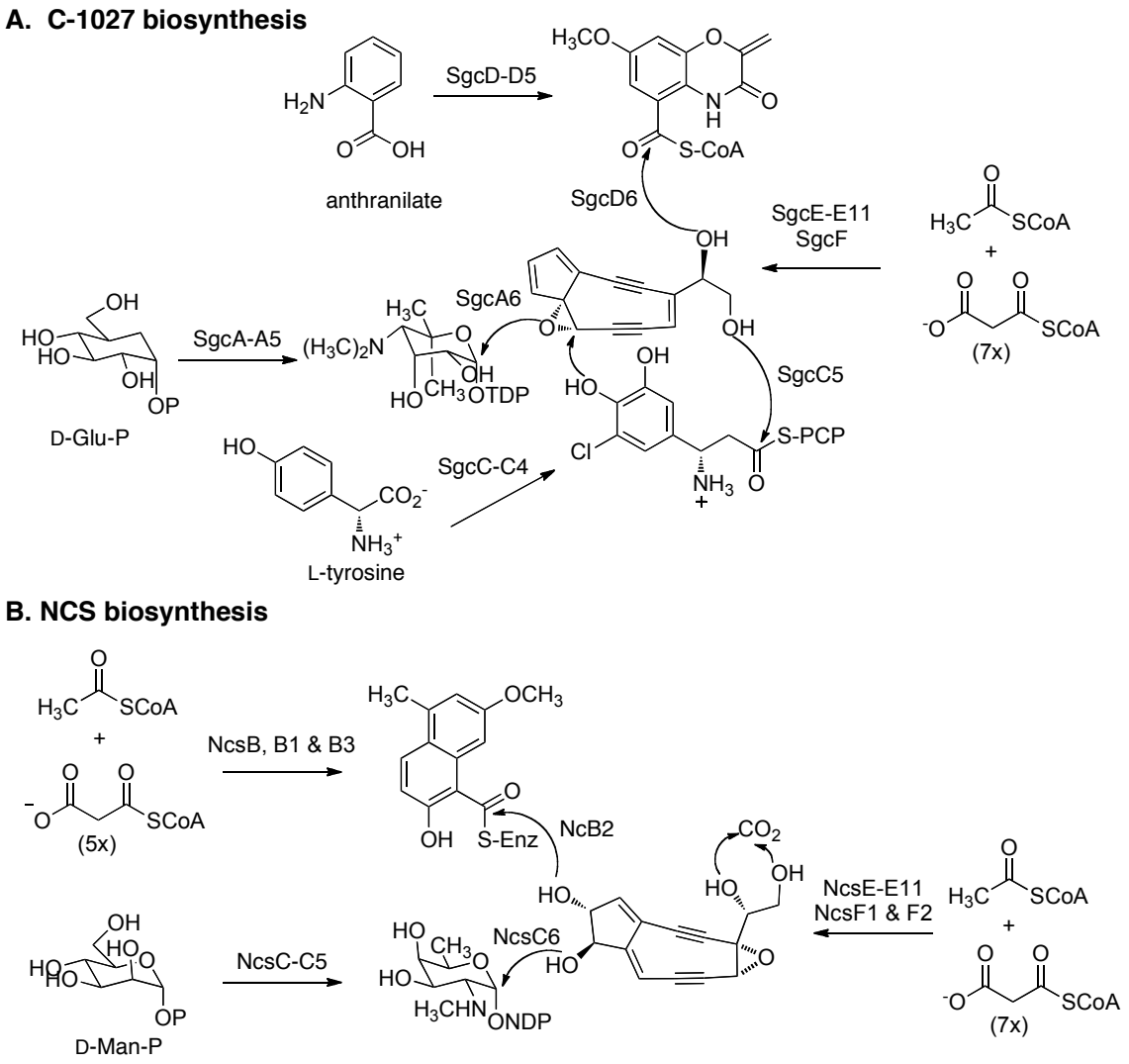


Figure 1.5: Early proposed convergent assembly. A) C-1027 requires three enzymes, SgcD6, SgcA6, and SgcC5, to attach the benzoxazolinate, deoxy-aminosugar, and β-tyrosine to the enediyne core. B) NCS is assembled with NcsB2 and NcsC6 catalyzing the attachment of the naphthoate and deoxy-aminosugar moieties to the core scaffold.

of the enzymes involved in the production or attachment of the peripheral moieties. The three enzymes that will be discussed in the following chapters are all candidates for rational engineering.

Chapter 2 will be a discussion of studies into the SAM-dependent *O*-methyltransferase NcsB1, which is involved in the biosynthesis of the naphthoate moiety of NCS (*48, 49*). Studies into the specificity of NcsB1 have proved that it regiospecifically methylates a variety of aromatic aryl acids. Structural work and site-directed mutagenesis has led to insight into specificity determinants. Chapter 3 involves studies into the promiscuity of NcsB2, a CoA-ligase originally believed to attach the functionalized naphthoic acid to the enediyne core (*50*). Studies of this enzyme have lead to significant changes to the proposed biosynthesis of NCS, which will be discussed in detail. Also, we have identified a range of substrates that are activated by NcsB2. This observed promiscuity suggests that this enzyme can be exploited to produce analogs of NCS by varying the naphthoic acid unit it activates and is subsequently attached to the enediyne core.

Chapter 4 involves studies of the enzyme *Sg*TAM, also known as SgcC4, an MIO-dependent tyrosine aminomutase responsible for producing β-tyrosine, which is subsequently functionalized by downstream enzymes and incorporated into C-1027 (*51*). This class of enzymes performs a difficult chemical transformation via a mechanism that is still under debate. Through site-directed mutagenesis and structural studies of *Sg*TAM, we are closer to understanding how it affects the conversion of L-tyrosine into β-tyrosine.

Finally, Appendix A includes a discussion of and experimentals for the cloning of three *ncsE*s and attempts at isolating, purifying and crystallizing the enzymes for which they encode. Appendix A also includes details of the synthesis of the benzoxazolinate moiety of C-1027. Appendix B is a review of MIO chemistry based largely on recent structural characterization of MIO-containing enzymes (*52*).

1.7 Future outlook

The development of new cancer treatments is a constant goal for pharmaceutical and research institutions. We hope the knowledge gained from studies of enediyne biosynthetic enzymes will aid in the design and engineering of useful therapeutics. Accessing complex natural products via organic synthesis is a significant challenge in terms of time and efficiency. Reactive enediynes are particularly difficult as synthesizing an array of unstable compounds is a challenge. Production of enediynes using rationally altered enzymatic machinery may be a facile way of accomplishing this goal. For example, we have synthesized a variety of analogs of the naphthoic acid moiety of NCS (see Chapter 3). Production of NCS analogs may be as simple as feeding the free naphthoic acid analogs into *S. carzinostaticus* media (Figure 1.6). With a combination of organic synthesis and molecular biology, accessing a variety of NCS analogs may only be limited by the specificity the NcsB2. Furthermore, rational engineering of enzymes upstream of those responsible for their convergent assembly may also be away to produce analogs. In addition to the enzymes in the naphthoic acid pathway of NCS, with proper engineering of the β-tyrosine producing *Sg*TAM, we may be able produce other β-amino acids, which may be incorporated to for C-1027 analogs. The possibilities are limited to

our knowledge of enzyme chemistry and further research will support the field of drug development.

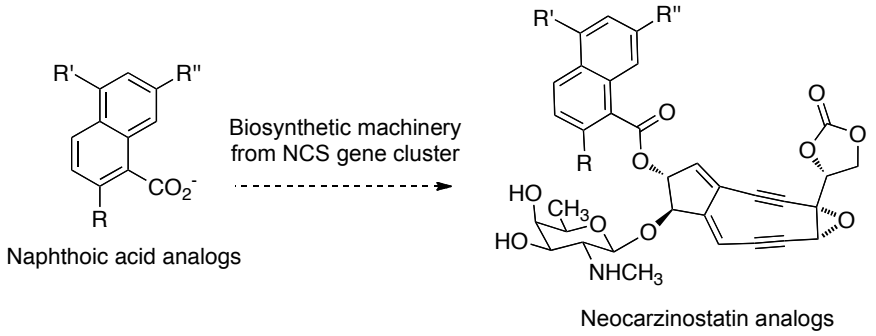


Figure 1.6: Production of NCS analogs using organic synthesis and molecular biology

1.8 References

1. Van Lanen, S. G., and Shen, B. (2008) Biosynthesis of Enediyne Antitumor Antibiotics, *Curr. Top. Med. Chem. 8*, 448-459.
2. Shen, B., Liu, W., and Nonaka, K. (2003) Enediyne natural products: biosynthesis and prospect towards engineering novel antitumor agents, *Curr. Med. Chem. 10*, 2317-2325.
3. Maeda, H., Aikawa, S., and Yamashita, A. (1975) Subcellular fate of protein antibiotic neocarzinostatin in culture of a lymphoid cell line from Burkitt's lymphoma, *Cancer Res. 35*, 554-559.
4. Liu, W., Ahlert, J., Gao, Q., Wendt-Pienkowski, E., Shen, B., and Thorson, J. S. (2003) Rapid PCR amplification of minimal enediyne polyketide synthase cassettes leads to a predictive familial classification model, *Proc. Natl. Acad. Sci. U.S.A. 100*, 11959-11963.
5. Liu, W., Nonaka, K., Nie, L., Zhang, J., Christenson, S. D., Bae, J., Van Lanen, S. G., Zazopoulos, E., Farnet, C. M., Yang, C. F., and Shen, B. (2005) The neocarzinostatin biosynthetic gene cluster from Streptomyces carzinostaticus ATCC 15944 involving two iterative type I polyketide synthases, *Chem. Biol. 12*, 293-302.
6. Udwary, D. W., Zeigler, L., Asolkar, R. N., Singan, V., Lapidus, A., Fenical, W., Jensen, P. R., and Moore, B. S. (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica, *Proc. Natl. Acad. Sci. U.S A. 104*, 10376-10381.
7. Liu, W., Christenson, S. D., Standage, S., and Shen, B. (2002) Biosynthesis of the enediyne antitumor antibiotic C-1027, *Science 297*, 1170-1173.
8. Van Lanen, S. G., Oh, T. J., Liu, W., Wendt-Pienkowski, E., and Shen, B. (2007) Characterization of the maduropeptin biosynthetic gene cluster from Actinomadura madurae ATCC 39144 supporting a unifying paradigm for enediyne biosynthesis, *J. Am. Chem. Soc. 129*, 13082-13094.
9. Ahlert, J., Shepard, E., Lomovskaya, N., Zazopoulos, E., Staffa, A., Bachmann, B. O., Huang, K., Fonstein, L., Czisny, A., Whitwam, R. E., Farnet, C. M., and Thorson, J. S. (2002) The calicheamicin gene cluster and its iterative type I enediyne PKS, *Science 297*, 1173-1176.
10. Ishida, N., Miyazaki, K., Kumagai, K., and Rikimaru, M. (1965) Neocarzinostatin, an antitumor antibiotic of high molecular weight. Isolation, physiochemical properties and biological activities., *J. Antibiot. (Tokyo) 18*, 68-76.
11. Edo, K., Mizugaki, M., Koide, Y., Seto, H., Furihata, K., Otake, N., and Ishida, N. (1985) The Structure of Neocarzinostatin Chromophore Possessing a Novel Bicyclo-[7,3,0]Dodecadiyne System, *Tet. Lett. 26*, 331-334.
12. Myers, A. G., Proteau, P. J., and Handel, T. M. (1988) Stereochemical Assignment of Neocarzinostatin Chromophore - Structures of Neocarzinostatin Chromophore Methyl Thioglycolate Adducts, *J. Am. Chem. Soc. 110*, 7212-7214.

13. Lee, M. D., Dunne, T. S., Siegel, M. M., Chang, C. C., Morton, G. O., and Borders, D. B. (1987) Calichemicins, a Novel Family of Antitumor Antiobiotics. 1. Chemistry and Partial Structure of Calichemicin *J. Am. Chem. Soc. 109*, 3464-3466.
14. Lee, M. D., Dunne, T. S., Chang, C. C., Ellestad, G. A., Siegel, M. M., Morton, G. O., McGahren, W. J., and Borders, D. B. (1987) Calichemicins, a Novel Family of Antitumor Antiobiotics. 2. Chemistry and Structure of Calichemicin., *J. Am. Chem. Soc. 109*, 3466-3468.
15. Golik, J., Clardy, J., Dubay, G., Groenewold, G., Kawaguchi, H., Konishi, M., Krishnan, B., Ohkuma, H., Saitoh, K., and Doyle, T. W. (1987) Esperamicins, a Novel Class of Potent Antitumor Antibiotics. 2. Structure of Esperamicin X., *J. Am. Chem. Soc. 109*, 3461-3462.
16. Golik, J., Dubay, G., Groenewold, G., Kawaguchi, H., Konishi, M., Krishnan, B., Ohkuma, H., Saitoh, K., and Doyle, T. W. (1987) Esperamicins, a Novel Class of Potent Antitumor Antibiotics. 3. Structures of Esperamicins A1, A2, and A1b, *J. Am. Chem. Soc. 109*, 3462-3464.
17. Konishi, M., Ohkuma, H., Matsumoto, K., Tsuno, T., Kamei, H., Miyaki, T., Oki, T., Kawaguchi, H., VanDuyne, G. D., and Clardy, J. (1989) Dynemicin A, a novel antibiotic with the anthraquinone and 1,5-diyn-3-ene subunit, *J. Antibiot. (Tokyo) 42*, 1449-1452.
18. Minami, Y., Yoshida, K.-i., Azuma, R., Saeki, M., and Otani, T. (1993) Structure of an aromatization product of C-1027 chromophore, *Tet. Lett. 34*, 2633-2636.
19. Yoshida, K.-i., Minami, Y., Azuma, R., Saeki, M., and Otani, T. (1993) Structure and cycloaromatization of a novel enediyne, C-1027 chromophore, *Tet. Lett. 34*, 2637-2640.
20. Hofstead, S. J., Matson, J. A., Malacko, A. R., and Marquardt, H. (1992) Kedarcidin, a new chromoprotein antitumor antibiotic. II. Isolation, purification and physico-chemical properties, *J Antibiot (Tokyo) 45*, 1250-1254.
21. Schroeder, D. R., Colson, K. L., Klohr, S. E., Zein, N., Langley, D. R., Lee, M. S., Matson, J. A., and Doyle, T. W. (1994) Isolation, Structure Determination, and Proposed Mechanism of Action for Artifacts of Maduropeptin Chromophore, *J. Am. Chem. Soc. 116*, 9351-9352.
22. Baker, J. R., Woolfson, D. N., Muskett, F. W., Stoneman, R. G., Urbaniak, M. D., and Caddick, S. (2007) Protein-small molecule interactions in neocarzinostatin, the prototypical enediyne chromoprotein antibiotic, *Chembiochem 8*, 704-717.
23. Hu, J. L., Xue, Y. C., Xie, M. Y., Zhang, R., Otani, T., Minami, Y., Yamada, Y., and Marunaka, T. (1988) A new macromolecular antitumor antibiotic, C-1027. I. Discovery, taxonomy of producing organism, fermentation and biological activity, *J. Antibiot. (Tokyo) 41*, 1575-1579.
24. Otani, T., Minami, Y., Marunaka, T., Zhang, R., and Xie, M. Y. (1988) A new macromolecular antitumor antibiotic, C-1027. II. Isolation and physico-chemical properties, *J. Antibiot. (Tokyo) 41*, 1580-1585.

25. Otani, T., Minami, Y., Marunaka, T., Zhang, R., and Xie, M. Y. (1988) C-1027-AG, a selective antagonist of the macromolecular antitumor antibiotic C-1027, *J. Antibiot. (Tokyo) 41*, 1696-1698.
26. Matsumoto, T., Okuno, Y., and Sugiura, Y. (1993) Specific interaction between a novel enediyne chromophore and apoprotein in macromolecular antitumor antibiotic C-1027, *Biochem. Biophys. Res. Comm. 195*, 659-666.
27. Okuno, Y., Otsuka, M., and Sugiura, Y. (1994) Computer modeling analysis for enediyne chromophore-apoprotein complex of macromolecular antitumor antibiotic C-1027, *J. Med. Chem. 37*, 2266-2273.
28. Briozzo, P., Inaka, K., Minami, Y., Otani, T., and Miki, K. (1993) Crystallization and preliminary X-ray diffraction studies of C-1027-AG, the apoprotein of the macromolecular antitumor antibiotic C-1027 from Streptomyces globisporus, *Acta. Crystallogr. D. Biol. Crystallogr. 49*, 372-374.
29. Napier, M. A., Kappen, L. S., and Goldberg, I. H. (1980) Effect of nonprotein chromophore removal on neocarzinostatin action, *Biochemistry 19*, 1767-1773.
30. Kim, K. H., Kwon, B. M., Myers, A. G., and Rees, D. C. (1993) Crystal structure of neocarzinostatin, an antitumor protein-chromophore complex, *Science 262*, 1042-1046.
31. Tanaka, T., Fukuda-Ishisaka, S., Hirama, M., and Otani, T. (2001) Solution structures of C-1027 apoprotein and its complex with the aromatized chromophore, *J. Mol. Biol. 309*, 267-283.
32. Okuno, Y., Iwashita, T., and Sugiura, Y. (2000) Structural Basis for Reaction Mechanism and Drug Delivery System of Chromoprotein Antitumor Antibiotic C-1027, *J. Am. Chem. Soc. 122*, 6848-6854.
33. Povirk, L. F., Dattagupta, N., Warf, B. C., and Goldberg, I. H. (1981) Neocarzinostatin chromophore binds to deoxyribonucleic acid by intercalation, *Biochemistry 20*, 4007-4014.
34. Dedon, P. C., and Goldberg, I. H. (1992) Free-radical mechanisms involved in the formation of sequence-dependent bistranded DNA lesions by the antitumor antibiotics bleomycin, neocarzinostatin, and calicheamicin, *Chem. Res. Toxicol. 5*, 311-332.
35. Lee, S. H., and Goldberg, I. H. (1989) Sequence-specific, strand-selective, and directional binding of neocarzinostatin chromophore to oligodeoxyribonucleotides, *Biochemistry 28*, 1019-1026.
36. Dedon, P. C., and Goldberg, I. H. (1990) Sequence-specific double-strand breakage of DNA by neocarzinostatin involves different chemical mechanisms within a staggered cleavage site, *J. Biol. Chem. 265*, 14713-14716.
37. Kappen, L. S., Ellenberger, T. E., and Goldberg, I. H. (1987) Mechanism and base specificity of DNA Breakage in intact cells by neocarzinostatin, *Biochemistry 26*, 384-390.
38. Sugiura, Y., and Matsumoto, T. (1993) Some characteristics of DNA strand scission by macromolecular antitumor antibiotic C-1027 containing a novel enediyne chromophore, *Biochemistry 32*, 5548-5553.

39. Xu, Y. J., Zhen, Y. S., and Goldberg, I. H. (1994) C1027 chromophore, a potent new enediyne antitumor antibiotic, induces sequence-specific double-strand DNA cleavage, *Biochemistry 33*, 5947-5954.
40. Maeda, H., Ueda, M., Morinaga, T., and Matsumoto, T. (1985) Conjugation of poly(styrene-co-maleic acid) derivatives to the antitumor protein neocarzinostatin: pronounced improvements in pharmacological properties, *J. Med. Chem. 28*, 455-461.
41. Maeda, H., Edo, K., and Ishida, N. (1997) *Neocarzinostatin: The past, present, and future of an anticancer drug* Springer-Verlag, New York.
42. Kobayashi, M. Y., JP), Ohtani, Go (Tokyo, JP), Sekino, Jun (Yaizu, JP), Konno, Toshimitsu (Kumamoto, JP), Maeda, Hiroshi (Kumamoto, JP). (1991) Lyophilized pharmaceutical composition of neocarzinostatin derivative, Yamanouchi Pharmaceutical Co., Ltd. (Tokyo, JP), United States.
43. Sievers, E. L., Appelbaum, F. R., Spielberger, R. T., Forman, S. J., Flowers, D., Smith, F. O., Shannon-Dorcy, K., Berger, M. S., and Bernstein, I. D. (1999) Selective ablation of acute myeloid leukemia using antibody-targeted chemotherapy: a phase I study of an anti-CD33 calicheamicin immunoconjugate, *Blood 93*, 3678-3684.
44. Brukner, I. (2000) C-1027 Taiho Pharmaceutical Co. Ltd., *Curr. Opin. Oncol. Endo. Meta. Inves. Drugs 2*, 344-353.
45. Beerman, T. A., Gawron, L. S., Shin, S., Shen, B., and McHugh, M. M. (2009) C-1027, a radiomimetic enediyne anticancer drug, preferentially targets hypoxic cells, *Cancer Res. 69*, 593-598.
46. Horsman, G. P., Van Lanen, S. G., and Shen, B. (2009) Iterative type I polyketide synthases for enediyne core biosynthesis, *Methods Enzymol. 459*, 97-112.
47. Zhang, J., Van Lanen, S. G., Ju, J., Liu, W., Dorrestein, P. C., Li, W., Kelleher, N. L., and Shen, B. (2008) A phosphopantetheinylating polyketide synthase producing a linear polyene to initiate enediyne antitumor antibiotic biosynthesis, *Proc. Natl. Acad. Sci. U.S.A. 105*, 1460-1465.
48. Luo, Y., Lin, S., Zhang, J., Cooke, H. A., Bruner, S. D., and Shen, B. (2008) Regiospecific O-methylation of naphthoic acids catalyzed by NcsB1, an O-methyltransferase involved in the biosynthesis of the enediyne antitumor antibiotic neocarzinostatin, *J. Biol. Chem. 283*, 14694-14702.
49. Cooke, H. A., Guenther, E. L., Luo, Y., Shen, B., and Bruner, S. D. (2009) Molecular Basis of Substrate Promiscuity for the SAM-Dependent O-Methyltransferase NcsB1, Involved in the Biosynthesis of the Enediyne Antitumor Antibiotic Neocarzinostatin, *Biochemistry 48*, 9590-9598.
50. Cooke, H. A., Zhang, J., Griffin, M. A., Nonaka, K., Van Lanen, S. G., Shen, B., and Bruner, S. D. (2007) Characterization of NcsB2 as a promiscuous naphthoic acid/coenzyme A ligase integral to the biosynthesis of the enediyne antitumor antibiotic neocarzinostatin, *J. Am. Chem. Soc. 129*, 7728-7729.
51. Christianson, C. V., Montavon, T. J., Festin, G. M., Cooke, H. A., Shen, B., and Bruner, S. D. (2007) The mechanism of MIO-based aminomutases in beta-amino acid biosynthesis, *J. Am. Chem. Soc. 129*, 15744-15745.

52. Cooke, H. A., Christianson, C. V., and Bruner, S. D. (2009) Structure and chemistry of 4-methylideneimidazole-5-one containing enzymes, *Curr. Opin. Chem. Biol. 13*, 453-461.

Chapter 2

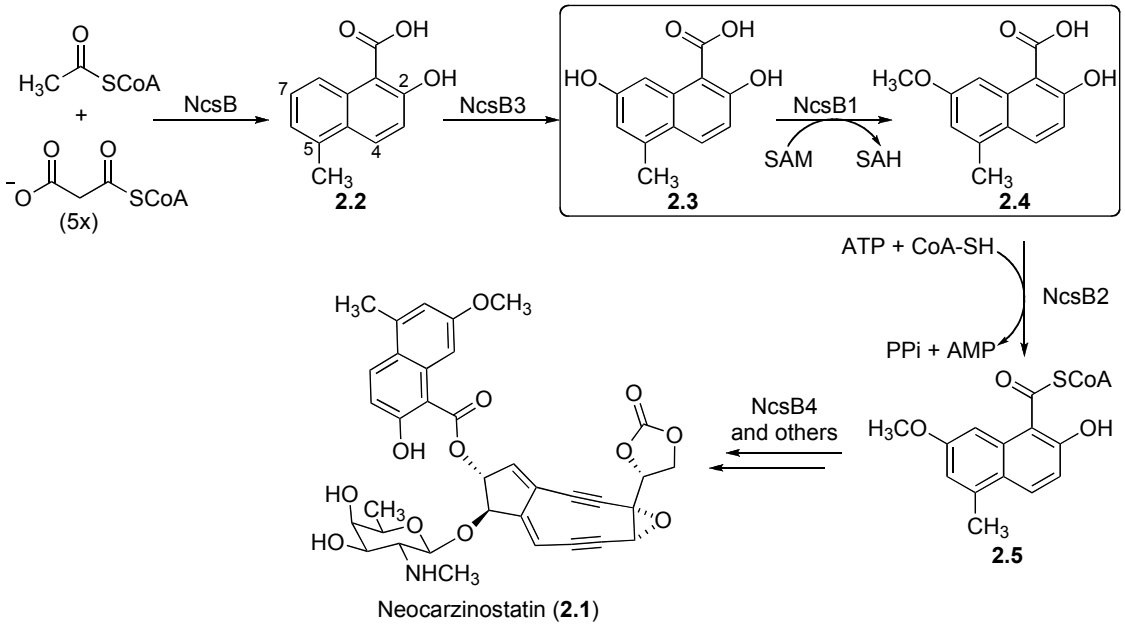
Molecular Basis of Substrate Promiscuity for the SAM-Dependent *O*-Methyltransferase NcsB1 Involved in the Biosynthesis of the Enediyne Antitumor Antibiotic Neocarzinostatin

Adapted from an article published in *Biochemistry* **2009**, 48, 9590-8 in collaboration with E.L. Guenther, Y. Luo, B. Shen, and S.D. Bruner

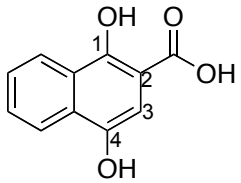
2.1 Introduction

The recent cloning and sequencing of the NCS biosynthetic pathway from *Streptomyces carzinostaticus* revealed a convergent assembly with two polyketide synthases playing central roles in building the carbon framework (*1-3*). The scaffold for the naphthoate moiety in NCS is constructed by an iterative type I polyketide synthase via a mechanism common to the biosynthesis of diverse aromatic products (*4-6*). Four additional gene products decorate and couple the naphthoic acid to the enediyne core (Scheme 2.1). The activation and attachment steps will be introduced here but will be further elaborated in Chapter 3.

Scheme 2.1: Biosynthesis of the naphthoic acid moiety and its attachment to the enediyne core with *O*-methyltransfer reaction catalyzed by NcsB1 boxed



The product of NcsB (**2.2**) is first hydroxylated at the C-7 position by a cytochrome P450 hydroxylase (NcsB3) to produce **2.3**. The resulting hydroxyl group is methylated by NcsB1, producing the methyl ether **2.4** present in the natural product. This functionalized naphthoic acid is activated via adenylation, converted to the aryl-CoA **2.5** by the CoA-ligase NcsB2, and is finally attached to the enediyne core by NcsB4 (a putative acyltransferase) to produce the natural product (**2.1**). The assignment of the naphthoic acid biosynthetic pathway is based on bioinformatic analysis (*1*) and biochemical characterization of NcsB2 as a CoA-ligase (*7*) and NcsB1 as an S-adenosylmethionine (SAM)-dependent *O*-methyltransferase (*O*-MTase) (*8*). NcsB1 was shown to catalyze the third step in the biosynthesis of the naphthoate moiety of the NCS chromophore (**2.3** to **2.4**). Biochemical analysis also revealed relaxed substrate flexibility towards substituted naphthoic acids, and methylated 3-hydroxy-2-naphthoic acids with the reactive hydroxyl group located at either the 5 or 7 position on the ring. NcsB1 also was demonstrated to methylate 1,4-dihydroxy-2-naphthoic acid (Figure 2.1, **2.6**) at the 4-hydroxyl group, which was surprising because of the large difference in the location of the reactive hydroxyl group from that on the natural substrate, necessitating a reorientation of the substrate in the active site. This observation led to the prediction of a model by which naphthoic acids flexibly bind the active site, and that the 1,2-hydroxy



2.6

Figure 2.1: 1,4-Dihydroxy-2-naphthoic acid, an alternate substrate

acid functionality was required for substrate binding and SAM-dependant methylation (*8*).

There are ~160 classes of SAM-dependent MTases comprised of thousands of identified members. These functionally diverse enzymes methylate a wide range of substrates, including nucleic acids for regulation of gene expression, DNA repair or protection against restriction enzymes, proteins for repair or control of signal transduction pathways, hormones and neurotransmitters, and biosynthetic intermediates to produce secondary metabolites. Structurally characterized MTases can be segregated into five general classes based on specific structural features (*9*). Among the several examples of MTases that act on small molecules, catechol *O*-MTase isolated from rat liver was the first to be structurally characterized (*10*). This structure exemplifies the basic SAM-MTase core fold: a mixed α, β secondary structure with a β-sheet flanked on each side by three α-helices (*11*). A common feature of the active site of MTases is a glycine-rich region within the SAM binding pocket. Despite low overall sequence homology amongst family members, structurally characterized SAM-MTases share this common fold and the biologically active unit typically exists as a homodimer.

The role of NcsB1 was assigned in the enediyne neocarzinostatin biosynthetic pathway based on sequence homology to small molecule SAM dependent *O*-MTases including DnrK, an enzyme involved in the biosynthesis of the aromatic polyketide antibiotic daunorubicin from *S. peucetius* (44% sequence identity) (*12*). The X-ray crystal structure of DnrK revealed a homodimer with each monomer exhibiting the conserved small molecule MTase fold (*13*). RdmB from rhodomycin biosynthesis in *S.*

purpurascens is also homologous to NcsB1, sharing 44% sequence identity and 58% similarity, but interestingly is a hydroxylase, not an *O*-MTase (*14, 15*). RdmB was used as a model for molecular replacement in solving the crystal structure of DnrK. In addition to DnrK, other bacterial small molecule *O*-MTases structures have recently been determined, including examples requiring a cation for catalysis. Cation-dependent *O*-MTases include CmcI from cephamycin biosynthesis (*16*), LiOMT from the pathogenic bacterium *Leptospira interrogans* (*17*), and BcOMT2 from *Bacillus cereus* (*18*). A limited number of cation-independent bacterial *O*-MTase structures have been structurally characterized and include DnrK and RebM from rebeccamycin biosynthesis (*19*).

We sought to determine the structural basis for NcsB1 activity in a variety of cocomplexes with SAM or S-adenosyl-L-homocysteine (SAH) with or without 5-methyl-2-hydroxy naphthoic acid (**2.2**), product (**2.4**), or 1,4-dihydroxy-2-naphthoic acid (**2.6**), an alternate substrate. The structures reveal that NcsB1 shares an overall architecture common to the large MTase family. The active site binding pocket is able to accommodate the natural substrate and structurally diverse analogs, allowing efficient methylation with distinct regiospecificity. The specificity determinants of the naphthoate binding pocket were probed using site-directed mutagenesis and alternate substrates. Based on the results, we identified residues that affect the substrate specificity of the enzyme and tested NcsB1 for *O*-methyltransferase activity with alternate substrates.

2.2 Crystallization and x-ray structure determination

The 34.5 kDa/332 residue NcsB1 was cocrystallized with various combinations of SAH, SAM, deshydroxy-substrate analog **2.2** and the methylated product **2.4**. The analog **2.2** was used in crystallization experiments due to the oxidative instability of the substrate (**2.3**) over extended time periods. Two crystal forms were generated depending on the bound small molecules. When cocrystallized with a substrate analog/product and SAH or SAM, the cocomplex crystallized in the $P6_5$ space group. In the absence of naphthoic acid, distinct crystals belonging to the $C222_1$ space group formed. The ternary complex of NcsB1/SAM/**2.2** was the initial structure solved using molecular replacement with a polyalanine dimer model of DnrK (44% identity) as the search model. This structure was then used as a search model for the alternate cocomplexes. Solution of the NscB1/SAH complex ($C222_1$ space group) by molecular replacement was not straightforward, suggesting an alternate overall structure. The final structural model was obtained by performing rotation and translation searches separately on the N- and C-terminal domains. While a 0.6 Å $RMSD_{C\alpha}$ was observed amongst the $P6_5$ structures, the NcsB1/SAH structure differed significantly from the others by an $RMSD_{C\alpha}$ of 1.8 Å. In addition, the two subunits of the NcsB1/SAH structure were quite distinct, with an $RMSD_{C\alpha}$ of 2.8 Å between the homodimer chains.

2.3 Overview of the structures

The resulting dimer models of NcsB1 display a high degree of structural similarity with DnrK ($RMSD_{C\alpha}$ = 1.8 Å). In the initial NcsB1/SAM/**2.2** structure, and others in the $P6_5$ space group, the two monomeric subunits exhibit a high degree of

symmetry. Each subunit of the NcsB1 dimer is made up of three domains totaling nineteen α-helices and eight β-sheets (Figure 2.2). The N-terminal domain is largely α-helical, with just two β-strands (β1 and β2), and constitutes most of the residues involved in the homodimer interface (~5300 $Å^2$ total surface). Of note, α5 (residues 52-65) has significant interactions with α18 (residues 284-292) from the C-terminal domain of the paired dimer. The dimer interface is made up of primarily hydrophobic residues, and a dimeric biological unit was supported by size-exclusion chromatography (Figure 2.3). A middle domain acts as a hinge-like region between the two larger terminal domains and is made up of α-helices 9-11. The C-terminal domain exhibits a Rossman-like fold (β3-9 flanked by α12-19), which is conserved in structural homologs and makes up the majority of the SAM binding site (*19*). Based on structural alignment using the DALI server, NcsB1 shares the highest structural similarity with SAM-dependent enzymes that act on small molecules (Table 2.1, Figure 2.4) (*20*). Homologs include DnrK (an *O*-MTase) and RdmB (a hydroxylase), isoflavone-O-MTase from alfalfa (*21*), a putative *O*-MTase from *Nostoc punctiforme*, a phenazine-specific N-MTase from *Pseudomonas aeruginosa* (*22*) and caffeic acid 3-*O*-MTase from alfalfa (*23*). Among the structures, an indicative glycine-rich region of the SAM binding pocket is conserved. These six structural homologs demonstrate the ability of the conserved scaffold to catalyze a variety of transformations.

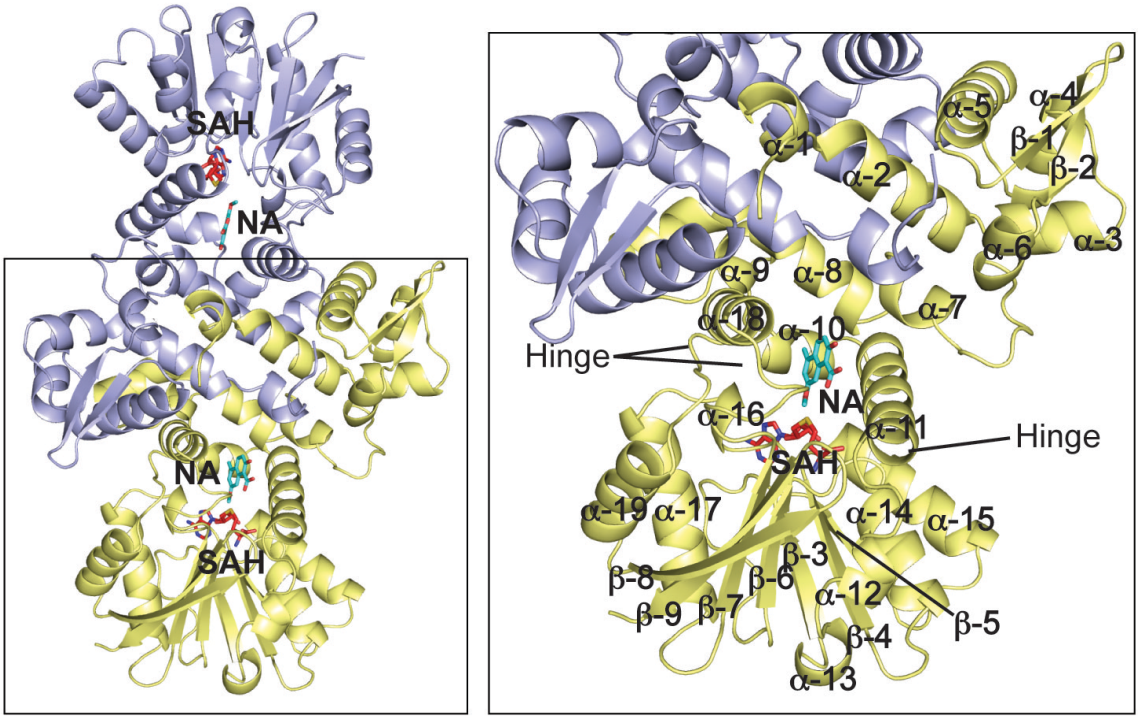


Figure 2.2: Cartoon representations of NcsB1. (A) Dimer of NcsB1 with the active site region indicated. Chain A is colored light blue and chain B yellow. Ligands are depicted in stick format with SAH in red and naphthoic acid (NA, **4**) in cyan. B) Close up view of monomer with secondary structural elements and hinge regions labeled.

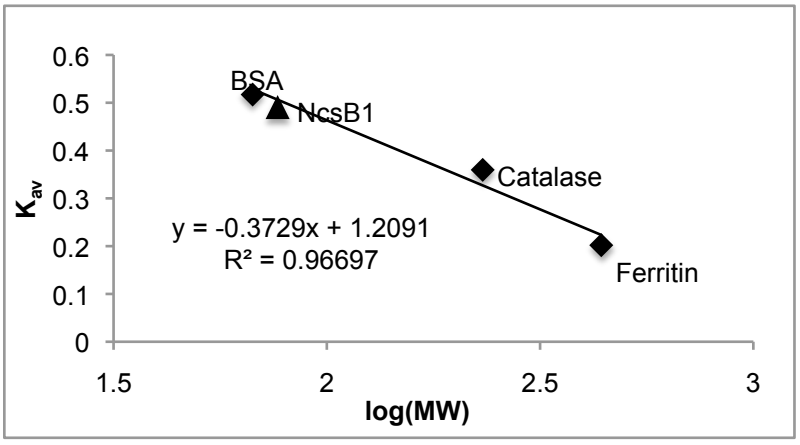


Figure 2.3: Size-exclusion chromatography calibration for estimating biological active unit of NcsB1 (▲). K_{av} for NcsB1 = 0.491; Calculated size = 84.3 kDa, ~2 monomers. Actual size = 69.0 kDA.

Table 2.1: Dali Structural Homolog Search (*20*)

Position[a]	PDB[b]	Z[c]	$RMSD_{Ca}$ (Å)	# eq. res.	# res.	%ID of eq. res.	Description
1 (*13*)	1TW3	39.2	1.8	324	340	44	DnrK, *Streptomyces peucetius*
5 (*14*)	1XDS	37.5	1.7	316	336	45	RdmB, *Streptomyces purpurascens*
10 (*21*)	1FPX	29.6	3.8	309	345	23	IOMT, alfalfa
12	2R3S	29.1	3.6	314	330	18	NPOMT, *Nostoc punctiforme*
13 (*22*)	2IP2	28.7	4.8	317	330	26	PhzM, *Pseudomonas aeruginosa*
16 (*23*)	1KYW	27.0	4.3	313	361	25	COMT, alfalfa

[a] Position in listing of structural homologs. Gaps indicate additional structures(s) of identical proteins. Reference in parentheses.

[b] Protein Data Bank code

[c] Strength of structural similarity

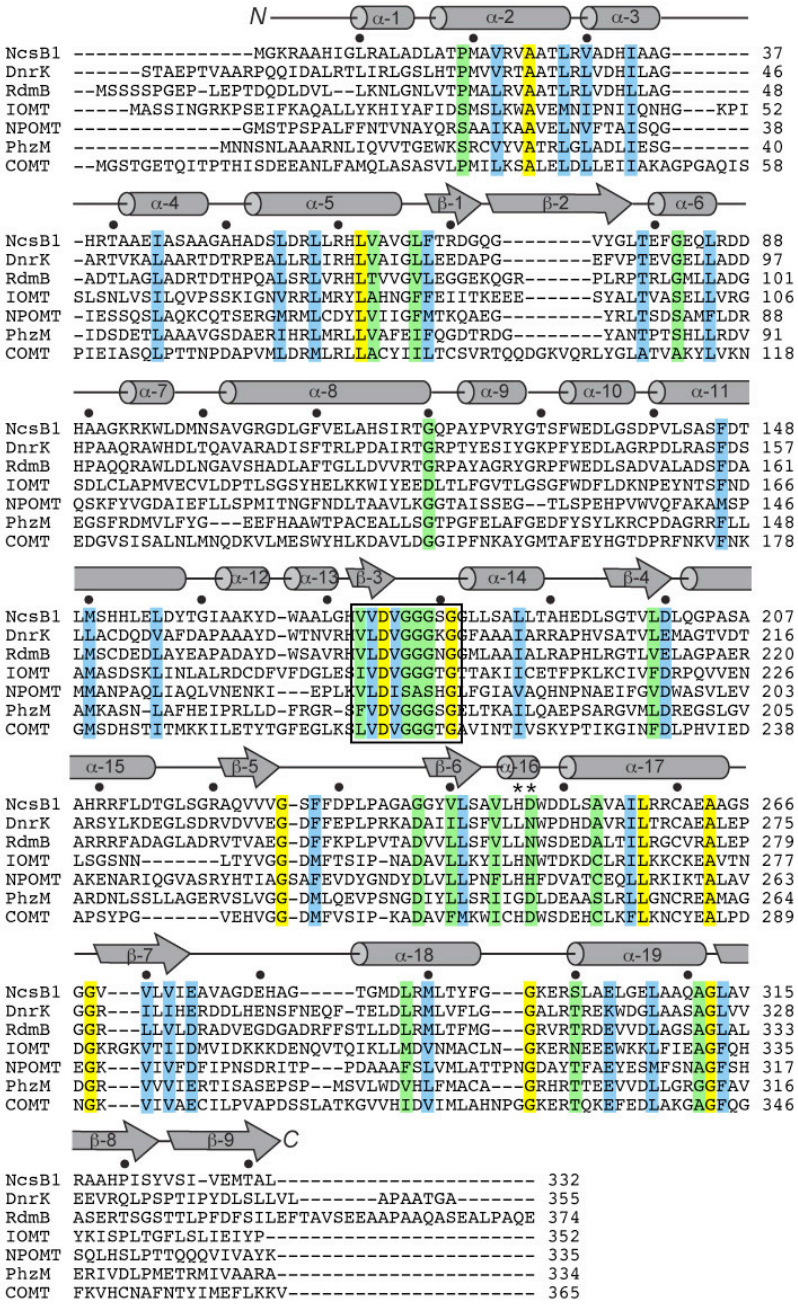


Figure 2.4: Sequence alignment of NcsB1 structural homologs as determined by DALI structural alignment server. Abbreviations used with accession codes from the RCSB Protein Data Bank in parentheses: DnrK, *S. peucetius* O-MTase (Q06528); RdmB, *S. purpurascens* hydroxylase (Q54527); IOMT, alfalfa O-MTase (O24529); NPOMT, *N. punctiforme* putative O-MTase (ZP_00112478); PhzM, *P. aeruginosa* N-MTase (Q9HWH2); COMT, alfalfa O-MTase (P28002). Completely conserved residues are yellow, highly conserved residues are blue, and conserved residues are green. Catalytic residues are indicated by an asterisk and conserved glycine region is boxed.

In the NcsB1/SAH cocomplex structure, the C-terminal domain of chain B is rotated by ~20° compared to NcsB1/SAH/**2.2**, displacing α13 by 6.8 Å and β8/β9 by 9.4 Å (Figure 2.5). This movement results in a more open conformation of the active site, possibly functioning to allow the entrance of substrate. A similar domain displacement was seen in the DnrK ternary complex where two crystals with different space groups were also observed (*13*). Both DnrK crystals were bound to SAH and product, but the B-factors were higher in one of the two subunits, suggesting lower occupancy of the ligands. One of the crystal structures exhibited a significant difference between the two monomers, with an $RMSD_{C\alpha}$ of 1.3 Å. Likewise, the RdmB structures exhibit a domain movement upon substrate binding, though in this cocomplex structure, both domains were displaced (*14*). The NcsB1/SAH cocomplex described here exhibits a larger difference, likely due to complete lack of substrate analog or product in the crystallization conditions. In addition, SAH bound in the open subunit (chain B) has higher relative B-factors. Several regions in chain B showed weak or no electron density, possibly indicating areas of flexibility on which the domain can twist. These hinge regions

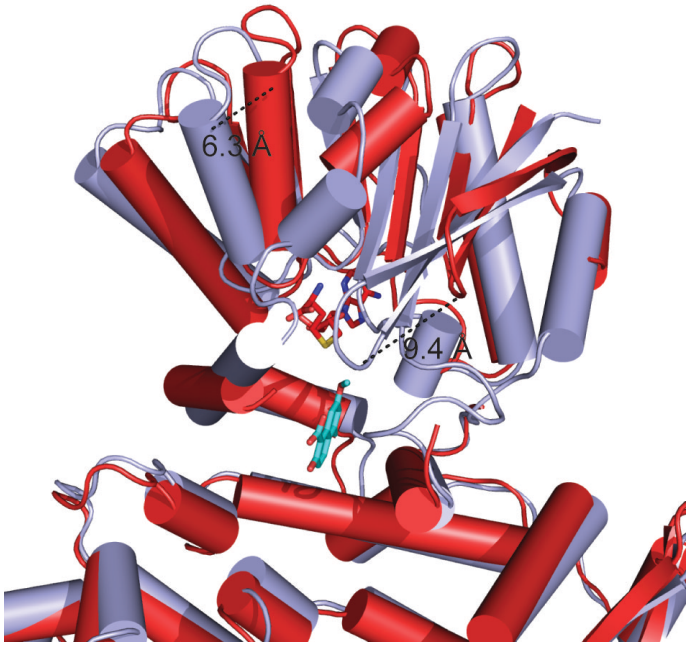


Figure 2.5: Overlay of NcsB1/SAH/**4** monomer (light blue) and NcsB1/SAH monomer (red).

include part of α11 (residues 150-159), the loop region between β7 and α18 (residues 274-282) and the loop between α18 and α19 (residues 297-301) (Figure 2.2).

2.4 SAM/SAH binding pocket

The SAM/SAH binding pocket is located at the C-terminus of the β-strands comprising part of the Rossman-like fold. Electron density for SAM or SAH is clearly defined in all four cocomplex structures and the SAM binding interactions are largely conserved among small molecule MTases. The adenine ring is involved in a hydrogen-bonding interaction with Ser227, a π interaction with Phe228 and the binding pocket is lined with additional aromatic and/or hydrophobic residues, including Trp133, Trp248, Phe229 and Leu201 (Figure 2.6). The ribosyl moiety is anchored by two hydrogen bonds between Asp200 and Ser143 and the homocysteine portion of SAM forms hydrogen bonding interactions with the sidechains of Asp175, Ser242, His153 and the backbone carbonyl of Gly177. The LDXGXGXG motif indicative of SAM-utilizing proteins is found in NcsB1 as VDVGGGSG and is located between β3 and α14. We previously reported that NcsB1 alkylates the substrate **2.3** with a variety of SAM analogs, including S-ethyl and S-n-propyl, to the corresponding 7-alkyl ether with reasonable efficiency (*8*). The nearest side chain to the methyl-group on SAM is the sterically small Ala243 (3.5 Å distance). Based on the large size and flexibility of the substrate binding pocket, it is not surprising that NcsB1 could accommodate larger SAM analogs.

2.5 Naphthoic acid binding pocket

The naphthoic acid binding site is located at the juncture of helices from all three subdomains, including α7, α10, α11, α16, and α18, and the pocket is lined primarily with hydrophobic/aromatic residues (Figure 2.6). Three methionine sidechain thioethers (Met150, 286 and 290) are present on either side of the naphthoate ring forming van der Waals interactions with the substrate. The carboxylate and 2-hydroxyl group of the substrate interacts with Arg11 directly and with Asp157 through an ordered water. This interaction represents the only hydrogen bonds between the enzyme and substrate and appears to anchor the naphthoic acid into the active site adjacent to the bound SAM. In order to probe the importance of Arg11 in enzyme activity, site-directed mutagenesis was used to alter this site and the activity was evaluated. This residue was mutated to Ala and Lys to assess the requirement of this hydrogen-bonding interaction for binding naphthoic

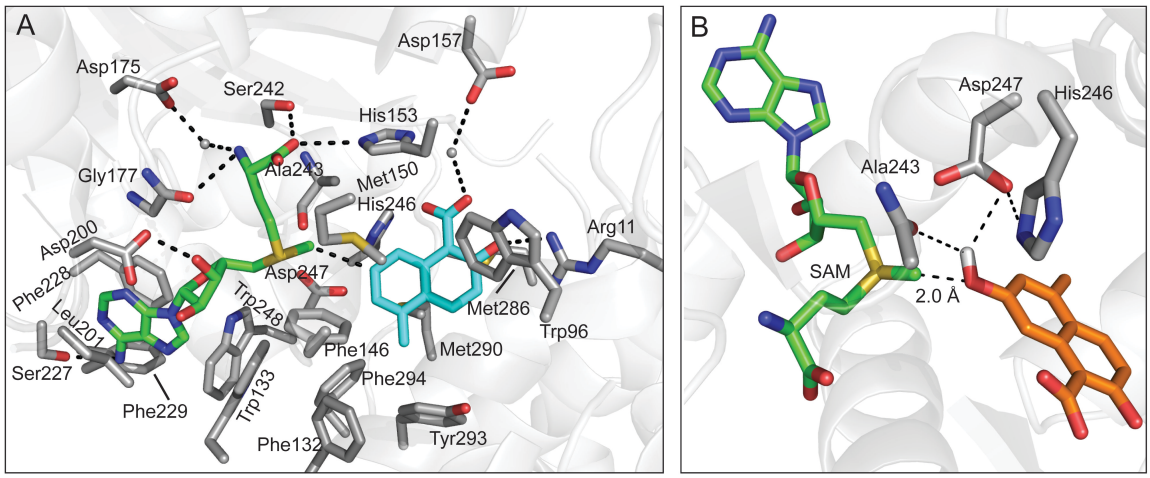


Figure 2.6: Active site of NcsB1/SAM/**2.2**. (A) Substrate and SAM binding pocket. SAH is shown in green and naphthoic acid **2.2** in cyan; residues in the substrate and SAM binding pockets are shown in grey with pertinent hydrogen-bonds indicated. (B) Close up of active site in NcsB1/SAM/**2** structure with 7-hydroxy group modeled onto **2.2**, with the likely catalytic residues shown in grey. SAM is shown in green and the naphthoic acid **2.3** in orange.

acids. Unexpectedly, the Arg11Ala mutant still effectively methylated naphthoic acid **2.3** (Table 2.2), although the results showed a doubling of K_M that was countered by a similar increase in k_{cat}. The Arg11Lys mutant had 110% of the catalytic efficiency as compared to the WT. This mutant likewise showed an increased rate of turnover. This observation suggests that the specific hydrogen bond from Arg11 is not entirely crucial for substrate turnover and can be compensated for by a similar interaction with Lys or by increased overall catalytic turnover. In addition, a simple benzoic acid lacking a hydroxyl group adjacent to the aryl acid (3-hydroxybenzoic acid, **2.9**, Table 2.5) was methylated by NcsB1, although at a significantly decreased rate, further suggesting that the 1,2-hydroxy acid motif is not absolutely required for enzyme-catalyzed chemistry.

Table 2.2: Kinetics of NcsB1 and mutant constructs

Protein/substrate	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M (rel to WT/NA)
WT/**2.3** (***8***)	206 ± 49	0.69 ± 0.05	1
Tyr293Ile/**2.3**	649 ± 9	0.76 ± 0.04	0.35
Arg11Trp/**2.3**	400 ± 68	0.60 ± 0.03	0.45
Arg11Ala/**2.3**	419 ± 61	1.24 ± 0.06	0.88
Arg11Lys/**2.3**	319 ± 40	1.18 ± 0.05	1.10

2.6 NcsB1 residues involved in methyltransferase chemistry

Roles for residues in the active site relevant for catalysis could include activation of the phenol for nucleophilic attack via acid/base chemistry. Histidine residues specifically have been implicated to play this role in RebM, a carbohydrate *O*-MTase in rebeccamycin biosynthesis, where mutation of two histidine residues in the substrate binding pocket to alanines led to a complete loss of activity, while single mutations showed a marked decrease in activity (*19*). In NcsB1, three residues are in close

proximity to the substrate phenol; the diad of His246 and Asp247 side chains and the backbone carbonyl of Ala243 may function to aid in activation and/or proton shuttling (Figure 2.6B).

2.7 Binding of 1,4-dihydroxy-2-naphthoic acid in active site of NcsB1

Previous exploration of substrate specificity of NcsB1 suggested a large degree of flexibility for naphthoate binding (*8*). A model was proposed that involved anchoring of the substrate by the 1,2-hydroxy acid motif and subsequent methylation of exposed phenols in proximity to bound SAM (Figure 2.7). In order to probe this hypothesis, the structure of NcsB1 bound to 1,4-dihydroxy-2-naphthoic acid (DHN, **2.6**) was determined. Methylation of this substrate would necessitate binding in a conformation distinct from the natural substrate while maintaining the interaction with the enzyme through the 1,2-hydroxy acid. In the cocrystal structure, the proximity and orientation of **2.6** to SAH and important hydrogen bonding partners were readily apparent, despite low resolution (Figure 2.8A). The substrate is bound in an appropriate position for regioselective methylation by SAM with an alternate orientation of the bound naphthoic acid. The methyl group on SAM is ~2.1

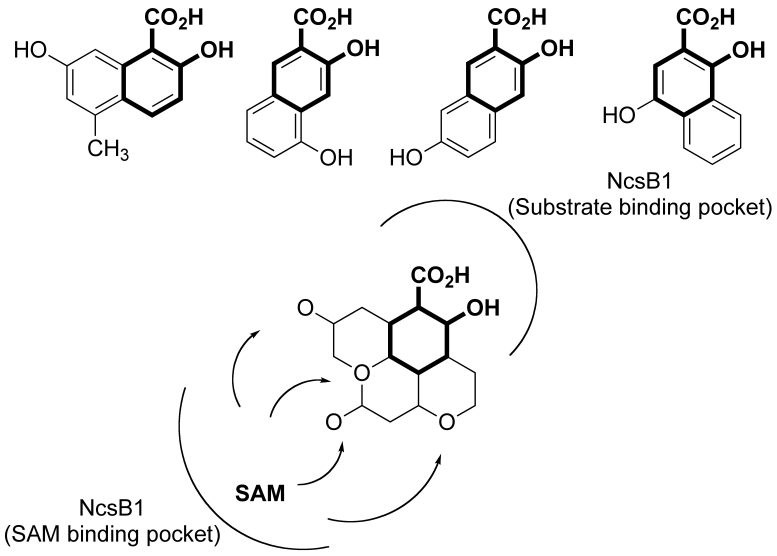


Figure 2.7: Substrate binding scheme for NcsB1

Å away from the 5-hydroxyl group of **2.6** and the carboxylate is hydrogen bonded to Arg11 via an ordered water (Figure 2.8B).

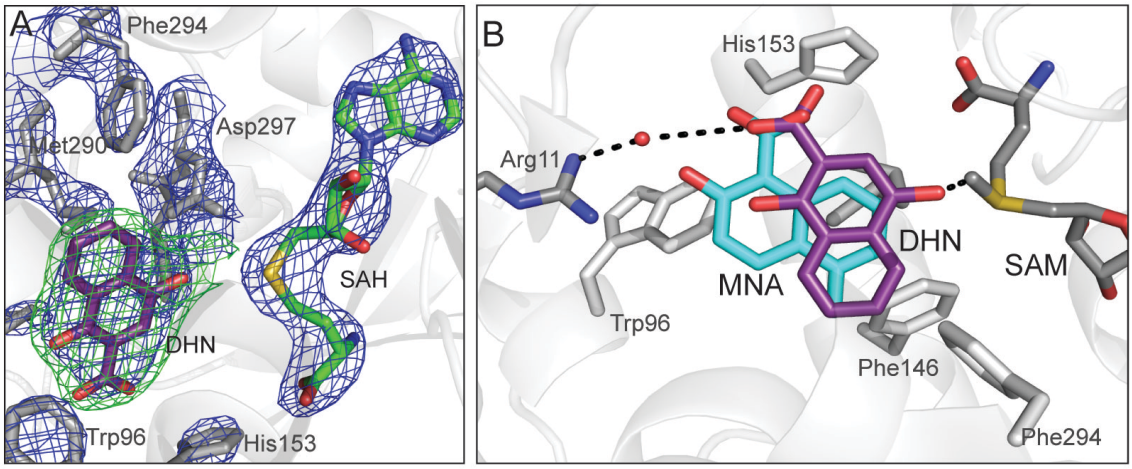
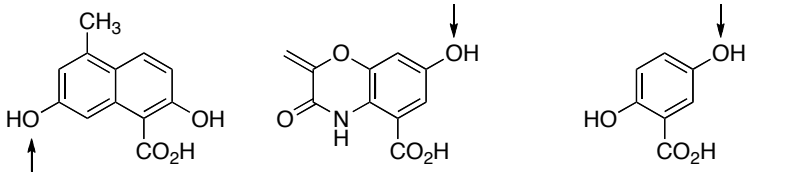


Figure 2.8: Binding of alternate substrate, DHN (**6**). (A) Electron density map of NcsB1/SAH/**2.6** active site, generated without **2.6**. $2F_o - F_c$ map (blue, 1.5 σ) and $F_o - F_c$ map (green, 3 σ) (B) Overlay of **2.2** and **2.6** in NcsB1 active site. Naphthoic acid **2.6** is hydrogen-bonded to Arg11 via a water and the 4-hydroxyl group is in sufficiently close proximity to SAM (from NcsB1/SAM/**2.2** structure) for methyl transfer to occur.

2.8 Probing the determinants of the substrate specificity for NcsB1

The biosynthetic gene cluster for the related enediyne C-1027 contains a gene (*sgcD4*) with high homology to ncsB1 (56% identity, 67% similarity), and the enzyme is proposed to methylate a benzoxazolinate (**2.7**, Figure 2.9) (*24*, *25*). Alignment of their



NcsB1 substrate naphthoic acid (**2.3**)

SgcD4 substrate benzoxazolinate (**2.7**)

2,5-dihydroxybenzoic acid (**2.8**)

Figure 2.9: NcsB1 and SgcD4 substrates and model substrate. Arrow indicates point of methylation.

corresponding amino acid sequences shows many residues pertinent to substrate and cofactor binding in common (Figure 2.10). Of nine residues in NcsB1 identified to play a role in substrate binding, four have identical counterparts in SgcD4 (Trp96, Phe146, Met286, and Met290), two are similar ($Met150_{NcsB1}$ is a Leu in SgcD4 and $Phe294_{NcsB1}$ is a Tyr) and two residues are distinct ($Arg11_{NcsB1}$ is a Trp and $Tyr293_{NcsB1}$ is an Ile).

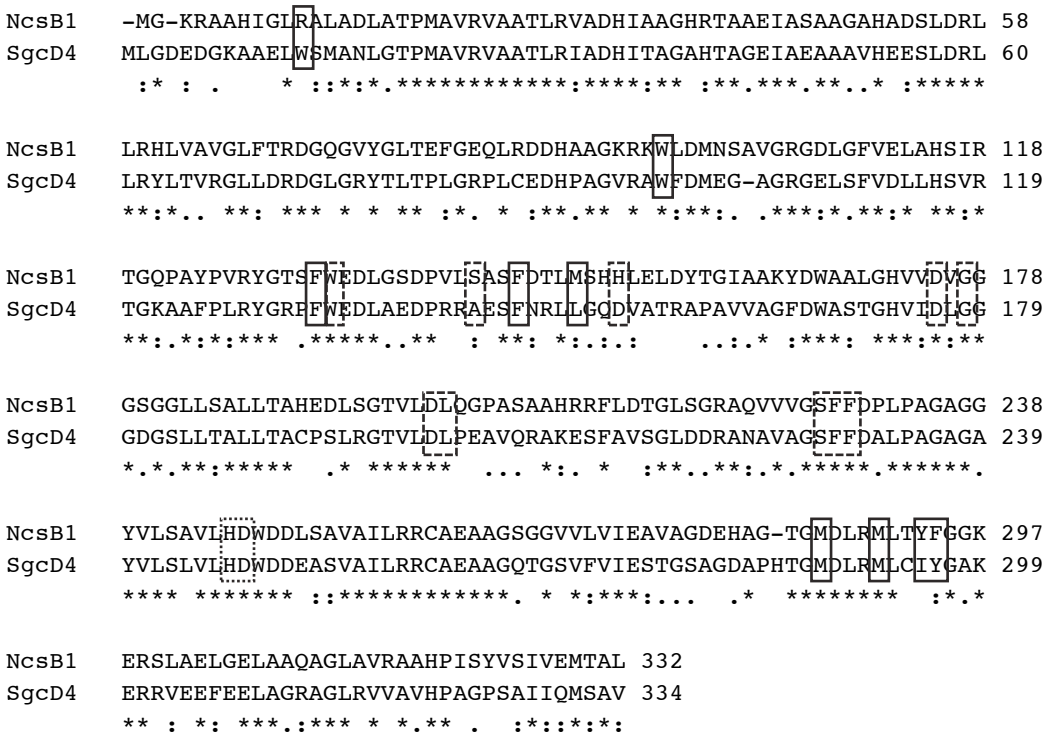


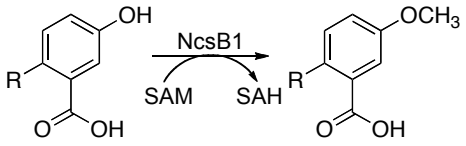
Figure 2.10: NcsB1/SgcD4 sequence alignment with key residues boxed. Naphthoic acid binding pocket, solid line; SAM binding pocket, dashed line; Catalytic residues, dotted line.

In order to probe the substrate specificity determinants of NcsB1, two rational mutations were made on the residues in the active site that differed significantly between NcsB1 and SgcD4: Tyr293Ile and Arg11Trp. These two additional mutant constructs were assayed against the natural substrate (**2.3**) (Table 2.2). Both mutants methylated **2.3**, though they showed significant increases in K_M compared to the wild-type enzyme.

A larger decrease in catalytic efficiency was seen with the Tyr293Ile mutant, which had a K_M approximately three times higher than wild-type. The K_M was doubled for the Arg11Trp. The ability of Arg11 mutants, especially the Trp and Ala mutants, to methylate naphthoic acid suggests that the aromatic binding pocket of NcsB1 is substantially more crucial to binding substrate than the hydrogen bond from Arg11. The reduced aromatic character of the Tyr293Ile mutant supports this theory.

We next tested a simple substrate analog of the benzoxazolinate, 2,5-dihydroxybenzoic acid (**2.8**). In this model compound, methylation occurs at the 5-position (**2.10**), although with significantly lower efficiency than with the natural substrate. **2.8** was then assayed against the panel of mutants in Table 2.2, and indeed alternate specificity was observed as compared to wild-type NcsB1 (Table 2.3). For example, although the relative measured activities were low, contrary to the wild-type enzyme, the Arg11Trp mutant has higher relative activity toward 2,5-dihydroxybenzoic acid while the Arg11Lys mutant was less efficient. These results implicate Arg11 in substrate discrimination for these related methyltransferases.

Table 2.3: Kinetics of benzoic acids with NcsB1 and mutant constructs



2.8, R=OH
2.9, R=H

2.10, R=OH
2.11, R=H

Protein/substrate	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m (rel to WT/NA)	k_{cat}/K_m (rel to WT/BA)
WT/**2.8**	280 ± 30	0.0043 ± 0.0001	0.005	1
WT/**2.9**	137 ± 22	0.00064 ± 0.00002	0.001	0.3
Tyr293Ile/**2.8**	558 ± 113	0.0033 ± 0.0003	0.002	0.4
Arg11Trp/**2.8**	183 ± 27	0.0071 ± 0.0003	0.012	2.6
Arg11Ala/**2.8**	216 ± 17	0.0022 ± 0.0005	0.003	0.7
Arg11Lys/**2.8**	348 ± 41	0.0040 ± 0.0002	0.003	0.7

NcsB1 was also demonstrated to methylate 2,4,5-trihydroxybenzoic acid. The mass of both a single methylated and a double methylated product was found via LC-MS. The ability of NcsB1 to not only methylate **2.8** and **2.9**, but also 2,4,5-trihydroxybenzoic acid shows that it is a relatively flexible *O*-methyltransferase and has the ability to transfer more than one methyl group onto a substrate.

2.9 Conclusion

In summary, we have solved the crystal structures of NcsB1 in two conformations with SAH/SAM bound, with and without a substrate analog or product. These structures revealed a large displacement of the C-terminal domain when not bound to substrate, a movement that likely opens up the active site for naphthoate binding. Additionally, the ternary complex structure of 1,4-dihydroxy naphthoic acid and SAH bound to NcsB1 was solved and showed a rotation of this alternate substrate in the binding pocket, allowing

for methylation of the hydroxyl group at the 4 position. These results led us to probe substrate binding using active site mutants, demonstrating altered substrate specificity and revealing the importance of key residues in substrate binding.

2.10 Experimentals

Protein expression and purification

NcsB1 was overproduced as an N-terminal His_6-tagged fusion protein using the expression plasmid pBS5039 in *E. coli* BL21(DE3) cells as reported previously (*8*). Cells were grown in 1 L Luria-Bertani media at 37 ºC, 150 RPM to an OD_{600} = 0.5-0.8. Overexpression was induced with 50 μM isopropyl β-D-thiogalactopyranoside at 18 ºC for 16 hr. Cells were collected by centrifugation (20 min, 2000 g), resuspended in 25 mL of 20 mM Tris-HCl pH 7.5 and 500 mM NaCl, and flash frozen at -78 ºC. Cell pellets were thawed and lysed by two passes through a French Press cell disruptor at 1000 psi. The lysate was then clarified by centrifugation (20 min, 10,000 g) and the protein was purified batchwise using Ni-NTA resin (QIAGEN, Valencia, CA). NcsB1 used in biochemical assays was dialyzed into 20 mM Tris-HCl pH 7.5, 50 mM NaCl, and 1 mM β-mercaptoethanol, then concentrated using an Amicon Ultra-4 concentrator (10 kDa MW cutoff, GE Healthcare) and frozen with 40% supplemented glycerol at -25 ºC. For crystallography, the Ni-NTA purified protein was dialyzed into 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM $CaCl_2$, 1 mM β-mercaptoethanol) and concentrated to ~ 1 mL. The His_6-tag was cleaved by incubation with the protease Factor Xa for 36 h at 4 ºC (monitored by SDS-PAGE). The protein solution was then diluted to 5 mL and purified

on a HiTrap-Q ion exchange column (0-1 M NaCl gradient in 50 mM Tris-HCl pH 7.5 and 1 mM β-mercaptoethanol) followed by a Superdex 200 gel filtration column (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM β-mercaptoethanol, 10% glycerol). The purified protein was concentrated to ~10 mg/mL for crystallography (a total yield of 4 mg/L of cells). Protein concentration was determined using UV absorption at 280 nm and the Bradford assay (*26*).

Site-directed mutagenesis of ncsB1

The Quikchange II Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA) was used to generate mutant constructs of NcsB1 using the following DNA oligonucleotide primers (IDT DNA, Coralville, IA) (mutant codon is underlined): Arg11Ala For: 5' - GGC TGC ACA CAT CGG ATT GGC GGC GCT GGC CGA TCT GGC GAC – 3' and Rev: 5' - GTC GCC AGA TCG GCC AGC GCC GCC AAT CCG ATG TGT GCA GCC – 3'; Arg11Trp For: 5' - GGC TGC ACA CAT CGG ATT GTG GGC GCT GGC CGA TCT GGC GAC – 3' and Rev: 5' - GTC GCC AGA TCG GCC AGC GCC CAC AAT CCG ATG TGT GCA GCC – 3'; Arg11Lys For: 5' - GGC TGC ACA CAT CGG ATT GAA GGC GCT GGC CGA TCT GGC GAC – 3' and Rev: 5' - GTC GCC AGA TCG GCC AGC GCC TTC AAT CCG ATG TGT GCA GCC – 3'; Tyr293Ile For: 5' - GCG CAT GCT CAC CAT CTT CGG AGG CAA GGA ACG C – 3' and Rev: 5' - GCG TTC CTT GCC TCC GAA GAT GGT GAG CAT GCG C – 3'. The Quikchange protocol was followed as described with the addition of DMSO to the reaction to a final concentration of 5%. Mutations were confirmed by sequencing (Genewiz, S. Plainfield, NJ) and

mutant constructs were transformed into E. coli BL21(DE3) cells, over-expressed and purified as described above.

Crystallization and X-ray data collection

Purified NcsB1 was crystallized with a substituted naphthoic acid and/or SAM/SAH by the hanging drop vapor diffusion method at 20 ºC with ~4 M sodium formate (7 M stock solution from Hampton Research, see Table 2.4 for crystallization conditions).

Table 2.4: Crystallization Conditions for NcsB1

	SAH	**SAM and 2.2**	**SAH and 2.4**	**SAH and 2.6**
NcsB1 (mg/mL)	2	2	2	2
cofactor (mM)	2	2	2	2
NA (mM)	0	2	2	1
βME (mM)	1	1	1	10
Sodium Formate (M)	3.70	3.60	3.66	3.90

Stock solutions of the small molecules for cocrystallization were in 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM β-mercaptoethanol, 10% glycerol. Naphthoic acids were synthesized as described previously (*7, 27*) and in chapter 3 or purchased from Aldrich. Typically, NcsB1 crystals appeared within a day and reached full size within one week. Mature crystals were transferred to a cryoprotectant solution (4.0 M sodium formate, 15% glycerol) and soaked briefly before flash freezing in liquid nitrogen. X-ray diffraction data was collected on the X25 beamline at the National Synchrotron Light Source at Brookhaven National Laboratories with an ADSC Q315 CCD X-ray detector. Diffraction intensities were indexed, integrated and scaled using HKL2000 (*28*) as

summarized in Table 2.5. The crystals belonged to the space group $P6_5$ except for the NcsB1-SAH cocomplex, which belonged to the $C222_1$ space group.

Table 2.5: Data collection and refinement statistics

	SAH	**SAM and 2.2**	**SAH and 2.4**	**SAH and 2.6**
Space group	$C222_1$	$P6_5$	$P6_5$	$P6_5$
Unit cell parameters (Å)	a=91.3,b=161.6, c=98.9	a=b=109.1, c=206.9	a=b=108.4, c=210.3	a=b=108.0, c=211.9
		Data collection		
Resolution (Å)	25-2.08 (2.15-2.08)	50-2.6 (2.69-2.60)	30-2.69 (2.79-2.69)	50.0-3.0 (3.1-3.0)
Wavelength (Å)	1	1	1	1
No. of reflections (measured/unique)	345090/44038	967889/42568	658394/38115	35878/27798
% Completeness*	99.6 (99.6)	99.8 (100.0)	98.5 (100.0)	99.8 (100.0)
R_{merge}*	0.052 (0.349)	0.077 (0.441)	0.094 (0.49)	0.076 (0.523)
I/σ*	32.2 (4.4)	49.6 (6.4)	24.0 (7.0)	29.3 (4.5)
Redundancy*	7.8 (7.4)	22.7 (22.9)	17.3 (17.8)	12.6 (12.8)
		Refinement		
No. of reflections (total/test)	41442/2071	41357/2064	37165/3704	26923/2698
R_{work}	0.240	0.206	0.212	0.217
R_{test}	0.272	0.227	0.247	0.245
No. amino acids (chains A/B)	325/303	325/325	328/328	328/328
No. atoms				
Protein (chains A/B)	2379/2193	2379/2379	2406/2406	2406/2406
Ligands (cofactor/substrate/ glycerol)	52/-/6	54/30/12	52/34/12	50/30/12
Water Molecules	305	249	163	95
Mean B (protein/SAM/ NA/water)	46.0/36.0/ -/50.1	53.0/42.3/ 51.6/55.6	71.4/59.6/ 81.4/74.3	69.9/59.9/ 57.2/65.4
RMSD from ideal geometry				
RMSD bonds (Å)	0.006	0.006	0.006	0.0013
RMSD angles (º)	1.2	1.3	1.2	1.7
Ramachandran plot by PROCHECK (%)				
Core region	90.6	93.2	91.7	88.8
Allowed region	9.0	6.2	7.9	10.4
Generously allowed	0.2	0.5	0.2	0.7
Disallowed	0.2	0.0	0.2	0.0
PDB code	3I53	3I5U	3I58	3I64

*overall (highest resolution shell)

Structure determination

An initial solution for the NcsB1/SAM/**2.2** ternary complex was obtained using the molecular replacement program PHASER (*29*), part of the CCP4 suite (*30*). A polyalanine dimer model of DnrK from Streptomyces peucetius (PDB ID code 1TW2) truncated to residues 15-350 was used as the search model. Manual building of the NcsB1 model was performed using the program COOT (*31*) and the structure was refined using CNS (*32, 33*). This solution was used to solve phases in subsequent datasets in the $P6_5$ space group using molecular replacement. For the NcsB1-SAH cocomplex ($C222_1$ space group), a partial solution (chain A and the N-terminus of chain B) was found using PHASER. The remaining C-terminus of chain B was placed into the model using the program MOLREP (*34*). All final structures were subjected to multiple rounds of building and refinement until the R-values converged. Non-crystallographic restraints were used for all $P6_5$ space group structures until final stages of refinement. Ligands were fit into the electron density maps and the PRODRG server (*35*) was used for generating topology and parameter files; SAM and SAH coordinates were obtained from the HICup server (*36*) and were subsequently altered to fit electron density. The program PyMOL (Delano Scientific, San Carlos, CA) was used to generate graphic images. RMSD values were calculated via structural comparison using TopMatch (*37, 38*).

The NcsB1/SAH structure ($C222_1$ space group) was solved at a resolution of 2.08 Å and refined to an R_{work} and R_{free} of 24.0 and 27.2% (Table 2.5). The asymmetric unit consisted of two monomers. For chain A, there was sufficient electron density to build in residues 5-330 and the bound SAH molecule. For chain B, the final model consisted of

residues 9-150, 159-274, 282-297 and 302-330, as insufficient electron density was available to build the remaining 29 residues. Electron density for the SAH molecule was well defined, albeit with higher B factors for the SAH molecule in chain B (Figure 2.10A). In addition, the dimer contained one glycerol molecule. The NcsB1 cocomplex with SAM and the substrate analog (**2.2**) was solved at a resolution of 2.60 Å with R_{work} and R_{free} of 20.6 and 22.7%, respectively. For both chains, N-terminal residues 1-4 could not be built, nor could residues 279-281, though sufficient electron density was available for both SAM and **2.2** to be built into each active site (Figure 2.10B). NcsB1 with SAH and product (**2.4**) diffracted to 2.69 Å and was refined to an R_{work} and R_{free} of 21.2 and 24.7%, respectively. All residues of both chains could be built in as well as SAH, the product in the active site, and four glycerol molecules (Figure 2.10C).

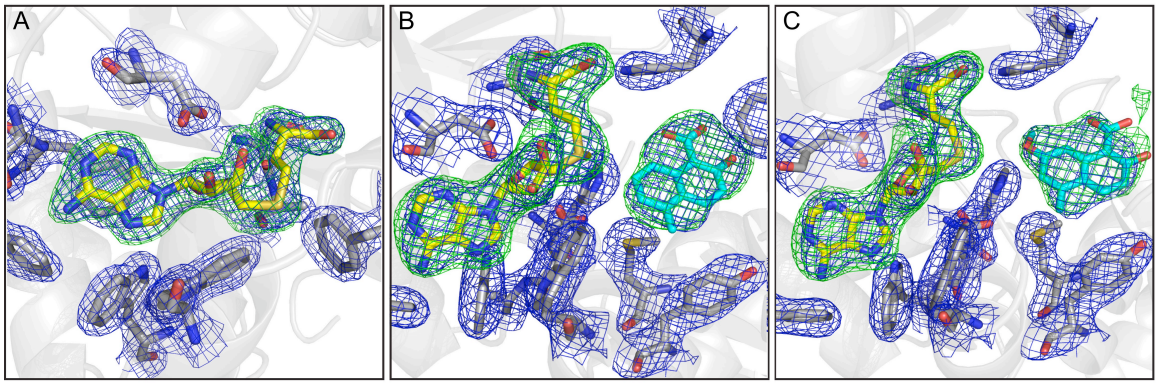


Figure 2.11: $2F_o - F_c$ maps contoured to 1.5 σ and $F_o - F_c$ map contoured to 3.0 σ. Maps generated without ligands modeled into active site (A) NcsB1/SAH. (B) NcsB1/SAM/**2.2**. (C) NcsB1/SAH/**2.4**.

The final structure was a cocomplex of NcsB1 with SAH and an alternate substrate, 1,4-dihydroxy-naphthoic acid (**2.6**). The crystal diffracted to 3.0 Å, likely a result of reduced crystal quality, as 10 mM β-mercaptoethanol was required for crystal growth because of problematic oxidation of the substrate. The final structure refined to an R_{work} and R_{free} of 21.7 and 24.5% and included residues 4-332, SAH and **2.6** in the final model.

Size-exclusion chromatography calibration

Size-exclusion chromatography was used to determine the biologically active entity of NcsB1 using BSA, catalase, and ferritin as standards. A standard solution was made up with 10 mg of BSA, 13.5 mg of catalase, and 1.4 mg of ferritin and was diluted up to final volume of 2 mL with gel filtration buffer (see section on protein purification). About 1 mL of the standard solution was injected onto the gel filtration column and was run at 1 mL/min for 1.5 column volumes. Observed elution times were used to calculate the K_{av} and plotted against the log of molecular weight (Figure 2.11). The calculated size of NcsB1 in solution was 84.3 kDa, which is roughly equal to two monomers.

Kinetics of NcsB1 mutants

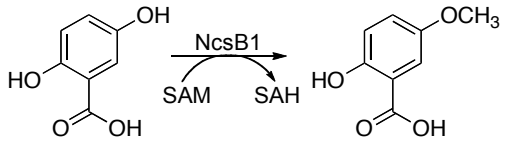
To determine the kinetic parameters of all mutant constructs with the natural substrate, the previously reported biochemical assay was used with minor modifications (*8*). Reactions were run at 25 ºC and contained 10 μM enzyme, 2.5 mM SAM, 100 mM phosphate buffer at pH 6.5 and varying concentrations of **2.3** ranging from 50 μM to 2.5 mM for a total volume of 50 μL (see Chapter 3 for synthesis of **2.3**). Reactions were

quenched with trifluoroacetic acid to a final concentration of 16% after 15 minutes, except for WT and Arg11Ala, which were quenched at 10 min. Samples were centrifuged at 14,000 g for 2 minutes and the supernatant was analyzed by HPLC at 340 nm (eluent A: water with 0.1% trifluoroacetic acid; eluent B: acetonitrile; gradient: 0-12 min at 10-60% B, 12-17 min at 0% B at a flow rate of 1 mL/min on a VYDAC C18 Protein and Peptide column, 4.6 x 250 mm). To obtain kinetic parameters, initial rates were plotted against substrate concentration and were fitted to the Michaelis-Menten equation using KaleidaGraph (Synergy Software, Reading, PA).

Determination of substrate specificity of NcsB1 for benzoic acids

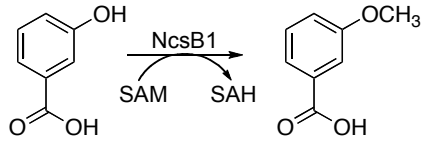
To measure the ability of NcsB1 to methylate hydroxybenzoic acids **2.8** and **2.9**, the previously reported biochemical assay was used with minor modifications (*8*). Benzoic acid substrates were purchased from Aldrich and used without purification. A 100 μL reaction containing 100 mM sodium phosphate pH 6.0, 2.5 mM SAM, and 2.5 mM benzoic acid was initiated with 100 μM NcsB1 at 25 ºC. A 50 μL aliquot was quenched with trifluoroacetic acid to a final concentration of 16% (Aldrich) at 1.5 and 24 hr. Samples were centrifuged at 14,000 *g* for 2 min and the supernatant was analyzed by HPLC (see method above). UV detection was varied based on the UV_{max} of the substrate: **2.8** was observed at 320 nm, **2.9** was observed at 287 nm. The product(s) of the enzymatic reactions of NcsB1 with the various benzoic acid substrates were analyzed on an Agilent 6210 time-of-flight LC/MS in negative mode. Enzymatic reactions were set up as above. The reactions were quenched after 18 h at 25 ºC with trifluoroacetic acid and were submitted for LC-MS analysis (eluent A: water with 0.1% formic acid; eluent

B: acetonitrile with 0.1% formic acid; gradient: 0-12 min at 10-60% B, 12-15 min at 60% B, 15-17 min at 60-10% B, 17-20 min at 10% B at a flow rate of 0.2 mL/min). Products of the enzymatic reactions were isolated from a large-scale reaction, purified by HPLC and subjected to ^{1}H NMR and NOE analysis.



2.8 **2.10**

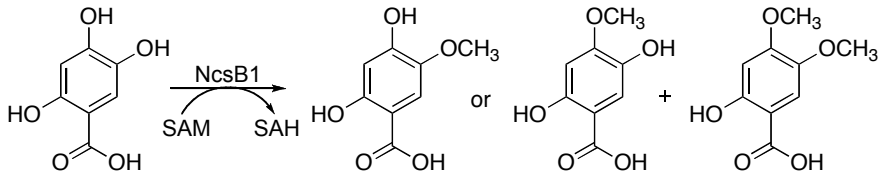
2,5-dihydroxybenzoic acid (**2.8**) to 2-hydroxy-5-methoxybenzoic acid (**2.10**) **LC-MS (ESI$^-$):** found m/z 153.1 and 167.1, respectively ([M-H]$^-$ for both). **^{1}H NMR (500MHz, CD_3OD) δ:** 7.36 (1H, d, *J*=5Hz, C6-H), 7.10 (1H, dd, *J*=10, 5Hz, C4-H), 6.87 (1H, d, *J*=10Hz, C3-H), 3.78 (3H, s, C5-OCH_3). **NOE (500MHz, CD_3OD):** correlations between Hs at δ 7.36/3.78 and 7.10/3.78.



2.9 **2.11**

3-hydroxybenzoic acid (**2.9**) to 3-methoxybenzoic acid (**2.11**) **LC-MS (ESI$^-$):** found *m/z* 137.1 and 151.1, respectively ([M-H]$^-$ for both).

2,4,5-trihydroxy benzoic acid to single methylated product **LC-MS (307 nm, ESI$^-$):** found *m/z* 169.1 and 183.1, respectively ([M-H]$^-$ for both) (trace amounts of 2-hydroxy-4,5-dimethoxy benzoic acid, found *m/z* 197.1, [M-H]$^-$ for overnight reaction).



Determination of kinetic parameters of NcsB1 with benzoic acids

To determine the kinetic parameters of NcsB1 with **2.8** and **2.9** and mutant constructs with **2.8**, 50 μL reaction mixtures containing 2.5 mM SAM, 100 mM phosphate buffer at pH 6.5 and varying concentrations of benzoic acid ranging from 50 μM to 2.0 mM were set up. Reactions were initiated by the addition of 100 mM enzyme, and run at 25 °C. Reactions were quenched with trifluoroacetic to a final concentration of 16% at a specified time to determine the initial rate. Reaction times were as follows: **2.8** with WT, Arg11Ala, and Arg11Lys for 90 min, **2.8** with Tyr293Ile and **2.9** with WT for 120 min, **2.8** with Ala11Trp for 60 min. Reactions were analyzed by UV detection as described above. To obtain kinetic parameters, initial rates were plotted against substrate concentration and were fitted to the Michaelis-Menten equation using KaleidaGraph (Synergy Software, Reading, PA).

2.11 References

1. Liu, W., Nonaka, K., Nie, L., Zhang, J., Christenson, S. D., Bae, J., Van Lanen, S. G., Zazopoulos, E., Farnet, C. M., Yang, C. F., and Shen, B. (2005) The neocarzinostatin biosynthetic gene cluster from Streptomyces carzinostaticus ATCC 15944 involving two iterative type I polyketide synthases, *Chem. Biol. 12*, 293-302.
2. Zhang, J., Van Lanen, S. G., Ju, J., Liu, W., Dorrestein, P. C., Li, W., Kelleher, N. L., and Shen, B. (2008) A phosphopantetheinylating polyketide synthase producing a linear polyene to initiate enediyne antitumor antibiotic biosynthesis, *Proc. Natl. Acad. Sci. U.S.A. 105*, 1460-1465.
3. Sthapit, B., Oh, T. J., Lamichhane, R., Liou, K., Lee, H. C., Kim, C. G., and Sohng, J. K. (2004) Neocarzinostatin naphthoate synthase: an unique iterative type I PKS from neocarzinostatin producer Streptomyces carzinostaticus, *FEBS Lett. 566*, 201-206.
4. Weitnauer, G., Muhlenweg, A., Trefzer, A., Hoffmeister, D., Sussmuth, R. D., Jung, G., Welzel, K., Vente, A., Girreser, U., and Bechthold, A. (2001) Biosynthesis of the orthosomycin antibiotic avilamycin A: deductions from the molecular analysis of the avi biosynthetic gene cluster of Streptomyces viridochromogenes Tu57 and production of new antibiotics, *Chem. Biol. 8*, 569-581.
5. Jia, X. Y., Tian, Z. H., Shao, L., Qu, X. D., Zhao, Q. F., Tang, J., Tang, G. L., and Liu, W. (2006) Genetic characterization of the chlorothricin gene cluster as a model for spirotetronate antibiotic biosynthesis, *Chem. Biol. 13*, 575-585.
6. Zhao, Q., He, Q., Ding, W., Tang, M., Kang, Q., Yu, Y., Deng, W., Zhang, Q., Fang, J., Tang, G., and Liu, W. (2008) Characterization of the azinomycin B biosynthetic gene cluster revealing a different iterative type I polyketide synthase for naphthoate biosynthesis, *Chem. Biol. 15*, 693-705.
7. Cooke, H. A., Zhang, J., Griffin, M. A., Nonaka, K., Van Lanen, S. G., Shen, B., and Bruner, S. D. (2007) Characterization of NcsB2 as a promiscuous naphthoic acid/coenzyme A ligase integral to the biosynthesis of the enediyne antitumor antibiotic neocarzinostatin, *J. Am. Chem. Soc. 129*, 7728-7729.
8. Luo, Y., Lin, S., Zhang, J., Cooke, H. A., Bruner, S. D., and Shen, B. (2008) Regiospecific O-methylation of naphthoic acids catalyzed by NcsB1, an O-methyltransferase involved in the biosynthesis of the enediyne antitumor antibiotic neocarzinostatin, *J. Biol. Chem. 283*, 14694-14702.
9. Schubert, H. L., Blumenthal, R. M., and Cheng, X. (2003) Many paths to methyltransfer: a chronicle of convergence, *Trends Biochem. Sci. 28*, 329-335.
10. Vidgren, J., Svensson, L. A., and Liljas, A. (1994) Crystal structure of catechol O-methyltransferase, *Nature 368*, 354-358.
11. Martin, J. L., and McMillan, F. M. (2002) SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold, *Curr. Opin. Struct. Biol. 12*, 783-793.

12. Madduri, K., Torti, F., Colombo, A. L., and Hutchinson, C. R. (1993) Cloning and sequencing of a gene encoding carminomycin 4-O-methyltransferase from Streptomyces peucetius and its expression in Escherichia coli, *J. Bacteriol. 175*, 3900-3904.
13. Jansson, A., Koskiniemi, H., Mantsala, P., Niemi, J., and Schneider, G. (2004) Crystal structure of a ternary complex of DnrK, a methyltransferase in daunorubicin biosynthesis, with bound products, *J. Biol. Chem. 279*, 41149-41156.
14. Jansson, A., Koskiniemi, H., Erola, A., Wang, J., Mantsala, P., Schneider, G., and Niemi, J. (2005) Aclacinomycin 10-hydroxylase is a novel substrate-assisted hydroxylase requiring S-adenosyl-L-methionine as cofactor, *J. Biol. Chem. 280*, 3636-3644.
15. Jansson, A., Niemi, J., Lindqvist, Y., Mantsala, P., and Schneider, G. (2003) Crystal structure of aclacinomycin-10-hydroxylase, a S-adenosyl-L-methionine-dependent methyltransferase homolog involved in anthracycline biosynthesis in Streptomyces purpurascens, *J. Mol. Biol. 334*, 269-280.
16. Oster, L. M., Lester, D. R., Terwisscha van Scheltinga, A., Svenda, M., van Lun, M., Genereux, C., and Andersson, I. (2006) Insights into cephamycin biosynthesis: the crystal structure of CmcI from Streptomyces clavuligerus, *J. Mol. Biol. 358*, 546-558.
17. Hou, X., Wang, Y., Zhou, Z., Bao, S., Lin, Y., and Gong, W. (2007) Crystal structure of SAM-dependent O-methyltransferase from pathogenic bacterium Leptospira interrogans, *J. Struct. Biol. 159*, 523-528.
18. Cho, J. H., Park, Y., Ahn, J. H., Lim, Y., and Rhee, S. (2008) Structural and functional insights into O-methyltransferase from Bacillus cereus, *J. Mol. Biol. 382*, 987-997.
19. Singh, S., McCoy, J. G., Zhang, C., Bingman, C. A., Phillips, G. N., Jr., and Thorson, J. S. (2008) Structure and mechanism of the rebeccamycin sugar 4'-O-methyltransferase RebM, *J. Biol. Chem. 283*, 22628-22636.
20. Holm, L., Kaariainen, S., Rosenstrom, P., and Schenkel, A. (2008) Searching protein structure databases with DaliLite v.3, *Bioinformatics 24*, 2780-2781.
21. Zubieta, C., He, X. Z., Dixon, R. A., and Noel, J. P. (2001) Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases, *Nat. Struct. Biol. 8*, 271-279.
22. Parsons, J. F., Greenhagen, B. T., Shi, K., Calabrese, K., Robinson, H., and Ladner, J. E. (2007) Structural and functional analysis of the pyocyanin biosynthetic protein PhzM from Pseudomonas aeruginosa, *Biochemistry 46*, 1821-1828.
23. Zubieta, C., Kota, P., Ferrer, J. L., Dixon, R. A., and Noel, J. P. (2002) Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase, *Plant Cell 14*, 1265-1277.
24. Liu, W., Christenson, S. D., Standage, S., and Shen, B. (2002) Biosynthesis of the enediyne antitumor antibiotic C-1027, *Science 297*, 1170-1173.

25. Van Lanen, S. G., Lin, S., and Shen, B. (2008) Biosynthesis of the enediyne antitumor antibiotic C-1027 involves a new branching point in chorismate metabolism, *Proc. Natl. Acad. Sci. U.S.A. 105*, 494-499.
26. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem. 72*, 248-254.
27. Ji, N., Rosen, B. M., and Myers, A. G. (2004) Method for the rapid synthesis of highly functionalized 2-hydroxy-1-naphthoates. Syntheses of the naphthoic acid components of neocarzinostatin chromophore and N1999A2, *Org. Lett. 6*, 4551-4553.
28. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Macromol. Crystallogr. Pt. A. 276*, 307-326.
29. Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J. Appl. Crystallogr. 40*, 658-674.
30. Bailey, S. (1994) The Ccp4 Suite - Programs for Protein Crystallography, *Acta Crystallogr. D. 50*, 760-763.
31. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr. D. 60*, 2126-2132.
32. Brunger, A. T. (2007) Version 1.2 of the Crystallography and NMR system, *Nature Prot. 2*, 2728-2733.
33. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D. 54*, 905-921.
34. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement, *J. Appl. Crystallog. 30*, 1022-1025.
35. Schuttelkopf, A. W., and van Aalten, D. M. F. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes, *Acta Crystallogr. D. 60*, 1355-1363.
36. Kleywegt, G. J. (2007) Crystallographic refinement of ligand complexes, *Acta Crystallogr. D. 63*, 94-100.
37. Sippl, M. J., and Wiederstein, M. (2008) A note on difficult structure alignment problems, *Bioinformatics 24*, 426-427.
38. Sippl, M. J. (2008) On distance and similarity in fold space, *Bioinformatics 24*, 872-873.

Chapter 3

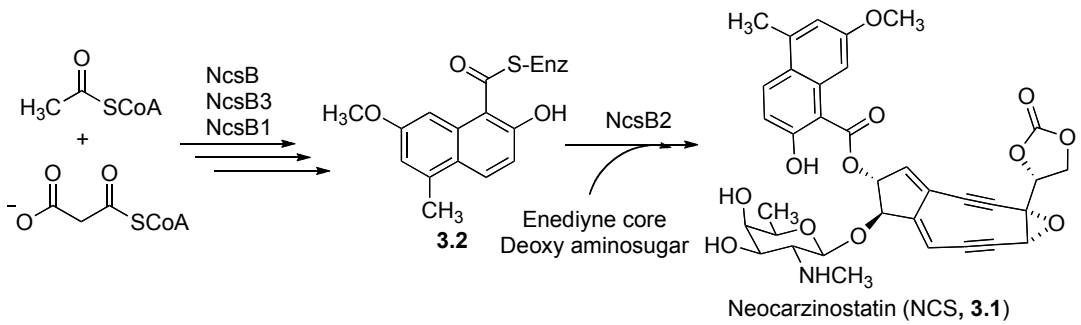
Characterization of NcsB2 as a Promiscuous Naphthoic Acid/Coenzyme A Ligase Integral to the Biosynthesis of the Enediyne Antitumor Antibiotic Neocarzinostatin

Adapted in part from an article published in *J. Am. Chem. Soc.* **2007**, *129*, 7728-9 in collaboration with J. Zhang, M.A. Griffin, K. Nonaka, S.G. Van Lanen, B. Shen, and S.D. Bruner, and including experimental work done by E.L. Guenther and E.J. Feris.

3.1 Introduction

When the NCS (**3.1**) biosynthetic gene cluster from *Streptomyces carzinostaticus* ATCC15944 was first sequenced, a convergent pathway featuring an NcsB2-catalyzed direct coupling between naphthoyl-*S*-NcsB and an enediyne core intermediate was proposed (Scheme 3.1) (*1*). This prediction was consistent with the sequence-based annotation of NcsB2 as a CoA-ligase, which would need to act in the reverse direction on a naphthoyl-*S*-NcsB tethered intermediate to perform the predicted coupling. Contrary to this hypothesis, when NcsB, a polyketide synthase (PKS), was overexpressed in heterologous hosts, accumulation of free 2-hydroxy-5-methyl-1-naphthoic acid was observed (*2*). This finding would argue against the involvement of a tethered intermediate in NCS biosynthesis as originally proposed.

Scheme 3.1: Original proposed biosynthesis



A reverse CoA-ligase would act on the enzyme-tethered aryl-thioester (**3.2**) with a donor hydroxyl group coming from, in this case, the enediyne core (Figure 3.1). NcsB2 is homologous to a number of biochemically and structurally characterized adenylate- and CoA-ligases. While there is little literature precedence for reverse CoA-ligase

activity, there is substantial bioinformatic support for NcsB2 playing a role in activation of the free naphthoic acid based on homology to characterized enzymes. DhbE is one such homolog (42% identity and 62% similarity). It is a standalone adenylation (A) domain involved in the biosynthesis of the catecholic siderophore bacillibactin (*3*, *4*). DhbE activates 2,3-dihydroxybenzoic acid via adenylation and subsequently attaches it to holo-DhbB (an isochorismate lyase and aryl-carrier protein) via a phosphopantetheinyl arm (Scheme 3.2). Subsequent transformations with the NRPS DhbF produce the natural product. It seems likely that NcsB2 would act in a similar manner to DhbE, though NcsB2 might instead transfer the functionalized naphthoic acid either directly onto the enediyne core or onto CoA. For the latter activity to occur, an additional enzyme-catalyzed step would be required to complete the biosynthesis of NCS.

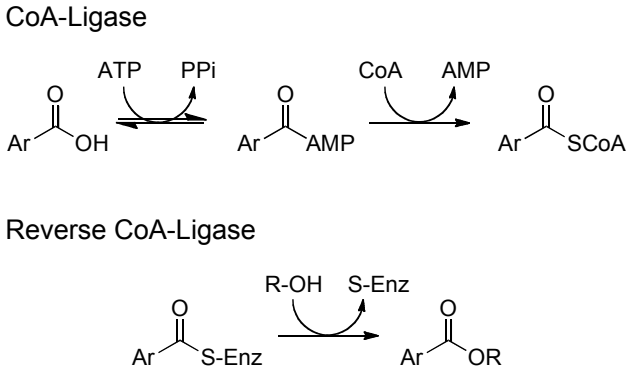
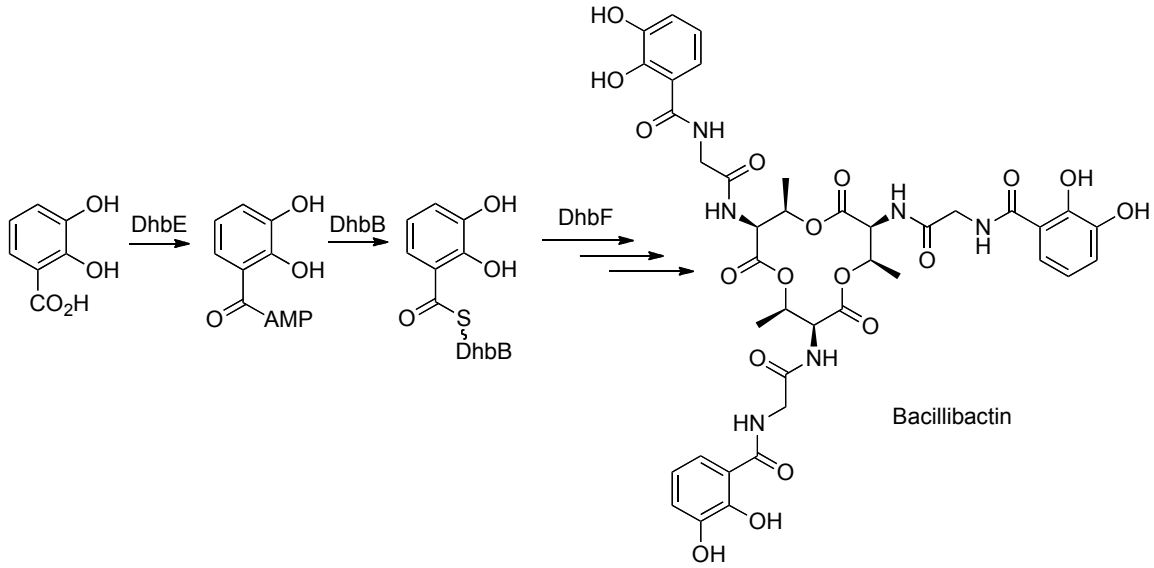


Figure 3.1: Reactions catalyzed by a CoA-ligase and a hypothetical reverse CoA-ligase. A CoA-ligase acts by first adenylating an aryl acid and then transfers it to the thiol of CoA to form a thioester. A reverse CoA-ligase would transfer an enzyme tethered aryl group to a donor hydroxyl group. In the case of NcsB2, R = enediyne core.

Scheme 3.2: Biosynthesis of bacillibactin



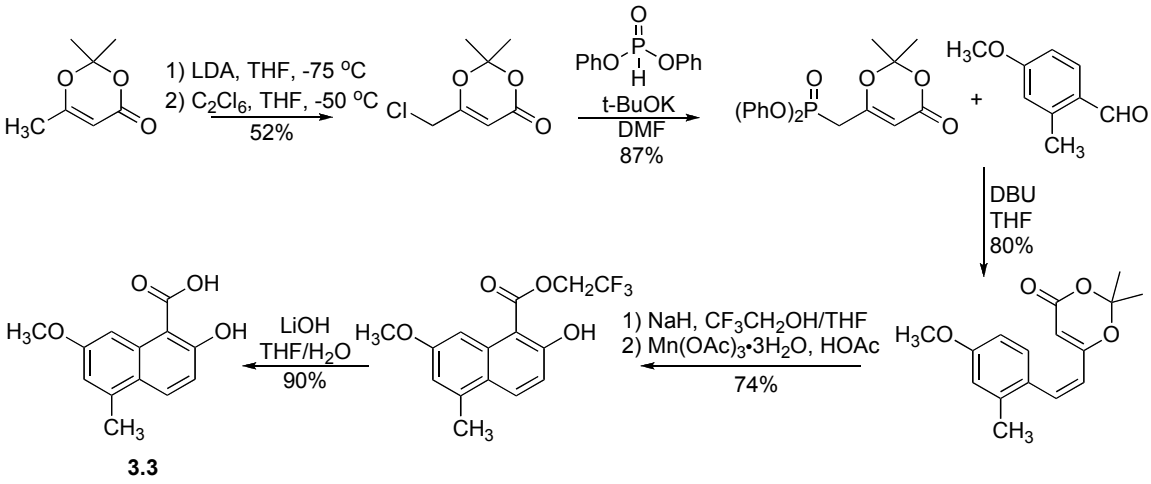
In this chapter, the *in vitro* characterization of NcsB2, including studies into its promiscuity as an adenylation domain and its activity following activation of naphthoic acid are discussed. A revised biosynthetic pathway of 2-hydroxy-7-methoxy-5-methyl naphthoic acid is presented featuring its incorporation into NCS with free naphthoic acids as intermediates. The results of these studies present an outstanding opportunity to produce novel analogs of NCS by engineering NCS biosynthesis.

3.2 Initial studies

It was first confirmed that NcsB2 is essential for NCS biosynthesis by inactivation and complementation of the resultant *ΔncsB2* mutants *in vivo* (*5*). Replacing *ncsB2* with a mutant copy yielded an *S. carzinostaticus* SB5006 mutant strain that completely lost its ability to produce NCS. Production was partially restored to SB5006 by overexpressing a functional copy of *ncsB2* in trans, alleviating any concern over potential polar effects

from *ncsB2* inactivation on the expression of downstream genes in SB5006. While these initial studies made it clear that NcsB2 is required for NCS biosynthesis, the actual role of NcsB2 had not yet been established. Whether NcsB2 acted on free naphthoic acid or on an enzyme tethered naphthoyl substrate would have considerable implications for the overall biosynthetic scheme. In order to fully understand NcsB2 activity, we needed to first synthesize the predicted free naphthoic acid substrate (**3.3**), which was done using a rapid synthesis developed by Ji *et al* (Scheme 3.3) (*6*). If NcsB2 could adenylate free naphthoic acids, the earlier prediction of its activity would likely be incorrect, necessitating revision of NCS biosynthesis.

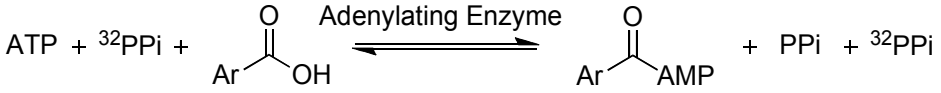
Scheme 3.3: Synthesis of 2-hydroxy-7-methoxy-5-methyl naphthoic acid



3.3 Characterization of NcsB2 as a CoA ligase

In order to determine the role of NcsB2 in NCS biosynthesis, we cloned the *ncsb2* gene from *S. carzinostaticus* genomic DNA, over-expressed NcsB2 in *E. coli*, purified it to homogeneity, and characterized it for naphthoyl-CoA ligase activity *in vitro*. CoA ligases have two activities: (1) ATP-dependent activation of carboxylic acids as acyl-AMP esters and (2) formation of acyl-CoAs by coupling the activated acyl group with the thiol nucleophile of CoA. We examined the first half of the enzymatic activity of NcsB2 by the ATP-[^{32}P]pyrophosphate exchange assay, which has been widely used in the characterization of adenylating enzymes (Scheme 3.4) (*7*). This assay relies on the incorporation of labeled pyrophosphate into unlabeled ATP, which occurs due to the

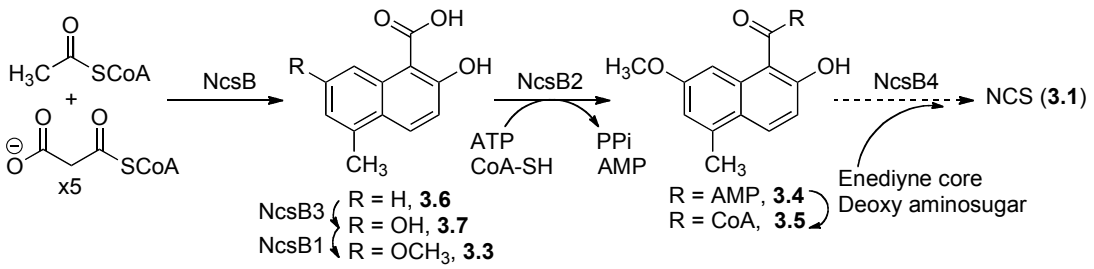
Scheme 3.4: ATP-pyrophosphate exchange assay



reversibility of the adenylation reaction. The reactions are quenched with acid and incubated on charcoal, for which ATP has a strong affinity. The charcoal is filtered and washed and is submitted for scintillation counting. The amount of radioactivity counted corresponds to the reaction rate. Efficient exchange of ATP was readily observed, indicative of the formation of naphthoyl-AMP ester **3.4** (Scheme 3.5). The reaction occurs with a k_{cat} of 38 min^{-1} ± 2 and K_M of 0.6 μM ± 0.2, values which are comparable to those for other characterized adenylation enzymes such as DhbE (k_{cat} of 167 min^{-1} and K_M of 7.6 μM) (*3*). At this point, we knew that NcsB2 acted by adenylating a free naphthoic acid substrate, making it very unlikely that NcsB2 also acted as a reverse CoA-ligase on

enzyme-tethered naphthoyl intermediates. Still, based on homology, we knew that NcsB2 might also catalyze a second step after adenylation. We verified the second half of the CoA ligase reaction of NcsB2 by directly monitoring the formation of **3.5** from **3.3** in the presence of ATP and CoA using HPLC. As shown in Scheme 3.5, **3.3** was converted to **3.5** via **3.4**, which was confirmed by LC-MS. CoA was the only nucleophile among a range of potential coupling partners tested that efficiently supported this conversion, including those that were considered analogs to the enediyne core, such as cyclopentanol.

Scheme 3.5: Revised biosynthetic pathway of naphthoic acid and NCS



3.4 Revision of biosynthetic pathway

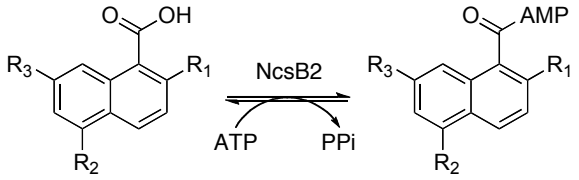
The characterization of NcsB2 as a naphthoyl-CoA ligase prompted us to revise this portion of the biosynthetic pathway, which includes the production of free naphthoic acid (**3.3**) and its subsequent incorporation into NCS. The functional assignments of NcsB as the naphthoic acid synthase, NcsB3 as the hydroxylase, and NcsB1 as the *O*-methyltransferase, all acting on naphthoyl-*S*-NcsB tethered intermediates, are consistent with the current finding of NcsB2 as a CoA ligase (*1*). However, we now prefer a scheme where NcsB produces the free naphthoic acid intermediate (*2*) for hydroxylation by NcsB3, followed by *O*-methylation by NcsB1 ((*8*) and as discussed in Chapter 2) to produce **3.3** (Scheme 3.5). The new proposal would require an additional enzyme to catalyze the coupling of **3.5** to the enediyne core. Re-examination of the NCS biosynthetic gene cluster indeed led to the identification of an additional open reading frame, *orf27*, immediately downstream of *ncsB3*, whose function was not assigned previously (*1*). We now renamed *orf27* as *ncsB4*. NcsB4 shows high sequence similarity to known esterases and acyltransferases, serving as a candidate to catalyze the coupling between **3.5** and the enediyne core to afford NCS.

3.5 Substrate specificity of NcsB2

Next, we directly compared the kinetic parameters of **3.3** as a substrate for NcsB2 with those of **3.6** and **3.7**, two alternative substrates for NcsB2 depending on the timing of the CoA-ester formation step in **3.5** biosynthesis. These alternate substrates and others were synthesized using the same synthetic scheme as that of the natural substrate, but the

functional groups on the benzaldehyde to be coupled to the phosphonate intermediate were varied (with the experimental assistance from M. Griffin). Both **3.6** and **3.7** can be activated by NcsB2 in the presence of ATP but with k_{cat}/K_m values ~2 fold lower than **3.3** (Table 3.1). Taken together, these results are consistent with the assignment of **3.3** as the preferred substrate of NcsB2 and support the timing of the individual steps as proposed for **3.5** biosynthesis (Scheme 3.5).

Table 3.1: Naphthoic acid substrate scope of the NcsB2 CoA ligase



3.3 (R_1 = OH, R_2 = CH_3, R_3 = OCH_3)
3.6 (R_1 = OH, R_2 = CH_3, R_3 = H)
3.7 (R_1 = OH, R_2 = CH_3, R_3 = OH)
3.8 (R_1 = OH, R_2 = H, R_3 = OCH_3)
3.9 (R_1 = OH, R_2 = NO_2, R_3 = H)
3.10 (R_1 = OH, R_2 = R_3 = OCH_3)
3.11 (R_1 = OH, R_2 = R_3 = H)
3.12 (R_1 = R_2 = R_3 = H)

Naphthoic acid	K_m (μM)	k_{cat} (min^{-1})	Rel. k_{cat}/K_m
3.3	0.59 ± 0.2	38 ± 2.0	1
3.6	0.25 ± 0.1	8.1 ± 0.6	0.47
3.7	3.1 ± 0.1	93 ± 0.6	0.50
3.8	2.3 ± 0.7	7.1 ± 0.8	4.8×10^{-2}
3.9	> 500	-	-
3.10	1.1 ± 0.4	25 ± 2.4	0.35
3.11	3.4 ± 0.9	2.8 ± 0.2	1.3×10^{-2}
3.12	18 ± 9	6.7 ± 1.4	5.8×10^{-3}

The fact that NcsB2 activated all three naphthoic acid biosynthetic intermediates inspired us to further probe its substrate specificity. Since NcsB2 acts on free naphthoic acids, as opposed to an enzyme-tethered substrate, we thought that perhaps it could activate a larger variety of naphthoic acids, which could be subsequently incorporated

into NCS to produce analogs. A variety of 1-naphthoic acid analogs with different substitutions at the 2-, 5-, or 7-position (**3.8–3.12**) were either prepared or were purchased. Each of the analogs was subjected to the ATP-[^{32}P]pyrophosphate exchange assay to examine if they would serve as substrates for NcsB2. Remarkably, substitutions at all three positions were well tolerated with their observed k_{cat} and K_M values ranging within 30-fold of the natural substrate (Table 3.1). Even unfunctionalized naphthoic acids (**3.11** and **3.12**) were adenylated, albeit at a significantly lower catalytic efficiency. Substrate **3.11** has a K_m value within the range of the other alternate substrates, while that of **3.12** is much higher. This result suggests that the hydroxyl group adjacent to the acid is an important feature for substrate recognition by the NcsB2. The one exception to the rather relaxed substrate specificity of NcsB2 was **3.9**, which showed essentially no detectable activity, presumably due to the strong electron-withdrawing effect of the nitro group. We also tested 2-hydroxybenzoic acid, benzoic acid, and the benzoxazolinate moiety from C-1027 (see Appendix A for synthesis details) as potential substrates for NcsB2 but no measurable activity was detected, suggesting the naphthalene ring as the minimal substrate recognition element for NcsB2.

3.6 Crystallization attempts

In addition to characterizing NcsB2 biochemically, we also sought to solve its X-ray crystal structure, as we hoped to gain insight into specificity determinants. Thus far there have been 47 structures solved of adenylate forming enzymes (*9*). Some examples include DhbE (46% identical, 62% similar, PDB code 1MDB (*4*)), firefly luciferase (25% identical, 41% similar, PDB code 1BA38 *(10)*) and 4-chlorobenzoyl-CoA ligase (25%

identical, 42% similar, PDB code 1T5D (*11*)). For the most part, sequence homology is low, typically around 20% identical. Nevertheless, the structures of DhbE, firefly luciferase, and of the phenylalanine activating module of the Gramicidin S NRPS (*12*) all share a common fold despite low sequence homology (*4*).

Initial attempts to crystallize NcsB2 included screening using Hampton Crystal Screens at 4 and 20 ºC. Unfortunately, no reproducible crystal hits were found; therefore addition of glycerol, ATP, AMP, and $MgCl_2$ in a variety of combinations and different temperatures was tried. Again, there were no reproducible hits, so the next step was to attempt to trap the enzyme in one conformation in hopes that this would encourage crystal formation.

Often, addition of a ligand, inhibitor, or stable analog of a catalytic intermediate will help crystallization by maintaining the enzyme in one conformation (*13*), and for an enzyme such as NcsB2 which binds to multiple substrates (naphthoic acid, ATP, CoA), a fair amount of movement may occur during one catalytic cycle. A variety of local and overall changes in conformation have been observed for adenylate ligase homologs (*4, 9, 14*). The first intermediate analog explored was a naphthoyl sulfamoyl adenosine. Sulfamoyl adenosines are analogous to the adenosine monophosphate intermediate produced by adenylation domains and tRNA synthetases but are non-hydrolyzable under biological conditions. This type of compound has been shown to inhibit amino and aryl acid adenylation enzymes (*15, 16*). To synthesize these inhibitors, adenosine is acetonide protected and is coupled to the sulfamoyl group by reacting with sulfamoyl chloride. Next, the amino or aryl acid is either coupled to the protected sulfamoyl adenosine using

standard peptide coupling reagents or is first activated with *N*-hydroxysuccinimide for subsequent coupling. TFA deprotection affords the aryl or amino sulfamoyl adenosine inhibitor in four linear steps. The first stable analog made was 1-naphthoyl sulfamoyl adenosine (Figure 3.2, **3.13**) and it was incubated with NcsB2. Unfortunately, this inhibitor did not encourage crystal growth. Next, the natural substrate, **3.3**, was coupled to sulfamoyl adenosine, producing **3.14** (synthesized and characterized by E. Feris). Once again, X-ray quality crystals were not obtained. The synthesis of a third stable intermediate analog, naphthoyl amino-CoA **3.15**, is currently being developed by E. Guenther, using a chemoenzymatic synthesis of amino CoA as described previously by Y. Liu (*17*), to provide a stable analog of the naphthoyl CoA thioester.

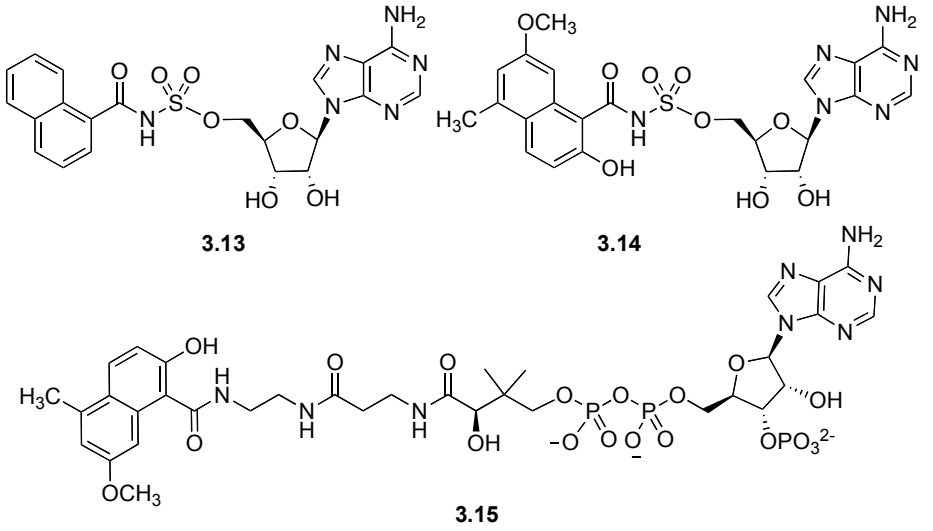


Figure 3.2: Stable intermediate analogs for cocrystallization with NcsB2

3.7 Truncated constructs of NcsB2

In addition to conformational changes being problematic to crystal formation, floppy ends or loops can also be detrimental to reaching this goal. Therefore, truncated constructs of NcsB2 were prepared based on sequence alignment with SgcD5 and DhbE to remove what might be disordered regions on the N- and C-termini (Figure 3.3). SgcD5 shares little sequence homology with NcsB2 but it is proposed to be a CoA-synthetase for the benzoxazolinate moiety on C-1027 (*18*), acting in a similar manner to NcsB2. Based on homology, NcsB2 has eight amino acids on the N-terminus and four on the C-terminus that might not contribute to the fold. We hoped that truncating NcsB2 would lower the amount of disorder in the protein, aiding in crystallization. Overall, NcsB2 is 19 amino acids longer than DhbE, much of which is in the middle of the enzyme. Grown under the same conditions as the original NcsB2 construct, over-expression and purification resulted in less than half of the typical amount of NcsB2 and attempts at improving protein yield by varying length and temperature of expression were both unsuccessful. Neither of the truncated constructs aided crystal growth and were fraught with over-expression problems.

Figure 3.3: Sequence alignment of NcsB2 with sequence homology DhbE (1MD9_A) and SgcD5 (AAL06665), a putative benzoxazolinate CoA synthetase from C-1027 biosynthesis.

3.8 Conclusion

The characterization of NcsB2 as a CoA ligase with promiscuous substrate specificity is exciting because it presents an outstanding opportunity to produce novel analogs of NCS by engineering NCS biosynthesis. Analogous to adenylation domains found in nonribosomal peptide biosynthesis, NcsB2 could be viewed as the "gate-keeper" that selects and activates naphthoic acids to be incorporated into NCS, assuming that the ensuing NcsB4 acyltransferase also possesses relaxed substrate specificity. Generation of novel enediynes has found success as exemplified by the production of C-1027 and calicheamicin analogs (*19, 20*), the availability of which has already unveiled new insight into the mode of action and drug discovery for the enediynes (*21*). This work now sets the stage to explore these possibilities within the NCS scaffold, one of two enediyne natural products currently used as anticancer drugs (*22, 23*).

3.9 Experimentals

General

DNA oligonucleotide primers were obtained from Integrated DNA Technology (Coralville, IA). Restriction enzymes and bovine serum albumin (BSA) were purchased from New England Biolabs (Ipswich, MA). Thrombin was purchased from Sigma-Aldrich (St. Louis, MO). DNA purification kits and Ni-NTA resin were from Qiagen (Valencia, CA). Tetrasodium [^{32}P] pyrophosphate was purchased from Perkin Elmer Life Science (Boston, MA). DNA sequencing was performed by Genewiz (New Brunswick, New Jersey). Columns for protein purification were from GE Healthcare. Competent *E. coli* cells were purchased from Invitrogen (Carlsbad, CA) and Fisher Scientific (Pittsburgh, PA). Liquid scintillation counting was performed on a Beckman Coulter LS 6500 Multi-purpose Scintillation Counter. Chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO) or VWR (West Chester, PA). Chemical reactions, unless otherwise noted, were run in solvents that were dried using stills by Glass Contour (Santa Monica, CA). Deuterated solvents were purchased from Cambridge Isotopes (Andover, MA). Silica gel used for all flash chromatography purifications was purchased from VWR, EMD brand Geduran. Mass Spectrometry (MS) was performed on a Waters LCT Classic. ^{1}H NMR spectra were recorded on a Varian 500, 400 or 300 MHz spectrometer. Proton chemical shifts are reported in ppm (δ) relative to internal tetramethylsilane (TMS, δ 0.0 ppm). Data are reported as follows: chemical shift (multiplicity [singlet (s), doublet (d), triplet (t), quartet (q), and multiplet (m)], coupling constants [Hz], integration). ^{13}C NMR spectra were recorded on a Varian 500 (126 MHz) or 400 (100

MHz) spectrometer with complete proton decoupling. Carbon chemical shifts are reported in ppm (δ) relative to the $CDCl_3$ or CD_3OD as the external standard (δ 77.26 ppm or 49.00 ppm, respectively).

Cloning of ncsB2 from genomic DNA.

A 2.4 kb fragment of the genomic DNA containing the *ncsB2* gene was amplified from *S. carzinostaticus* ATCC 15944 by PCR using the oligonucleotide primers 5'-GAA CTT GTC CGG CGA GTC GAG ATC-3' (forward) and 5'-CTG AGG TTC GTT CTC ACC GC-3' (reverse) with the following PCR program: 94.0 ºC for 5 min, then 30 cycles of 94.0 ºC for 1 min, 59.1 ºC for 1 min, and 72.0 ºC for 5 min, followed by 72.0 ºC for 10 min. The *ncsB2* gene was then amplified from the fragment using the primers 5'-GGC **CAT ATG** CAC GAG ACA GCA GCA GC-3' (forward, the *Nde*I site is shown in bold) and 5'-GGC **GGA TCC** TCA GGT TTC GGA CAT C-3' (reverse, the *Bam*HI site is shown in bold) with the following PCR program: 92.0 ºC for 2 min, then 30 cycles of 92.0 ºC for 1 min, 65.1 ºC for 1 min, and 72.0 ºC for 2 min, followed by 72.0 ºC for 10 min. The *ncsB2* gene was ligated into the pET28a vector as a 1.7 kb an *Nde*I-*Bam*HI fragment and transformed into competent cells. The resulting plasmid contained *ncsB2* as an *N*-terminal His_6-tagged fusion protein and was confirmed by double digest and sequencing.

Overproduction and purification of NcsB2

Plasmid pBS5035 was transformed into BL21(DE3)pLysS chemically competent cells. Cells were spread onto an LB agar plate containing 30 μg/mL of both

chloramphenicol and kanamycin; the plate was incubated at 37 ºC for 16 h. One colony was then used to inoculate a 1L flask of LB with the same antibiotics. The cells were grown at 37 ºC (125 RPM) until OD_{600} reached 0.55-0.80. Expression was induced at 25 ºC with 100 μM IPTG and cells continued growth for 6 h. Cells were collected by centrifugation (20 min, 2000 g), resuspended in 25 mL of buffer containing 20 mM Tris-HCl and 500 mM NaCl, pH 7.5 and frozen at -78 ºC. To obtain protein for assays, the cell pellet was thawed and lysed by two passes through a French Press cell disruptor at 1000 PSI. The lysed cells were then centrifuged (20 min, 10,000*g*) to remove cell debris and the protein was purified batchwise using Ni-NTA resin. The purified protein was dialyzed into low salt buffer (50 mM NaCl, 20 mM Tris-HCl, 1 mM β-mercaptoethanol, pH 7.5) diluted to 10 μM and stored at -78 ºC in 20% glycerol for use in assays.

For crystallography, the Ni-NTA purified protein was dialyzed into low salt buffer and concentrated to less than 1 mL. The His_6-tag was cleaved by incubation with 10 U thrombin for 2-3 h at 4 ºC in thrombin cleavage buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM $CaCl_2$). The cleavage was stopped by adding 30 μL 100 mM PMSF in methanol and incubated on ice for 10 min. The protein solution was then diluted, repurified on Ni-NTA resin, dialyzed into low salt buffer, purified on a HiTrap-Q ion exchange column (0-1 M NaCl gradient in 50 mM Tris-HCl pH 7.5 and 1 mM β-mercaptoethanol) and on a Superdex 200 gel filtration column (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM β-mercaptoethanol). The purified protein was concentrated to 10 mg/mL for crystallography (a total yield of 1.5 mg/L of cells or 7.3 mg/L of cells without

cleavage of His_6-tag). Protein concentration was determined using UV absorption at 280 nm.

Identification of ncsB4 as a candidate encoding an acyltransferase to catalyzes coupling between 2-hydroxy-7-methoxy-5-methyl-1-naphthoyl CoA and the enediyne core

The *ncsB4* gene (previously known as *ncs orf27*) is adjacent to and downstream of *ncsB3*, the gene product of which is the P450 monooxygenase predicted to be involved in the maturation of the naphthoic acid moiety. NcsB4 has sequence similarity to proteins found in COG0596, which are predicted hydrolases or acyltransferases of the α/β hydrolase superfamily. The closest functionally characterized homolog is EstC from *Lactobacillus casei* (22% identity/33% similarity; accession no. AAM90952). This family of enzymes mediates both the synthesis and hydrolysis of esters, with EstC displaying specificity towards fatty acid-phenyl esters.

Truncations of NcsB2

New primers were used to produce truncated constructs of NcsB2 with the same restriction sites as for the wild-type construct. N-terminal truncation primers were: 5'-GGC **CAT ATG** CCC GCA GGA TTC GTC CCC-3' (forward) and 5'-GGC **GGA TCC** TCA GGT TTC GGA CAT C-3' (reverse). C-terminal truncation primers were: 5'-GGC **CAT ATG** CAC GAG ACA GCA GCA GC-3' (forward) and 5'-CAT **GGA TCC** TCA ACC GGC GGC AGC AC-3' (reverse). PCR cycling program for the N-terminal truncation was 92.0 ºC for 5 min, then 30 cycles of 92.0 ºC for 1 min, 64.6 ºC for 1 min (annealing), and 72.0 ºC for 2 min (extension), followed by 72.0 ºC for 10 min. For the

C-terminal truncation, the successful annealing temperature was 64.9 ºC. Truncated protein constructs were purified using same purification steps as for native construct, yielding uncleaved N-terminal truncated and C-terminal truncated protein at 3.8 mg/L and 2.5 mg/L, respectively. Protein was screened for crystal formation using Hampton Crystal Screens with cleaved and uncleaved protein (cleavage resulted in a very low yield of pure protein). Screening resulted in no reproducible crystal hits.

ATP-[^{32}P]pyrophosphate exchange assay

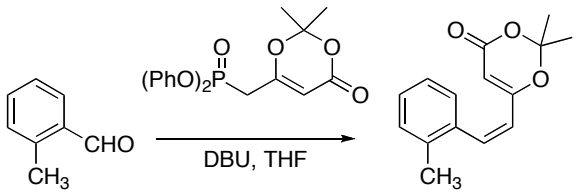
ATP-(^{32}P)PPi exchange by NcsB2 was monitored for the natural substrate (**3.3**) and analogs (**3.4**-**3.12**). K_M and k_{cat} typically were determined from seven different substrate concentrations for each naphthoic acid with a constant ATP concentration. The assay mixture contained 10 mM $MgCl_2$, 5 mM DTT, 5 mM ATP, 0.2 mg/mL BSA, 250 nM NcsB2, 1 mM tetrasodium pyrophosphate decahydrate (containing enough tetrasodium [^{32}P]pyrophosphate to detect activity)**,** 75 mM Tris-HCl, pH 8.0 and a range of substrate concentrations from 0.05 μM to 50 μM in a total volume of 0.1 mL. Reactions were initiated with the addition of 10 μL of a substrate stock. After incubation at 24 ºC for 3 minutes, the reaction was quenched with 0.4 mL of a mixture containing 3.5% perchloric acid, 1.6% (w/v in water) activated charcoal and 100 mM tetrasodium pyrophosphate. The suspension was centrifuged, the supernatant removed, and the charcoal was washed one time with 1 mL of wash solution containing 3.5% perchloric acid and 100 mM tetrasodium pyrophosphate. The charcoal was transferred into 3 mL of Ecolite(+) scintillation fluid using a total volume of 0.6 mL water and submitted for scintillation counting. Specific activity was determined by counting 3 μL of the reaction mixture in 3 mL of scintillation fluid. Significant substrate/product inhibition as well as protein instability caused large errors in the kinetics data, which was partially rectified by the addition of BSA.

Coenzyme A arylation assay

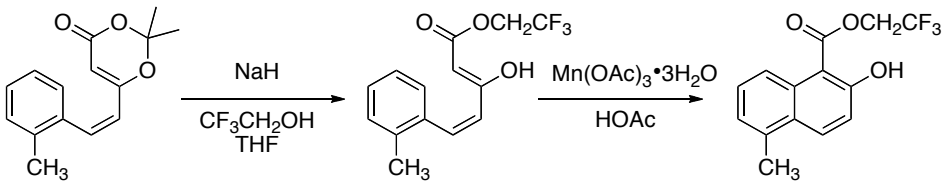
The ability of NcsB2 to couple the activated natural substrate (**3.4**) to CoA was assayed with similar conditions used by Linne et al.[3] An assay mixture containing 1 μM NcsB2, 4 mM ATP, 1 mM CoA, 0.2 mg/mL BSA, 0.5 mM naphthoate substrate (**3.3**) in assay buffer (50 mM HEPES-NaOH, 100 mM NaCl, 1 mM ethylendiaminetetraacetic acid, 10 mM $MgCl_2$, pH 7.5) was incubated at 37 ºC for 1.5 h. The assay was quenched by the addition of TFA to a final concentration of 20%. After incubation on ice for 5 min, the mixture was centrifuged for 2 min at 13,000 rpm and the supernatant was transferred to HPLC-vials. A control without CoA was run as well. The assay products were separated on a VYDAC C18 column (250 mm, ID=4.6 mm) at a flow rate of 1 mL/min using the following conditions (buffer A: 2 mM triethylamine/water; buffer B: 2 mM triethylamine/80% acetonitrile/20% water): 5% buffer B (5 min), linear gradient to 30% buffer B (over 20 min), linear gradient to 95% buffer B (over 4 min) followed by 100% B for 2 min. The products were identified via UV detection at 320 nm. LC-MS was subsequently performed using ESI in negative mode. **MS (ESI$^-$):** *m/z* calculated for $(C_{34}H_{45}N_7O_{19}P_3S)^-$ 980.2, found 979.8. Other nucleophiles were tested for activity, including β-mercaptoethanol, cyclopropanol and 2-indanol and showed no turnover.

Synthesis of substrates and analogs

All substrates were synthesized following the convergent procedure by Myers (*6*). Compounds made in our lab that were previously synthesized all matched published spectra. The following procedures are for the synthesis of three analogs (**3.6**, **3.7**, and **3.8**) tested not previously reported.

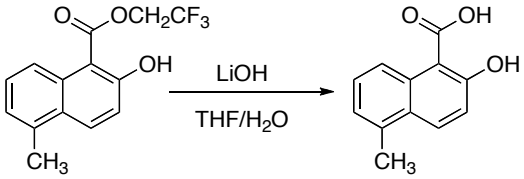


Methyl substituted (Z)-olefin. To a stirring solution of phosphonate (330 mg, 0.886 mmol, 1.4 eq) in THF (5 mL), DBU (188 μL, 1.26 mmol, 2 eq) was added dropwise via syringe at 23 ºC. The yellow solution was stirred for 20 min at 23 ºC. A solution of *o*-tolualdehyde (73.0 μL, 0.629 mmol, 1 eq) in THF (5 mL) was added dropwise to the stirring reaction mixture at 0 ºC over 1 min. The reaction was stirred overnight at 4 ºC and then was allowed to warm up to room temperature and stirred for another 4 h. The reaction was quenched with saturated aqueous ammonium chloride solution (20 mL) and the THF was removed *in vacuo*. The aqueous mixture was extracted with ethyl acetate (3 x 20 mL), the extracts were combined and washed with brine (2 x 20 mL) and the organic layer was dried over anhydrous sodium sulfate. The solution was filtered and concentrated. The crude mixture was purified by flash column chromatography (9:1 hexanes-ethyl acetate to 4:1) giving the product as a yellow oil in a 5:1 mixture with the (*E*)-olefin (by ^{1}H NMR, data annotated for (*Z*)-olefin) (185 mg, 99 % yield). **^{1}H NMR (500 MHz, $CDCl_3$) δ:** 7.20-7.13 (m, 4H), 7.01 (d, *J* = 12.5 Hz, 1H), 6.04 (d, *J* = 12.0 Hz, 1H), 5.25 (s, 1 H), 2.26 (s, 3H), 1.44 (s, 6H). **^{13}C NMR (125 MHz, $CDCl_3$) δ:** 163.9, 161.9, 140.4, 136.2, 135.6, 129.7, 128.7, 128.4, 125.3, 122.7, 106.3, 96.1, 25.4, 24.9, 20.1. **HRMS (ESI^+):** *m/z* calculated for $(C_{15}H_{16}O_3Na)^+$ 267.0997, found 267.0985



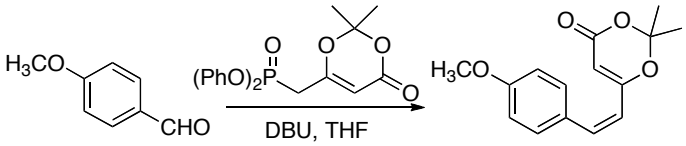
Trifluoroethyl 2-hydroxy-5-methyl-1-naphthoic acid ester. Trifluoroethanol (452 μL, 6.29 mmol, 10 eq) was dissolved in THF (5 mL), sodium hydride (60% dispersion in mineral oil, 90.5 mg, 3.77 mmol, 6 eq) was added to the solution. The reaction mixture was allowed to stir for 20 min at 23 ºC, resulting in a clear, orange solution. A solution of the (*Z*)-olefin (**12**) (154 mg, 0.629 mmol, 1 eq) in THF (5 mL) was added dropwise to the reaction mixture, which was then allowed to stir at 23 ºC for 50 min. The reaction was concentrated *in vacuo* to dryness and the resulting residue containing trifluoroethyl ester was taken up in acetic acid (3 mL). Manganese (III) acetate trihydrate (337 mg, 1.26 mmol, 2 eq) was added and the solution was stirred at 23 ºC for 20 h. The reaction was quenched with water (30 mL) and extracted with ethyl acetate (3 x 25 mL). The organic extracts were combined and washed with saturated sodium bicarbonate solution (2 x 40 mL) and the organic layer was dried over anhydrous sodium sulfate. The solution was filtered and concentrated. The residue was purified by flash column chromatography (2:1 hexanes-ethyl acetate) giving cyclized product (92.5 mg, 52% yield). **^{1}H NMR (500 MHz, $CDCl_3$) δ:** 11.59 (s, 1H), 8.53 (d, *J*=9.0 Hz, 1 H), 8.20 (d, *J*=9.0 Hz, 1 H), 7.48 (dd, *J* = 9.0, 7 Hz, 1H), 7.25 (d, *J* = 6.5 Hz, 1H), 7.213 (d, *J* = 9.0 Hz, 1H), 4.87 (q, *J* = 8.5 Hz, 2H), 2.69 (s, 3H). **^{13}C NMR (125 MHz, $CDCl_3$) δ:** 170.6, 164.3, 135.3, 133.7,

131.9, 128.7, 127.7, 125.4, 123.4, 121.4, 118.6 (m), 104.1, 61.0 (m), 19.9. **HRMS (ESI$^-$):** *m/z* calculated for $(C_{14}H_{10}O_3F_3)^-$ 283.0582, found 283.0580

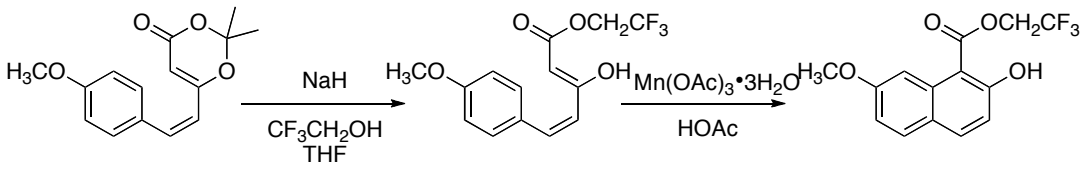


3.6

2-Hydroxy-5-methyl-1-naphthoic acid (3.6). To a stirring solution of naphthoic ester (92.5 mg, 0.0325 mmol, 1 eq) in 6 mL 2:1 THF-water, lithium hydroxide monohydrate (273 mg, 6.51 mmol, 20 eq) was added. The solution was stirred at 40 ºC for 22 h. The THF was removed *in vacuo* and the pH of the resulting slurry was adjusted with 1 M HCl until a pH of 2 was reached. The mixture was extracted with ethyl acetate (3 x 15 mL). The organic layer was washed with brine (2 x 10 mL), dried with sodium sulfate and then concentrated. The resulting residue was purified via flash column chromatography (1:2 hexanes-ethyl acetate with 1% acetic acid), yielding the naphthoic acid (**3.6**) (65.7 mg, 99% yield). **^{1}H NMR (400 MHz, $CDCl_3$) δ:** 12.1 (s, 1H), 8.78 (d, *J* = 8.8 Hz, 1H), 8.22 (d, *J* = 9.2 Hz, 1H), 7.50 (dd, *J* = 7.2, 8.8 Hz, 1H), 7.25 (d, *J* = 7.6 Hz, 1H), 7.22 (d, *J* = 9.6 Hz, 1H), 2.70 (s, 3H). **^{13}C NMR (100 MHz, $CDCl_3$) δ:** 175.6, 165.2, 135.4, 134.1, 132.7, 128.8, 127.9, 125.5, 124.1, 119.1, 104.1, 20.2. **HRMS (ESI$^-$):** *m/z* calculated for $(C_{12}H_9O_3)^-$ 201.0552, found 201.0558

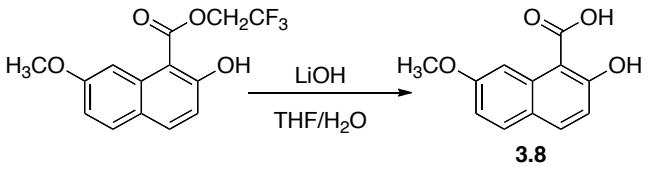


Methoxy substituted (*Z*)-olefin. To a stirring solution of phosphonate (500 mg, 1.34 mmol, 1.4 eq) in THF (5 mL), DBU (284 μL, 1.91 mmol, 2 eq) was added dropwise via syringe at 23 ºC. The yellow solution was stirred for 20 min at 23 ºC. A solution of 4-methoxy benzaldehyde (116 μL, 0.954 mmol, 1 eq) in THF (5 mL) was added dropwise to the stirring orange solution at 0 ºC over 1 min. The reaction was stirred overnight at 4 ºC. The reaction was quenched with saturated aqueous ammonium chloride solution (20 mL) and the THF was removed *in vacuo*. The aqueous mixture was extracted with ethyl acetate (3 x 30 mL) and the extracts were combined, washed with brine (2 x 20 mL), and dried over anhydrous sodium sulfate. The solution was filtered and concentrated. The crude mixture was purified by flash column chromatography (1:1 hexanes-ethyl acetate) giving product in a 1:1.5 inseparable mixture with the (*E*)-olefin (by ^{1}H NMR) (248 mg, 99 % yield). **^{1}H NMR (500 MHz, $CDCl_3$) δ:** 7.33 (d, *J* = 8.5 Hz, 2H), 6.85 (d, *J* = 9 Hz, 2H), 6.82 (d, *J* = 12.5 Hz, 1H), 5.83 (d, *J* = 12.5 Hz, 1H), 5.37 (s, 1H), 3.83 (s, 3H), 1.62 (s, 6H). **^{13}C NMR (125 MHz, CDCl3) δ:** 164.5, 162.5, 160.5, 131.5, 129.8, 128.1, 120.0, 113.8, 96.1, 55.8, 25.6. **HRMS (ESI^+):** *m/z* calculated for $(C_{15}H_{16}O_4Na)^+$ 283.0946, found 283.0950



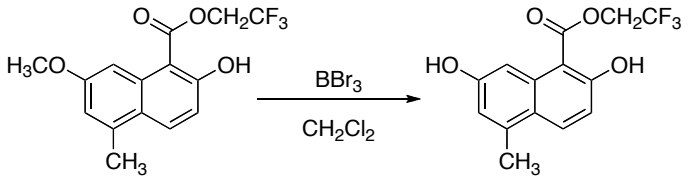
Trifluoroethyl 2-hydroxy-7-methoxy-1-naphthoic acid ester. Trifluoroethanol (277 μL, 3.85 mmol, 10 eq) was dissolved in THF (5 mL), and sodium hydride (60% dispersion in mineral oil, 55.4 mg, 2.31 mmol, 6 eq) was added. The solution was stirred for 20 min at 23 ºC, turning clear and orange. A solution of the (*Z*)-olefin (mixture with (*E*)-olefin) (100 mg, 0.384 mmol, 1 eq) in THF (5 mL) was added dropwise to the reaction and stirred at 23 ºC for 50 min. The reaction was concentrated *in vacuo* and the resulting residue containing trifluoroethyl ester was taken up in acetic acid (3 mL). Manganese (III) acetate trihydrate (337 mg, 1.26 mmol, 2 eq) was added and the solution was stirred at 23 ºC for 20 h. The reaction was quenched with water (30 mL) and extracted with ethyl acetate (3 x 25 mL). The organic extracts were combined and washed with saturated sodium bicarbonate solution (2 x 40 mL) and the organic layer was dried over anhydrous sodium sulfate. The solution was filtered, concentrated, and the brown residue was purified by flash column chromatography (gradient 3:1 to 1:1 hexanes-ethyl acetate) giving product (82.0 mg, 71 % yield). **^{1}H NMR (500 MHz, $CDCl_3$) δ:** 8.12 (d, *J* = 2.5 Hz, 1H), 7.86 (d, *J* = 9 Hz, 1H), 7.65 (d, *J* = 9 Hz, 1H), 7.04 (dd, *J* = 2.5, 9.0 Hz, 1H) 7.02 (d, *J* = 9.0 Hz, 1H), 4.87 (q, *J* = 8.5 Hz, 2H), 3.92 (s, 3H)**.** **^{13}C NMR (125 MHz, $CDCl_3$) δ:** 171.1, 166.3, 160.8, 138.2, 133.6, 131.0, 124.1, 116.5,

116.4, 105.2, 101.5, 61.8 (m), 55.4**. HRMS (ESI$^-$):** *m/z* calculated for $(C_{14}H_{10}O_4F_3)^-$ 299.0531, found 299.539.

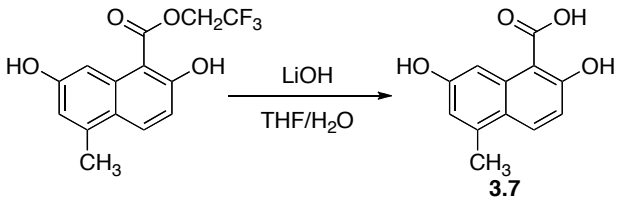


3.8

2-Hydroxy-7-methoxy-1-naphthoic acid (3.8). To a stirring solution of naphthoic ester (82.0 mg, 0.273 mmol, 1 eq) in 6 mL 2:1 THF-water, lithium hydroxide monohydrate (229 mg, 5.46 mmol, 20 eq) was added. The solution was stirred at 40 ºC for 22 h. The THF was removed *in vacuo* and the pH of the resulting slurry was adjusted with 1 M HCl until a pH of 2 was reached. The mixture was extracted with ethyl acetate (3 x 15 mL). The organic layer was washed with brine (2 x 10 mL), dried with sodium sulfate and then concentrated. The resulting residue was purified via flash column chromatography (2:1 hexanes-ethyl acetate with 1% acetic acid), yielding the naphthoic acid (**3.8**) (39.9 mg, 61 % yield). **^{1}H NMR (400 MHz, $CDCl_3$) δ:** 12.01 (s, 1H), 8.34 (s, 1H), 7.88 (d, *J* = 9.2 Hz, 1H), 7.68 (d, *J* = 8.4 Hz, 1H), 7.04 (t, *J* = 9.2 Hz , 1H), 3.97 (s, 3H). **^{13}C NMR (125 MHz, $CDCl_3$) δ:**175.9, 166.7, 160.5, 138.1, 134.1, 130.8, 127.3, 124.0, 116.7, 115.3, 106.4, 55.6. **HRMS (ESI$^-$):** *m/z* calculated for (C12H9O4)$^-$ 217.0501, found 217.0494.

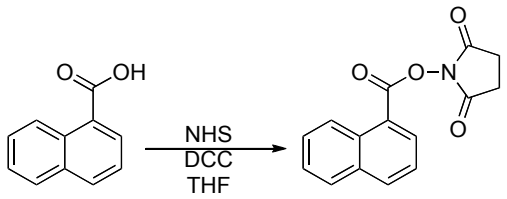


2,7-Dihydroxy-5-methyl-1-naphthoic ester. Dissolved naphthoic ester (24.8 mg, 0.0832 mmol, 1 eq) in 3 mL dichloromethane and cooled to -10 ºC while stirring under nitrogen. A 1 M solution of boron tribromide in dichloromethane (83.2 μL, 0.0832 mmol, 1 eq) was added dropwise, turning the reaction from clear to yellow. The reaction was allowed to stir at 23 ºC for 20 h. The reaction was quenched with 5 mL of methanol and then concentrated to dryness. Dissolved residue in methanol (2 x 5 mL) and concentrated to dryness with each addition. The brown residue was purified by flash column chromatography (2:1 hexanes-ether with 1% acetic acid) yielding the demethylated product (17.4 mg, 90% yield). **^{1}H NMR (400 MHz, $CDCl_3$) δ:** 11.70 (s, 1H), 8.09 (d, *J* = 9.6 Hz, 1H), 7.92 (s, 1H), 7.04 (d, *J* = 9.2 Hz, 1H), 6.88 (s, 1H), 4.98 (s, 1H), 4.87 (q, *J* = 8.4 Hz, 2H), 2.64 (s, 3H). **^{13}C NMR (100 MHz, CD_3OD) δ:** 170.5, 161.3, 158.5, 138.3, 135.5, 132.5, 126.3, 123.7, 117.8, 117.6, 115.9, 115.6, 106.7, 62.0 (m), 20.9. **HRMS (ESI^-):** *m/z* calculated for $(C_{14}H_{10}O_4F_3)^-$ 299.0531, found 299.0538.



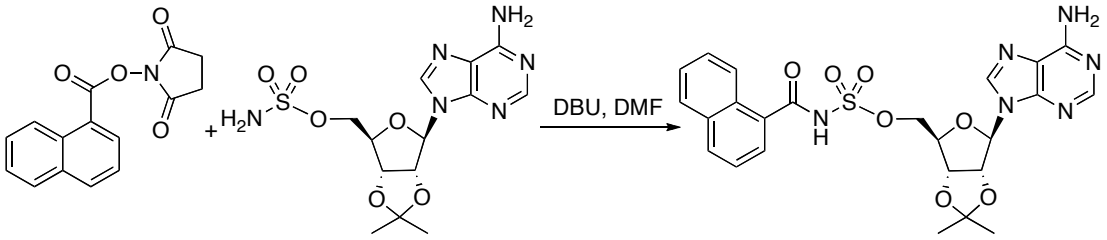
3.7

2,7-Dihydroxy-5-methyl-1-naphthoic acid (3.7). To a stirring solution of naphthoic ester (17.4 mg, 0.075 mmol, 1 eq) in 4 mL 1:1 THF-water, lithium hydroxide monohydrate (0.063 g, 1.45 mmol, 20 eq) was added. The yellow solution was heated to 40 ºC with stirring overnight. The THF was removed *in vacuo* and the pH of the resulting slurry was adjusted with 1 M HCl to pH 2. The mixture was extracted with dichloromethane (3 x 10 mL). The organic layer was washed with brine (1 x 30 mL), dried with sodium sulfate and then concentrated. The resulting residue was purified via flash column chromatography (1:1 hexanes-ether with 1% acetic acid), resulting in the naphthoic acid (**3.7**) (4.4 mg, 99% yield). **^{1}H NMR (400 MHz, CD_3OD) δ:** 8.15 (d, *J* = 2.4 Hz, 1H), 8.04 (d, *J* = 9.2 Hz, 1H), 6.92 (d, *J* = 9.2 Hz, 1H), 6.80 (d, *J* = 0.8 Hz, 1H), 2.59 (s, 3H). **^{13}C NMR (100 MHz, CD_3OD) δ:** 165.0, 162.8, 158.2, 137.9, 136.6, 132.5, 123.8, 117.1, 116.2, 108.2, 106.7, 20.2. **HRMS (ESI^-):** *m/z* calculated for $(C_{12}H_9O_4)^-$ 217.0501, found 217.0506.



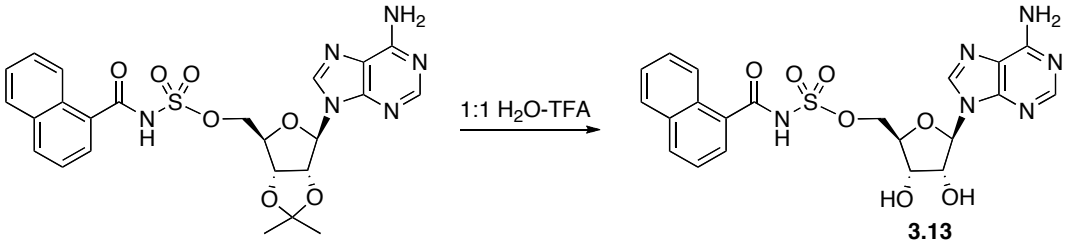
3.12

Naphthoyl-NHS. To a round bottom flask equipped with a stir bar, **3.12** (251 mg, 1.46 mmol, 1 eq) and *N*-hydroxysuccinimide (169 mg, 1.46 mmol, 1 eq) in 5 mL THF, *N*,*N*'-dicyclohexylcarbodiimide (300.9 mg, 1.458 mmol, 1 eq) was added. The clear solution was allowed stirred at room temperature until a fine white precipitate formed (30 min). The crude slurry was filtered through a pad of celite and concentrated under reduced pressure. The residue was resuspended in ethyl acetate and filtered again through celite. The resulting crude product was concentrated under reduced pressure and purified via flash column chromatography (3:1 hexanes-ethyl acetate to 1:1 then to 100% ethyl acetate), resulting in activated 1-naphthoyl-*N*-hydroxysuccinimide ester (390 mg, 99% yield). **^{1}H NMR (400 MHz, $CDCl_3$) δ:** 8.79 (d, *J* = 8.4 Hz, 1H), 8.44 (d, *J* = 8.4 Hz, 1H), 8.13 (d, *J* = 8.0 Hz, 1H), 7.89 (d, *J* = 8.0 Hz, 1H), 7.65 (m, 1H) 7.56 (m, 3H), 2.93 (s, 4H).



Acetonide protected naphthoyl sulfamoyl adenosine. Acetonide protected sulfamoyl adenosine (128 mg, 0.332 mmol, 1 eq, synthesized by published literature procedure

(*24*)) was dissolved in 3 mL dry DMF. 1,8-Diazabicyclo[5.4.0]undec-7-ene (121 μL, 0.797 mmol, 2.4 eq) was added to clear solution. Upon addition of activated naphthoic acid (134 mg, 0.498 mmol, 1.5 eq), solution turned from clear to yellow. The reaction was stirred under nitrogen at room temperature. After 24 h, the crude reaction was concentrated to dryness with slight heating. The crude residue was purified via flash column chromatography (9:1 chloroform-methanol with 1% triethylamine (TEA)) resulting in a TEA salt (115 mg, 0.212 mmol, 53.8% yield). A second column was required to obtain pure product. **^{1}H NMR (400 MHz, CD_3OD) δ:** 8.54 (m, 2H), 8.18 (s, 1H), 7.79 (m, 2H), 7.77 (d, *J* = 7.2 Hz, 1H), 7.65 (m, 3H), 6.27 (d, *J* = 3.2 Hz, 1H), 5.42 (dd, *J* = 6.0, 3.2 Hz, 1H), 5.26 (dd, *J* = 6.0, 2.4 Hz, 1H), 4.62 (t, *J* = 2.4 Hz, 1H), 4.41 (d, *J* = 4.0 Hz, 2H), 1.62 (s, 3 H), 1.25 (s, 3H). **HRMS (ESI^+):** *m/z* calculated for $(C_{24}H_{24}NaN_6O_7S)^+$ 563.1319, found 563.1505.



3.13

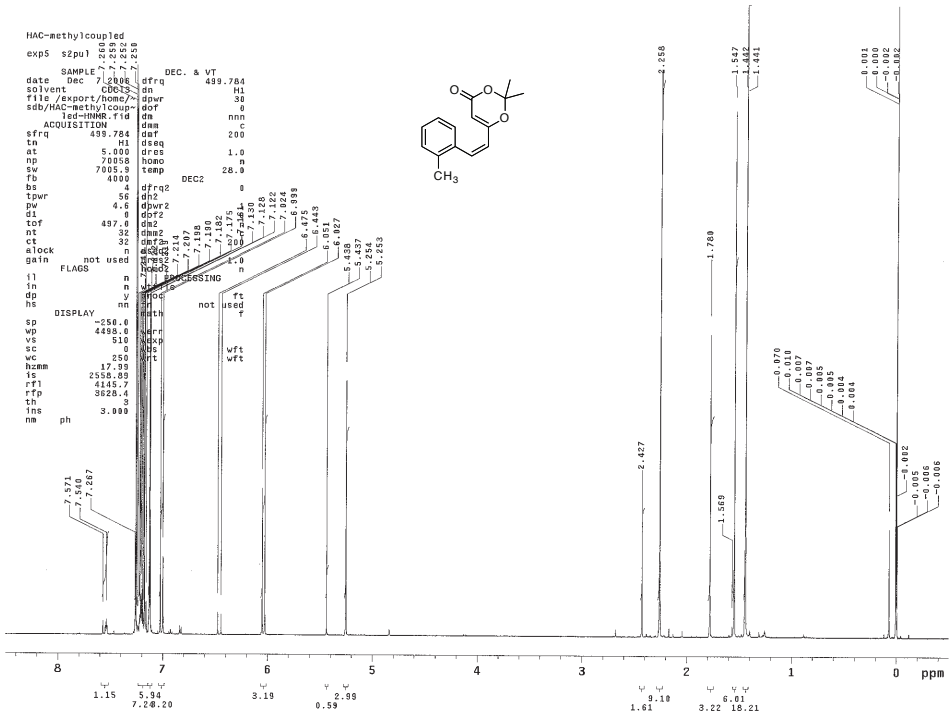
Naphthoyl sulfamoyl adenosine (3.13). Starting material (111 mg, 0.173 mmol, 1 eq) was dissolved in 4 mL 1:1 H_2O-TFA. Stirred orange slurry at room temperature for 2 h. Quenched with small amount of methanol, evaporated solvent *in vacuo*. Purified by flash column chromatography (4:1 dichloromethane-methanol to 3:1), yielding **3.13** in 99% yield. **^{1}H NMR (400 MHz, CD_3OD) δ:** 8.41 (s, 1H) 8.40 (d, 1H), 8.12 (s, 1H), 7.79 (m, 3H), 7.37 (m, 3H) 6.06 (d, *J* = 5.2 Hz, 1H), 4.68 (t, *J* = 5.2 Hz, 1H), 4.51 (m, 3H), 4.36

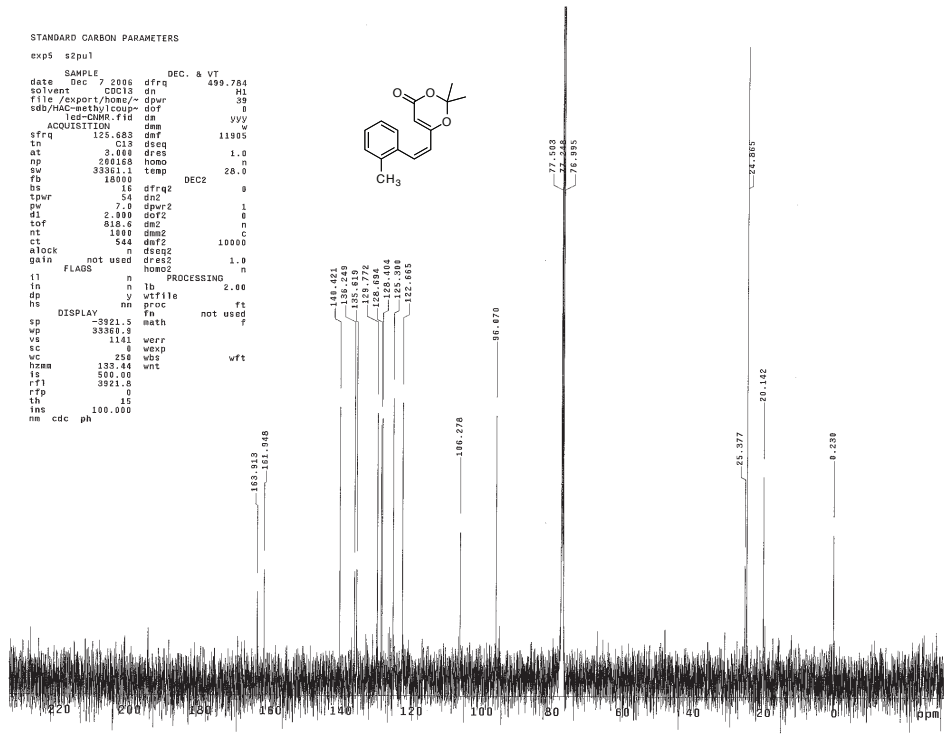
(d, 1H). **^{13}C NMR (100 MHz, CD_3OD) δ:** 175.8, 163.0, 156.9, 153.4, 150.7, 141.3, 137.0, 135.2, 131.7, 129.3, 128.2, 127.7, 127.0, 125.8, 120.3, 116.8, 89.9, 84.3, 75.9, 72.3, 70.4. **HRMS (ESI^+):** *m/z* calculated for ($C_{21}H_{20}N_6NaO_7S$) 523.1012, found 523.1031.

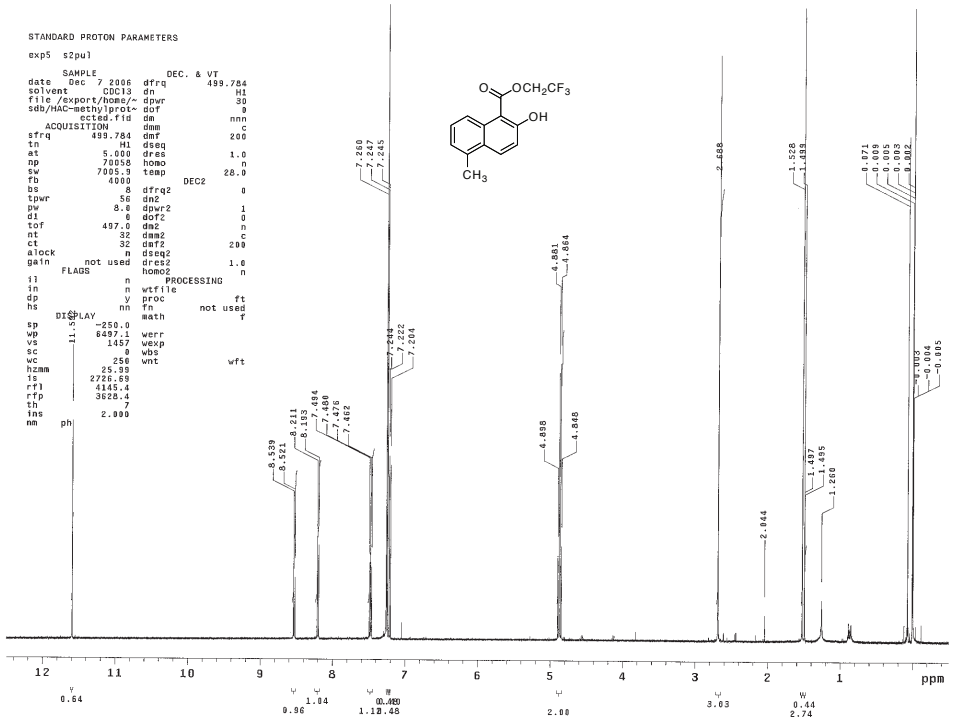
Crystallization attempts

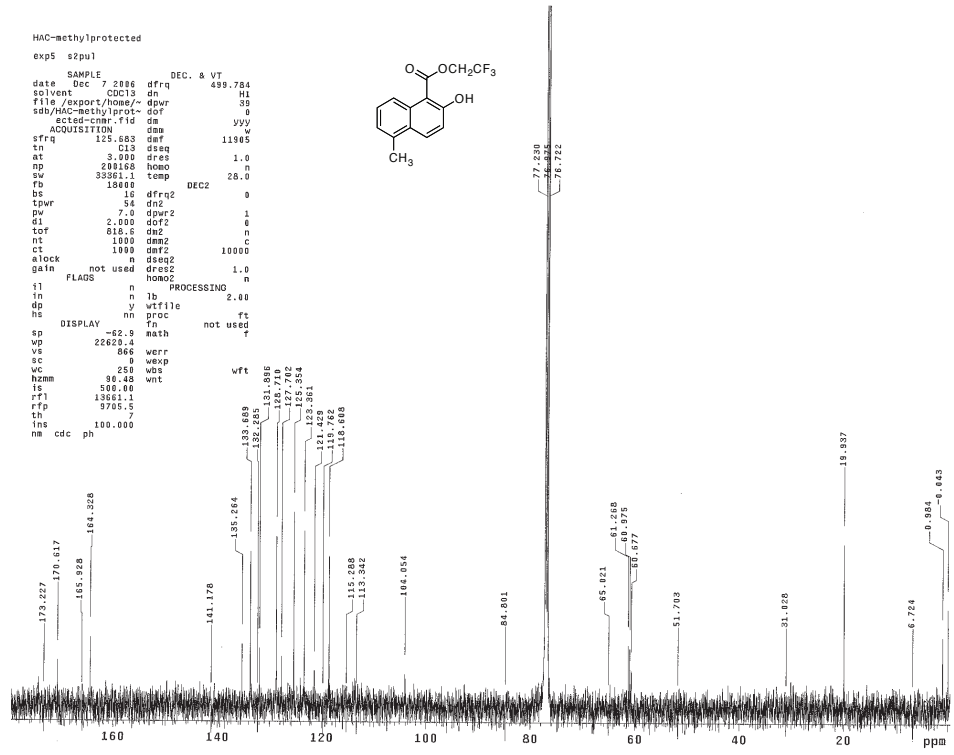
NcsB2 was screened for crystallization using cleaved and uncleaved protein with Hampton Crystal Screens (Hampton Research, Aliso Viejo, CA) as well as Wizard Crystal Screens I and II (Emerald Biosystems, Bainbridge Island, WA) at 4 and 20 ºC. Next, cleaved NcsB2 was screened for crystallization with 1.6 mM **3.3**, with and without 5 mM ATP, 2 mM $MgCl_2$ at 4 and 20 ºC. Truncated constructs of NcsB2 were screened with the Hampton Crystal Screens without substrate at 4 and 20 ºC.

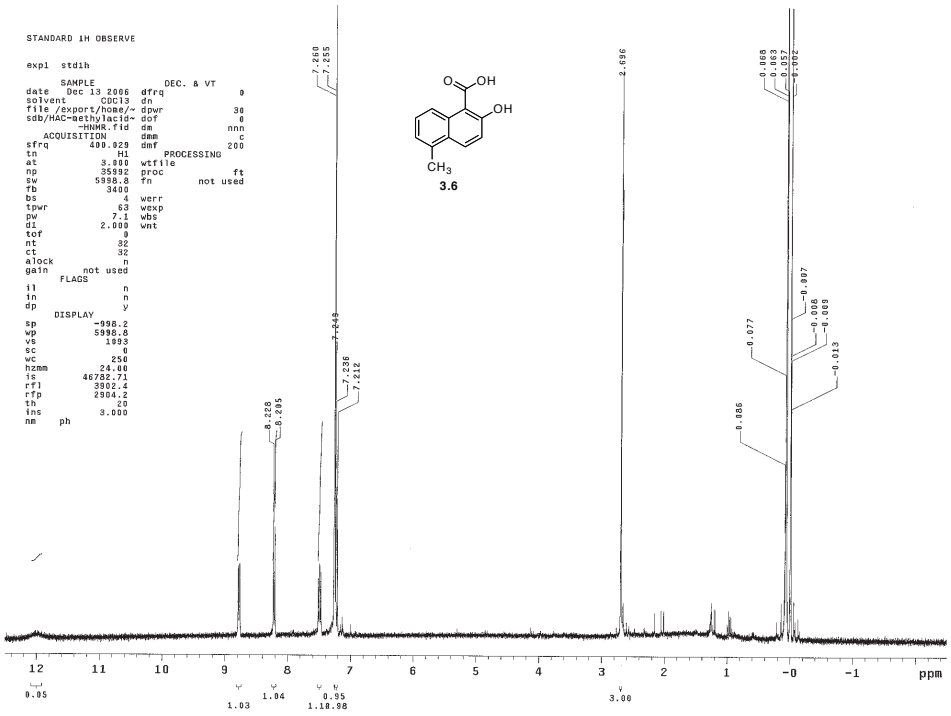
NcsB2 was also screened with both sulfamoyl adenosine inhibitors (**3.13** and **3.14**). A 10 mM stock solution of inhibitor was made up in water. NcsB2 (10 mg/mL) was incubated on ice for 1 h with 1.61 mM naphthoyl sulfamoyl adenosine. Two Hampton Crystal Screens were set up at 4 ºC and 20 ºC. No crystal hits resulted.

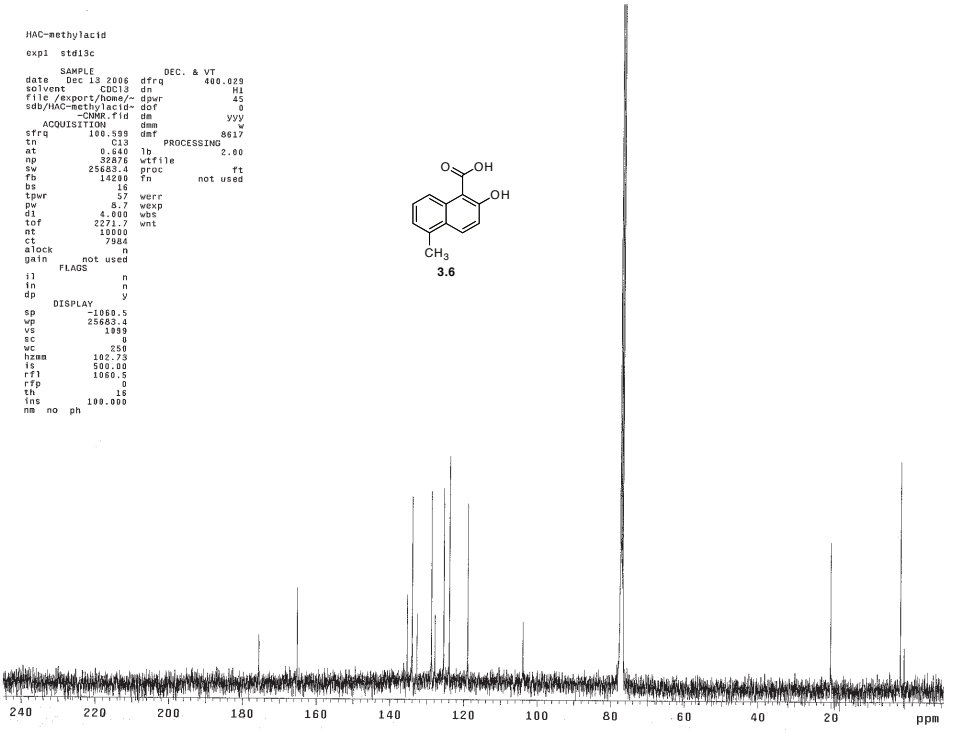

HAC-methylcoupled
exp5 s2pul
SAMPLE
date Dec 7 2006
solvent CDCl3
file /export/home/~
sdb/HAC-methylcoup~
led-HNMR.fid
ACQUISITION
sfrq 499.784
tn H1
at 5.000
np 70058
sw 7005.9
fb 4000
bs 4
tpwr 56
pw 4.6
d1 0
tof 497.0
nt 32
ct 32
alock n
gain not used
FLAGS
DISPLAY
sp -250.0
wp 4498.0
vs 510
sc 0
wc 250
hzmm 17.99
is 2558.89
rf1 4145.7
rfp 3628.4
th 3
ins 3.000
nm ph
DEC. & VT
dfrq 499.784
dn H1
dpwr 30
dof 0
dm nnn
dmm c
dmf 200
dseq
dres 1.0
homo n
temp 28.0
DEC2
dfrq2 0
dn2
dpwr2
dof2
dm2
dmm2
dmf2
dseq2
dres2 1.0
homo2 n
PROCESSING
ft
not used
math f
wft
wft
2.256
1.780
1.542
1.488
1.441
2.427
1.569
7.571
7.540
7.487
6.475
6.443
6.027
5.438
5.437
5.263
0.001
0.000
-0.002
-0.192
0.070
0.010
0.007
0.007
0.005
0.005
0.004
0.004
-0.082
-0.085
-0.006
-0.006
1.15
5.94
7.20
4.20
3.19
2.99
0.59
9.10
1.61
6.01
3.22
18.21
ppm

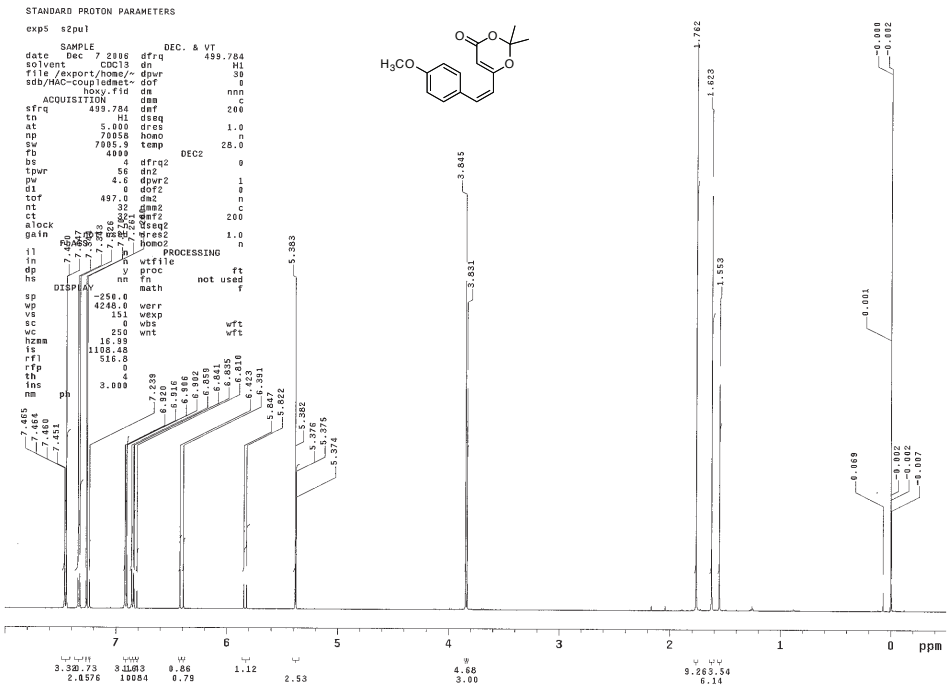

STANDARD CARBON PARAMETERS
exp5 s2pul
SAMPLE
date Dec 7 2006
solvent CDCl3
file /export/home/~
sdb/HAC-methylcoup~
led-CNMR.fid
ACQUISITION
sfrq 125.683
tn C13
at 3.088
np 200188
sw 33361.1
fb 18000
bs 16
tpwr 54
pw 7.0
d1 2.000
tof 818.6
nt 1000
ct 544
alock n
gain not used
FLAGS
DISPLAY
sp -3921.5
wp 33360.8
vs 1141
sc 0
wc 256
hzmm 133.44
is 500.00
rf1 3921.8
rfp 0
th 15
ins 100.000
nm cdc ph
DEC. & VT
dfrq 499.784
dn H1
dpwr 39
dof 0
dm yyy
dmm w
dmf 11905
dseq
dres 1.0
homo n
temp 28.0
DEC2
dfrq2 0
dn2
dpwr2 1
dof2 0
dm2 n
dmm2 c
dmf2 10000
dseq2
dres2 1.0
homo2 n
PROCESSING
lb 2.00
wtfile
proc ft
fn not used
math f
wexp
wbs wft
wnt
77.583
77.266
76.995
24.865
20.142
25.377
0.230
96.870
106.278
163.913
161.948
148.421
136.249
135.679
129.772
128.694
128.404
128.208
122.665
220
200
180
160
140
120
100
80
60
40
20
0
ppm

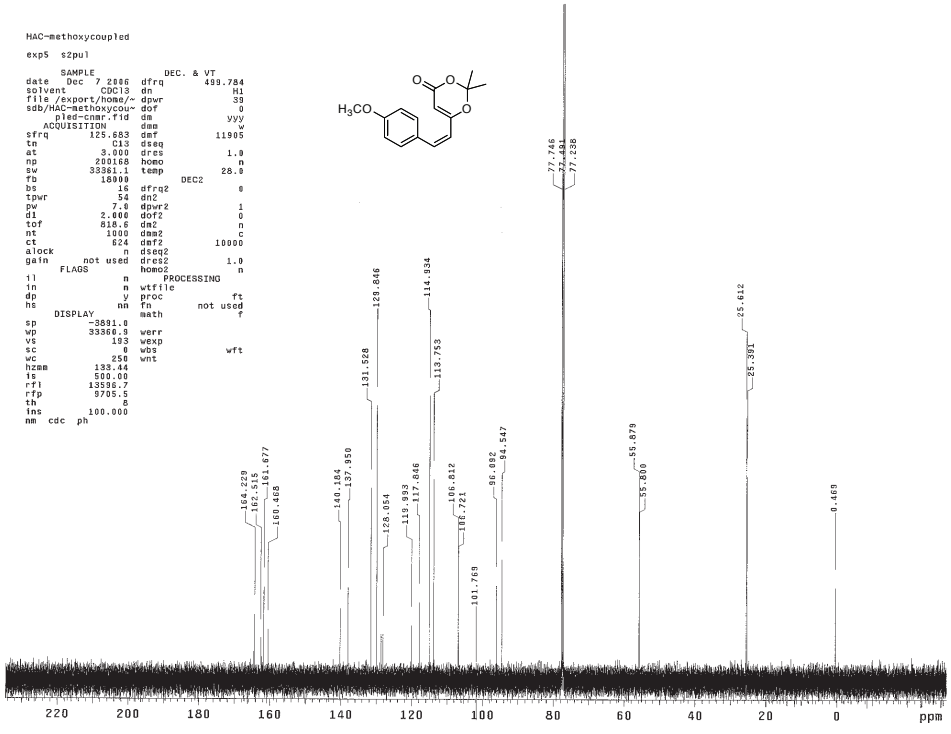

STANDARD PROTON PARAMETERS
exp5 s2pul
SAMPLE
date Dec 7 2006
solvent CDCl3
file /export/home/~
sdb/HAC-methylprot~
ected.fid
ACQUISITION
sfrq 499.784
tn H1
at 5.000
np 70058
sw 7005.9
fb 4000
bs 8
tpwr 56
pw 8.0
d1 0
tof 497.0
nt 32
ct 32
alock n
gain not used
FLAGS
il n
in n
dp y
hs nn
DISPLAY
sp -250.0
wp 6497.1
vs 1457
sc 0
wc 250
hzmm 25.99
is 2726.69
rfl 4145.4
rfp 3628.4
th 7
ins 2.000
nm ph
DEC. & VT
dfrq 499.784
dn H1
dpwr 30
dof 0
dm nnn
dmm c
dmf 200
dseq
dres 1.0
homo n
temp 28.0
DEC2
dfrq2 0
dn2
dpwr2 1
dof2 0
dm2 n
dmm2 n
dmf2 200
dseq2
dres2 1.0
homo2 n
PROCESSING
wtfile
proc ft
fn not used
math f
werr
wexp
wbs
wnt wft
O
OCH2CF3
OH
CH3
7.260
7.247
7.245
7.244
7.222
7.204
8.539
8.521
8.211
8.193
7.494
7.480
7.476
7.462
4.861
4.864
4.496
4.848
2.688
2.044
1.528
1.493
1.497
1.495
1.260
0.071
0.009
0.005
0.003
0.002
0.003
0.004
0.005
12
11
10
9
8
7
6
5
4
3
2
1
ppm
0.64
0.96
1.04
1.12
0.48
2.00
3.03
0.44
2.74

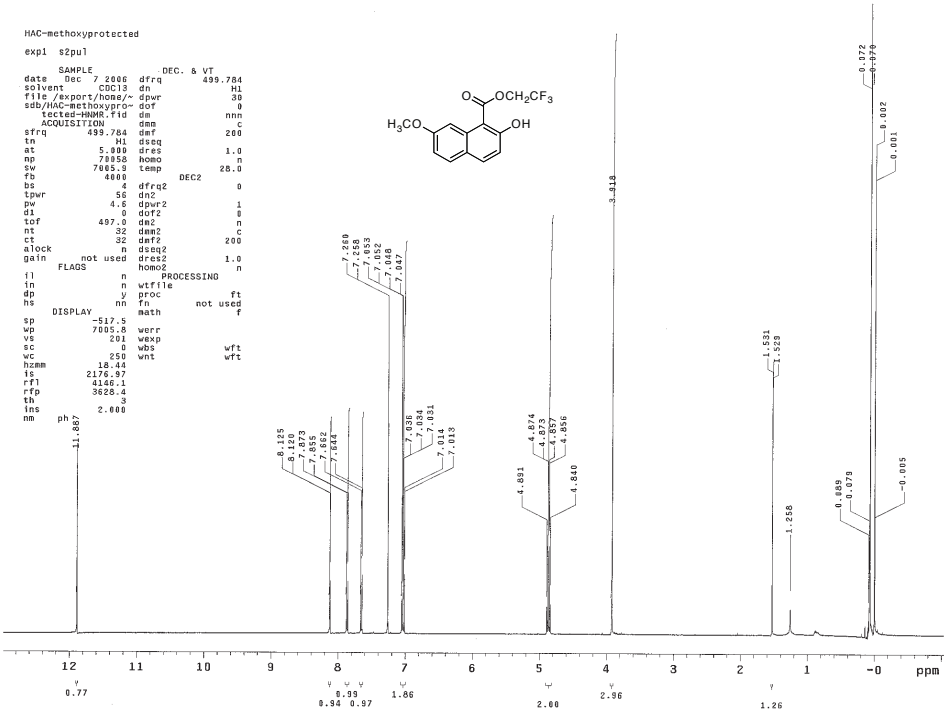

HAC-methylprotected
exp5 s2pul
SAMPLE
date Dec 7 2006
solvent CDCl3
file /export/home/~
sdb/HAC-methylprot~
ected-cnmr.fid
ACQUISITION
sfrq 125.683
tn C13
at 3.000
np 200168
sw 33361.1
fb 18600
bs 16
tpwr 54
pw 7.0
d1 2.000
tof 818.6
nt 1000
ct 1000
alock n
gain not used
FLAGS
il n
in n
dp y
hs nn
DISPLAY
sp -62.9
wp 22628.4
vs 866
sc 0
wc 250
hzmm 90.48
is 500.00
rfl 13661.1
rfp 9705.5
th 7
ins 100.000
nm cdc ph
DEC. & VT
dfrq 499.784
dn H1
dpwr 39
dof 0
dm yyy
dmm w
dmf 11905
dseq
dres 1.0
homo n
temp 28.0
DEC2
dfrq2 0
dn2
dpwr2 1
dof2 0
dm2 n
dmm2 n
dmf2 10000
dseq2
dres2 1.0
homo2 n
PROCESSING
lb 2.00
wtfile
proc ft
fn not used
math f
werr
wexp
wbs
wnt wft
O
OCH2CF3
OH
CH3
77.230
76.925
76.922
173.227
170.817
165.928
164.328
141.178
135.264
133.689
132.285
131.996
128.718
127.702
125.354
123.361
121.428
119.762
118.608
115.288
115.042
104.054
84.801
65.021
61.268
60.975
60.677
51.703
31.828
19.937
6.724
0.984
0.843
160
140
120
100
80
60
40
20
ppm

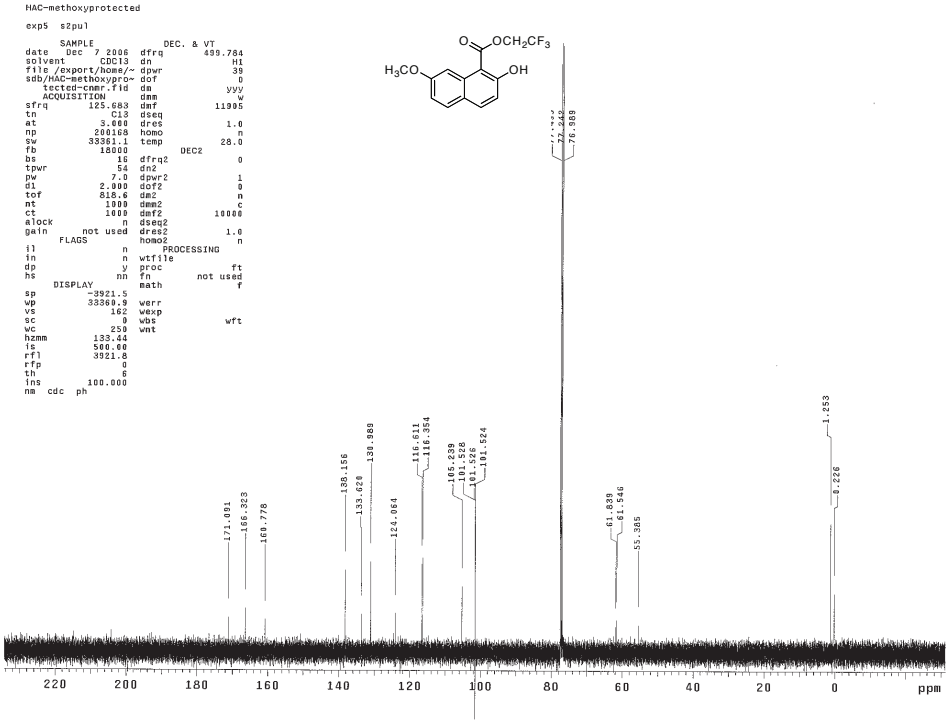

STANDARD 1H OBSERVE
expl stdih
SAMPLE
date Dec 13 2006
solvent CDCl3
file /export/home/~
sdb/HAC-methylacid~
-HNMR.fid
ACQUISITION
sfrq 400.029
tn H1
at 3.998
np 35992
sw 5998.8
fb 3400
bs 4
tpwr 63
pw 7.1
d1 2.000
tof 0
nt 32
ct 32
alock n
gain not used
FLAGS
il n
in n
dp y
DISPLAY
sp -998.2
wp 5998.8
vs 1993
sc 0
wc 250
hzmm 24.00
is 46782.71
rfl 3902.4
rfp 2904.2
th 20
ins 3.000
nm ph
DEC. & VT
dfrq 0
dn
dpwr 30
dof 0
dm nnn
dmm c
dmf 200
PROCESSING
wtfile
proc ft
fn not used
werr
wexp
wbs
wnt
7.260
7.255
7.248
7.236
7.212
8.226
8.205
2.696
0.068
0.063
0.057
0.002
0.007
0.006
0.005
0.077
0.013
0.086
OH
OH
CH3
3.6
12
11
10
9
8
7
6
5
4
3
2
1
-0
-1
ppm
0.05
1.03
1.04
0.95
1.18.98
3.00

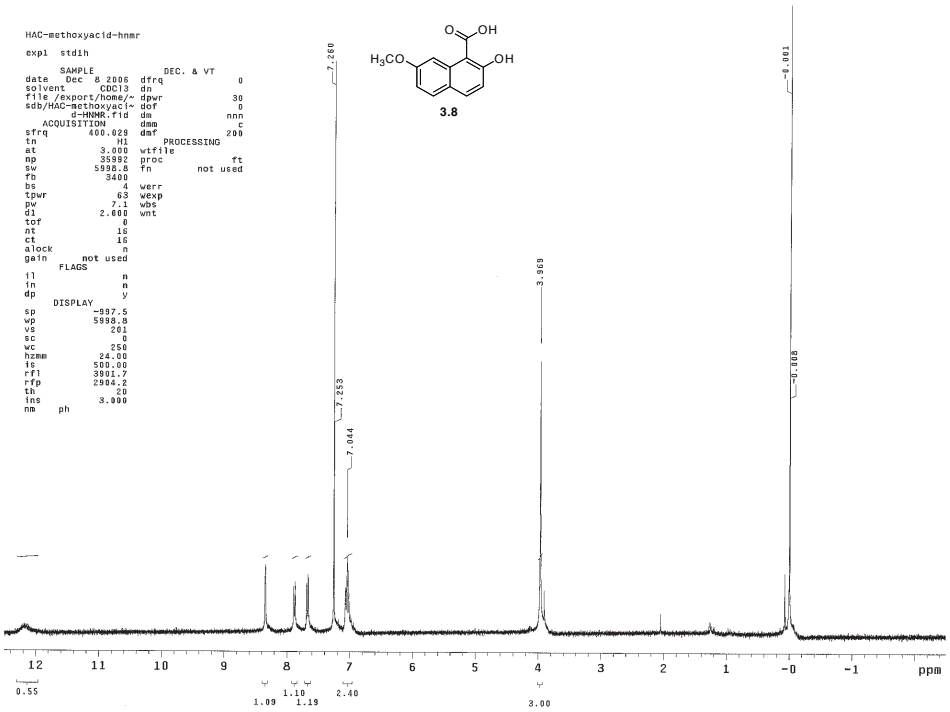

HAC-methylacid
expl stdi3c
SAMPLE
date Dec 13 2006
solvent CDCl3
file /export/home/~
sdb/HAC-methylacid~
-CNMR.fid
ACQUISITION
sfrq 100.599
tn C13
at 0.640
np 32876
sw 25683.4
fb 14200
bs 16
tpwr 57
pw 8.7
d1 4.000
tof 2271.7
nt 10000
ct 7984
alock n
gain not used
FLAGS
il n
in n
dp y
DISPLAY
sp -1060.5
wp 25683.4
vs 1089
sc 0
wc 250
hzmm 102.73
is 500.00
rfl 1060.5
rfp 0
th 16
ins 100.000
nm no ph
DEC. & VT
dfrq 400.029
dn H1
dpwr 45
dof 0
dm yyy
dmm w
dmf 8617
PROCESSING
lb 2.00
wtfile
proc ft
fn not used
werr
wexp
wbs
wnt
OH
OH
CH3
3.6
240
220
200
180
160
140
120
100
80
60
40
20
ppm

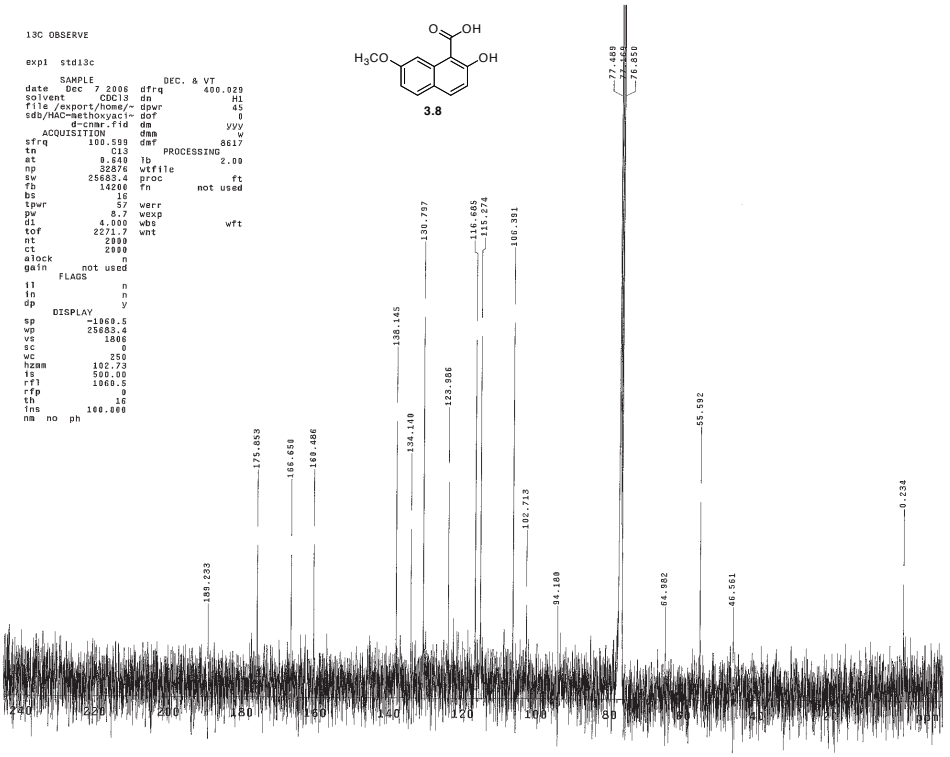

STANDARD PROTON PARAMETERS
expS s2pul
SAMPLE
DEC. & VT
PROCESSING
DISPLAY
ppm

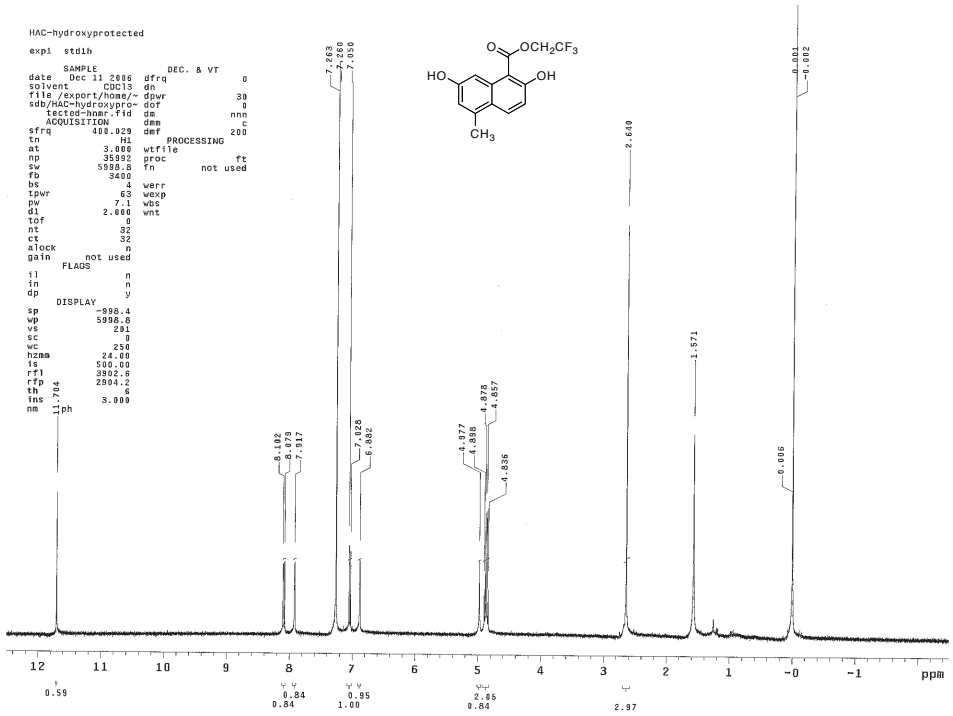

HAC-methoxycoupled
expS s2pul
SAMPLE
DEC. & VT
PROCESSING
DISPLAY
ppm

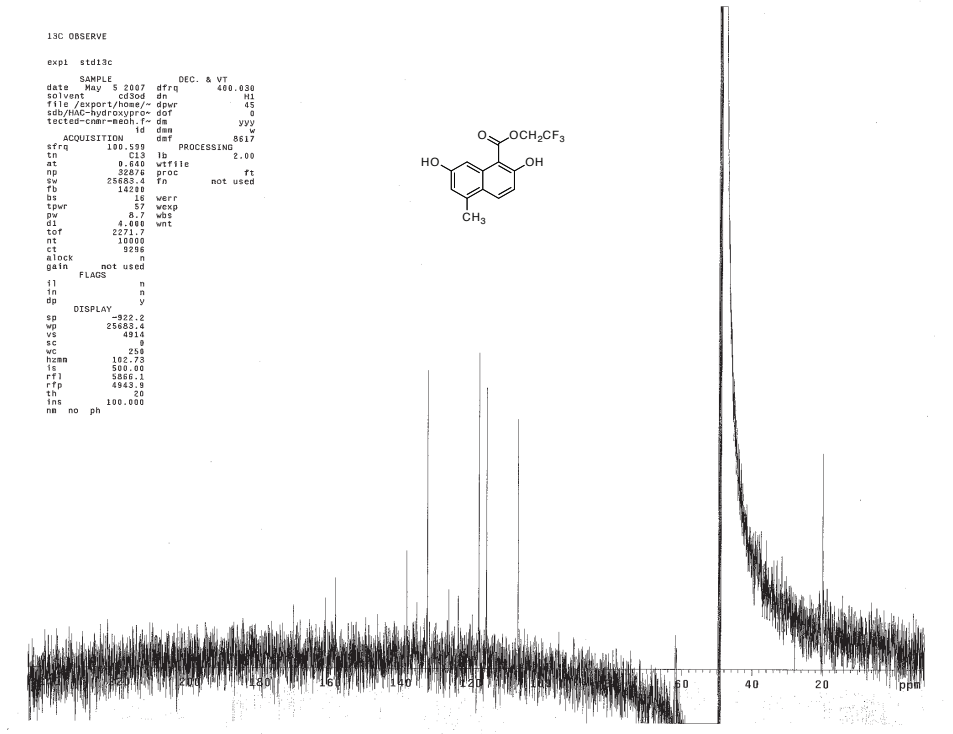

HAC-methoxyprotected
exp1 s2pul
SAMPLE
date Dec 7 2006
solvent CDCl3
file /export/home/~
sdb/HAC-methoxypro~
tected-HNMR.fid
ACQUISITION
sfrq 499.784
tn H1
at 5.000
np 78958
sw 7085.9
fb 4000
bs 4
tpwr 56
pw 4.6
d1 0
tof 497.0
nt 32
ct 32
alock n
gain not used
FLAGS
il n
in n
dp y
hs nn
DISPLAY
sp -517.5
wp 7005.8
vs 201
sc 0
wc 250
hzmm 14.44
is 2176.97
rfl 4146.1
rfp 3628.4
th 3
ins 2.000
nm ph
DEC. & VT
dfrq 499.784
dn H1
dpwr 38
dof 0
dm nnn
dmm c
dmf 200
dseq
dres 1.0
homo n
temp 28.0
DEC2
dfrq2 0
dn2
dpwr2 1
dof2 0
dm2 n
dmm2 c
dmf2 200
dseq2
dres2 1.0
homo2 n
PROCESSING
wtfile
proc ft
fn not used
math f
werr
wexp
wbs wft
wnt wft
OCH2CF3
H3CO
OH
11.867
8.125
8.120
7.873
7.855
7.682
7.644
7.269
7.258
7.255
7.052
7.048
7.047
7.036
7.034
7.031
7.014
7.013
4.891
4.874
4.873
4.857
4.856
4.840
3.918
1.531
1.529
1.258
0.372
0.370
0.002
0.001
0.005
0.889
0.079
ppm
0.77
0.99
0.94 0.97
1.86
2.00
2.96
1.26

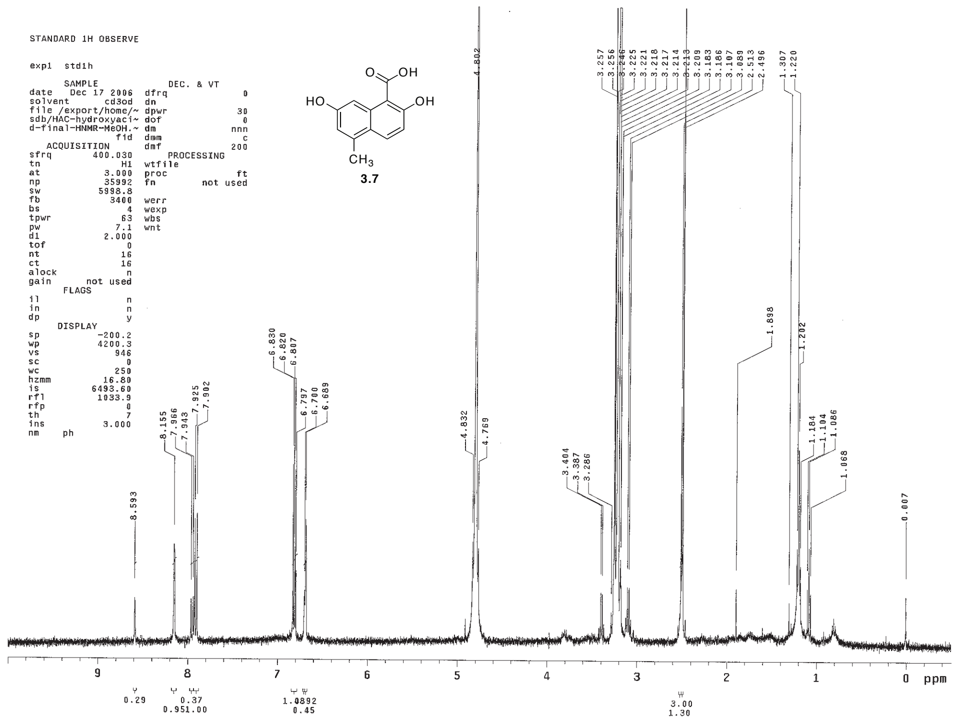

HAC-methoxyprotected
exp5 s2pul
SAMPLE
date Dec 7 2006
solvent CDCl3
file /export/home/~
sdb/HAC-methoxypro~
tected-cmnr.fid
ACQUISITION
sfrq 125.683
tn C13
at 3.000
np 200168
sw 33361.1
fb 18000
bs 16
tpwr 54
pw 7.0
d1 2.000
tof 818.6
nt 1000
ct 1000
alock n
gain not used
FLAGS
il n
in n
dp y
hs nn
DISPLAY
sp -3921.5
wp 33360.9
vs 162
sc 0
wc 250
hzmm 133.44
is 500.00
rfl 3921.8
rfp 0
th 6
ins 100.000
nm cdc ph
DEC. & VT
dfrq 499.784
dn H1
dpwr 38
dof 0
dm yyy
dmm w
dmf 11905
dseq
dres 1.0
homo n
temp 28.0
DEC2
dfrq2 0
dn2
dpwr2 1
dof2 0
dm2 n
dmm2 c
dmf2 10000
dseq2
dres2 1.0
homo2 n
PROCESSING
wtfile
proc ft
fn not used
math f
werr
wexp
wbs wft
wnt
OCH2CF3
H3CO
OH
77.242
76.989
171.091
166.323
160.778
135.156
133.620
130.988
124.064
116.611
116.354
105.239
101.528
101.526
101.524
61.838
61.546
55.385
1.253
0.228
ppm

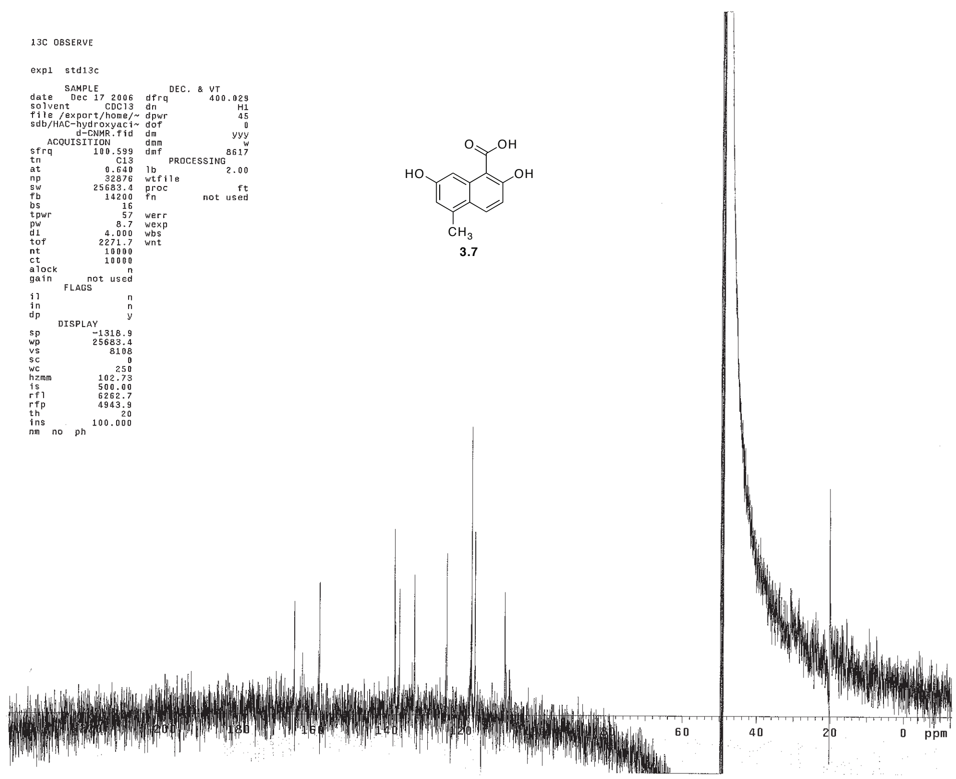

HAC-methoxyacid-hnmr
expl stdih
SAMPLE
date Dec 8 2006
solvent CDCl3
file /export/home/~
sdb/HAC-methoxyaci~
d-HNMR.fid
ACQUISITION
sfrq 400.028
tn H1
at 3.000
np 35992
sw 5998.8
fb 3400
bs 4
tpwr 63
pw 7.1
d1 2.000
tof 0
nt 16
ct 16
alock n
gain not used
FLAGS
il n
in n
dp y
DISPLAY
sp -997.5
wp 5998.8
vs 201
sc 0
wc 250
hzmm 24.00
is 500.00
rfl 3901.7
rfp 2904.2
th 20
ins 3.000
nm ph
DEC. & VT
dfrq 0
dn
dpwr 30
dof 0
dm nnn
dmm c
dmf 200
PROCESSING
wtfile
proc ft
fn not used
werr
wexp
wbs
wnt
O
OH
H3CO
OH
3.8
7.260
7.253
7.044
3.869
0.001
0.000
12
11
10
9
8
7
6
5
4
3
2
1
-0
-1
ppm
0.55
1.09
1.10
1.19
2.40
3.00

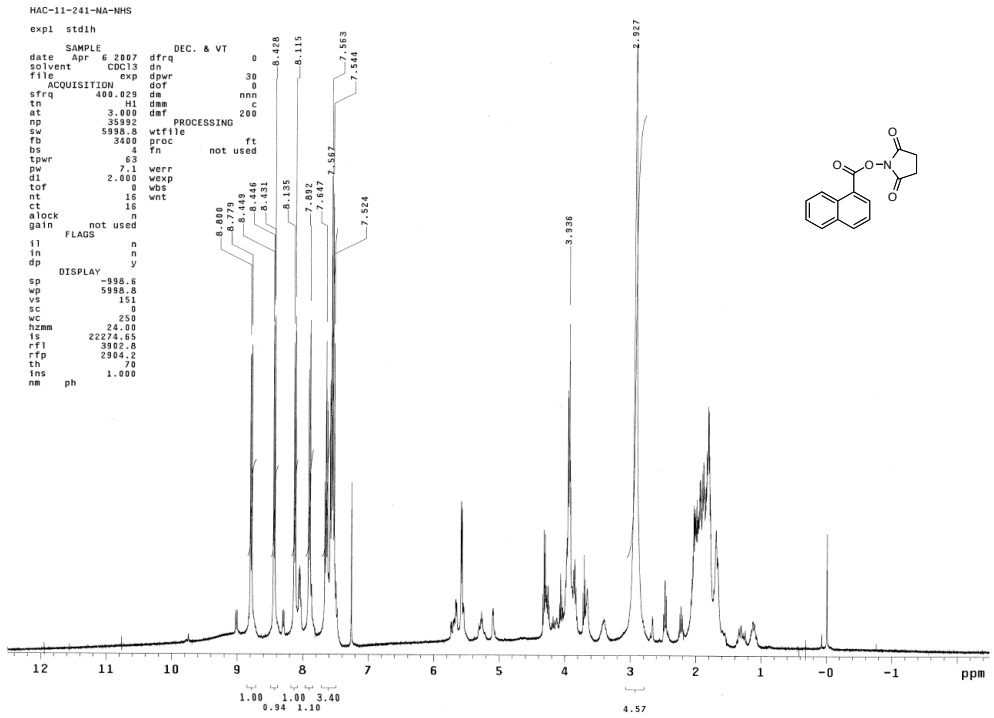

13C OBSERVE
expl stdi3c
SAMPLE
date Dec 7 2006
solvent CDCl3
file /export/home/~
sdb/HAC-methoxyaci~
d-cnmr.fid
ACQUISITION
sfrq 100.599
tn C13
at 0.640
np 32876
sw 25683.4
fb 14200
bs 16
tpwr 57
pw 8.7
d1 4.000
tof 2271.7
nt 2090
ct 2090
alock n
gain not used
FLAGS
il n
in n
dp y
DISPLAY
sp -1069.5
wp 25683.4
vs 1806
sc 0
wc 250
hzmm 102.73
is 500.00
rfl 1060.5
rfp 0
th 16
ins 100.000
nm no ph
DEC. & VT
dfrq 400.029
dn H1
dpwr 45
dof 0
dm yyy
dmm w
dmf 8617
PROCESSING
lb 2.00
wtfile
proc ft
fn not used
wexp
wbs wft
wnt
O
OH
H3CO
OH
3.8
77.488
77.168
76.850
175.853
168.650
160.486
138.145
134.140
130.797
123.986
116.685
115.274
106.391
102.713
94.180
85.233
64.982
55.592
46.561
0.234
240
220
200
180
160
140
120
100
80
60
40
20
0
ppm

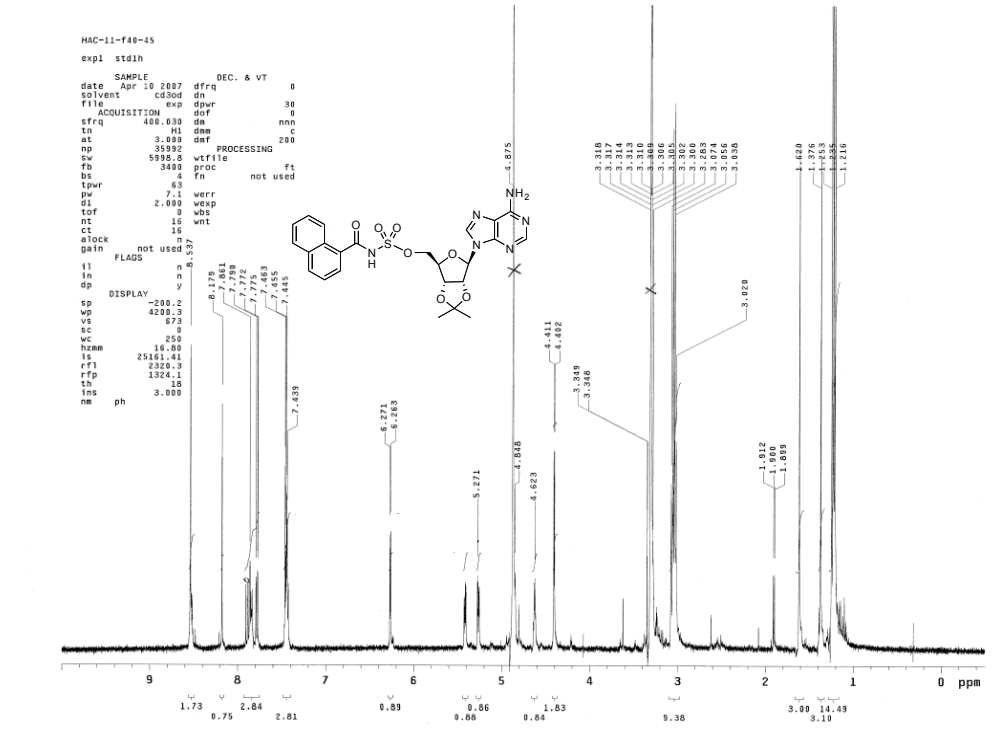

HAC-hydroxyprotected
expl std1h
SAMPLE
date Dec 11 2006
solvent CDCl3
file /export/home/~
sdb/HAC-hydroxypro~
tected-hnmr.fid
ACQUISITION
sfrq 400.029
tn H1
at 3.000
np 35992
sw 5998.8
fb 3400
bs 4
tpwr 63
pw 7.1
d1 2.000
tof 0
nt 32
ct 32
alock n
gain not used
FLAGS
il n
in n
dp y
DISPLAY
sp -998.4
wp 5998.8
vs 291
sc 0
wc 250
hzmm 24.00
is 500.00
rfl 3902.8
rfp 2904.2
th 6
ins 3.000
nm ph
DEC. & VT
dfrq 0
dn
dpwr 30
dof 0
dm nnn
dmm c
dmf 200
PROCESSING
wtfile
proc ft
fn not used
werr
wexp
wbs
wnt
7.263
7.260
7.050
8.102
8.079
7.917
7.028
6.882
4.977
4.898
4.878
4.857
4.836
2.640
1.571
0.011
-0.002
0.006
HO
OH
OCH2CF3
CH3
12
11
10
9
8
7
6
5
4
3
2
1
-0
-1
ppm
0.59
0.84
0.84
0.95
1.00
2.05
0.84
2.97

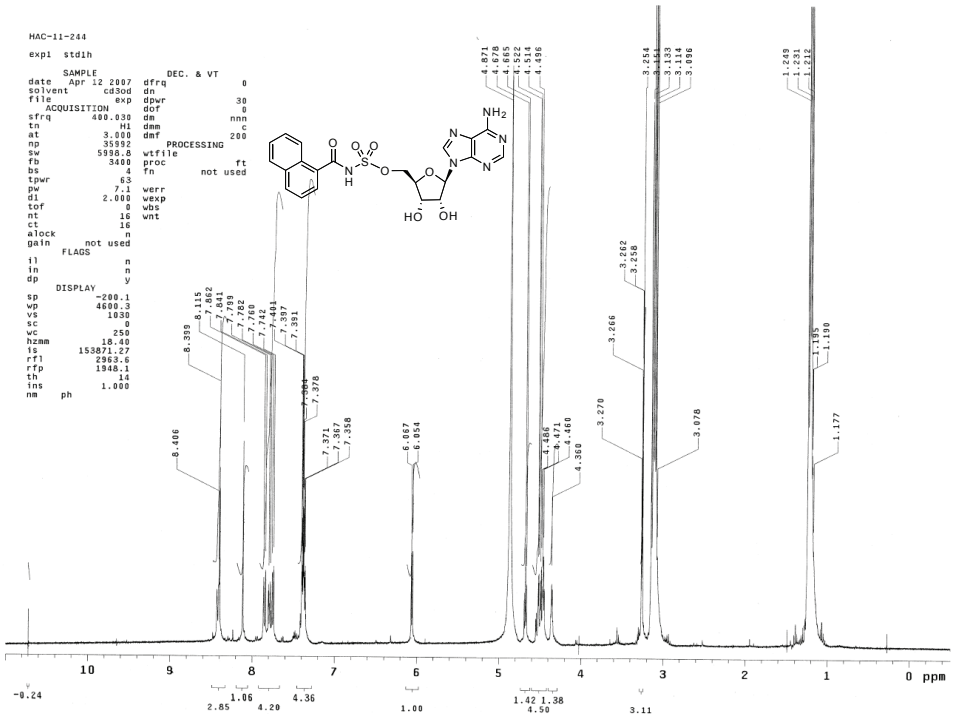

13C OBSERVE
expl std13c
SAMPLE
date May 5 2007
solvent cd3od
file /export/home/~
sdb/HAC-hydroxypro~
tected-cmhr-meoh.f~
id
ACQUISITION
sfrq 100.599
tn C13
at 0.640
np 32076
sw 25683.4
fb 14280
bs 16
tpwr 57
pw 8.7
d1 4.000
tof 2271.7
nt 10000
ct 9296
alock n
gain not used
FLAGS
il n
in n
dp y
DISPLAY
sp -922.2
wp 25683.4
vs 4914
sc 0
wc 250
hzmm 102.73
is 500.00
rfl 5856.1
rfp 4943.8
th 20
ins 100.000
nm no ph
DEC. & VT
dfrq 400.030
dn H1
dpwr 45
dof 0
dm yyy
dmm w
dmf 8617
PROCESSING
lb 2.00
wtfile
proc ft
fn not used
werr
wexp
wbs
wnt
HO
OH
OCH2CF3
CH3
200
180
160
140
120
100
80
60
40
20
ppm

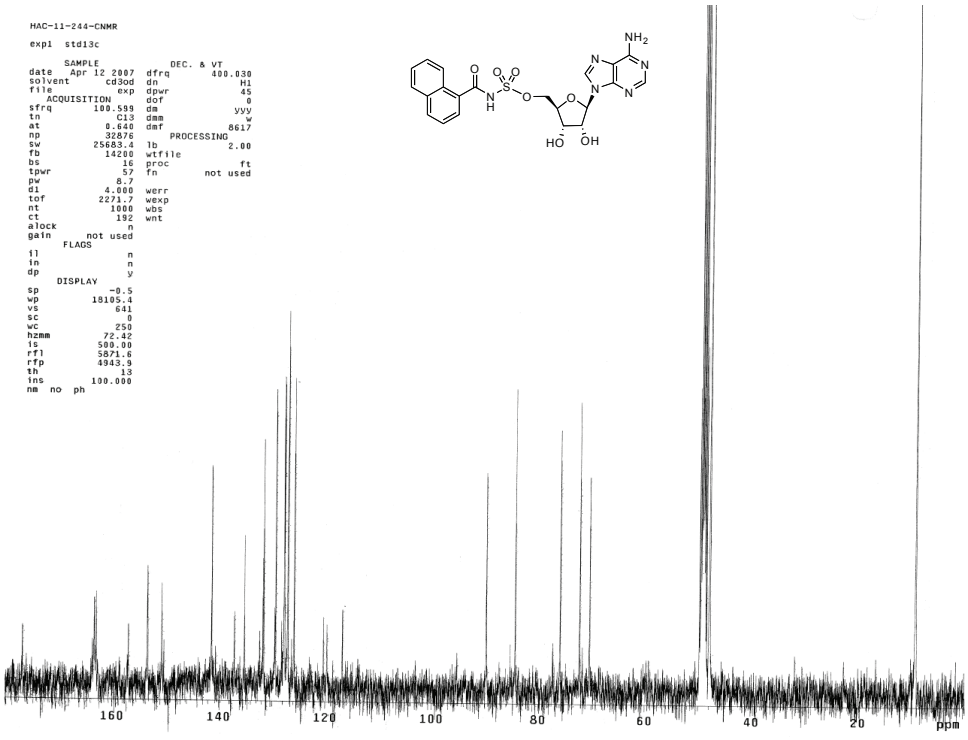

STANDARD 1H OBSERVE
HO
OH
CH3
3.7
ppm


13C OBSERVE
HO
OH
CH3
3.7
ppm


HAC-11-241-NA-NHS
expl stdlh
SAMPLE
date Apr 6 2007
solvent CDCl3
file exp
ACQUISITION
sfrq 400.029
tn H1
at 3.000
np 35992
sw 5998.8
fb 3400
bs 4
tpwr 63
pw 7.1
d1 2.000
tof 0
nt 16
ct 16
alock n
gain not used
FLAGS
il n
in n
dp y
DISPLAY
sp -998.6
wp 5998.8
vs 151
sc 0
wc 250
hzmm 24.00
is 22274.65
rfl 3902.8
rfp 2904.2
th 70
ins 1.000
nm ph
DEC. & VT
dfrq 0
dn
dpwr 30
dof 0
dm nnn
dmm c
dmf 200
PROCESSING
wtfile
proc ft
fn not used
werr
wexp
wbs
wnt
8.800
8.779
8.449
8.446
8.431
8.428
8.135
8.115
7.892
7.647
7.567
7.563
7.544
7.524
3.936
2.927
12 11 10 9 8 7 6 5 4 3 2 1 -0 -1 ppm
1.00 1.00 3.40
0.94 1.10
4.57


HAC-11-f40-45
expl stdlh
SAMPLE
date Apr 16 2007
solvent cd3od
file exp
ACQUISITION
sfrq 400.030
tn H1
at 3.000
np 35992
sw 5998.8
fb 3400
bs 4
tpwr 63
pw 7.1
d1 2.000
tof 0
nt 16
ct 16
alock n
gain not used
FLAGS
il n
in n
dp y
DISPLAY
sp -200.2
wp 4200.3
vs 673
sc 0
wc 250
hzmm 16.80
is 25161.41
rfl 2320.3
rfp 1324.1
th 18
ins 3.000
nm ph
DEC. & VT
dfrq 0
dn
dpwr 30
dof 0
dm nnn
dmm c
dmf 200
PROCESSING
wtfile
proc ft
fn not used
8.537
8.179
7.781
7.780
7.772
7.775
7.463
7.455
7.445
7.439
6.271
6.263
5.271
4.875
4.848
4.623
4.411
4.402
3.349
3.348
3.318
3.317
3.314
3.313
3.310
3.309
3.306
3.305
3.302
3.300
3.263
3.074
3.055
3.038
3.020
1.912
1.900
1.899
1.630
1.376
1.373
1.235
1.216
9 8 7 6 5 4 3 2 1 0 ppm
1.73 2.84 0.89 0.86 1.83 9.38 3.00 14.49
0.75 2.81 0.88 0.84 3.10


HAC-11-244
expl std1h
SAMPLE
date Apr 12 2007
solvent cd3od
file exp
ACQUISITION
sfrq 400.030
tn H1
at 3.000
np 35992
sw 5998.8
fb 3400
bs 4
tpwr 63
pw 7.1
d1 2.000
tof 0
nt 16
ct 16
alock n
gain not used
FLAGS
il n
in n
dp y
DISPLAY
sp -200.1
wp 4600.3
vs 1030
sc 0
wc 250
hzmm 18.40
is 153871.27
rfl 2963.6
rfp 1948.1
th 14
ins 1.000
nm ph
DEC. & VT
dfrq 0
dn
dpwr 30
dof 0
dm nnn
dmm c
dmf 200
PROCESSING
wtfile
proc ft
fn not used
werr
wexp
wbs
wnt
8.406
8.399
8.115
7.862
7.841
7.780
7.762
7.760
7.741
7.441
7.397
7.391
7.381
7.378
7.371
7.367
7.358
6.887
6.854
4.871
4.678
4.665
4.522
4.514
4.496
4.486
4.471
4.460
4.360
3.270
3.266
3.262
3.258
3.254
3.151
3.133
3.114
3.096
3.078
1.249
1.231
1.212
1.195
1.190
1.177
-0.24
1.06
2.85
4.20
4.36
1.00
1.42 1.38
4.50
3.11
10 9 8 7 6 5 4 3 2 1 0 ppm


HAC-11-244-CNMR
expl std13c
SAMPLE
date Apr 12 2007
solvent cd3od
file exp
ACQUISITION
sfrq 100.599
tn C13
at 0.640
np 32876
sw 25683.4
fb 14200
bs 16
tpwr 57
pw 8.7
d1 4.000
tof 2271.7
nt 1000
ct 192
alock n
gain not used
FLAGS
il n
in n
dp y
DISPLAY
sp -0.5
wp 18105.4
vs 641
sc 0
wc 250
hzmm 72.42
is 500.00
rfl 5871.6
rfp 4943.9
th 13
ins 100.000
nm no ph
DEC. & VT
dfrq 400.030
dn H1
dpwr 45
dof 0
dm yyy
dmm w
dmf 8617
PROCESSING
lb 2.00
wtfile
proc ft
fn not used
160 140 120 100 80 60 40 20 ppm

3.10 References

1. Liu, W., Nonaka, K., Nie, L., Zhang, J., Christenson, S. D., Bae, J., Van Lanen, S. G., Zazopoulos, E., Farnet, C. M., Yang, C. F., and Shen, B. (2005) The neocarzinostatin biosynthetic gene cluster from Streptomyces carzinostaticus ATCC 15944 involving two iterative type I polyketide synthases, *Chem. Biol. 12*, 293-302.
2. Sthapit, B., Oh, T. J., Lamichhane, R., Liou, K., Lee, H. C., Kim, C. G., and Sohng, J. K. (2004) Neocarzinostatin naphthoate synthase: an unique iterative type I PKS from neocarzinostatin producer Streptomyces carzinostaticus, *FEBS Lett. 566*, 201-206.
3. May, J. J., Wendrich, T. M., and Marahiel, M. A. (2001) The dhb operon of Bacillus subtilis encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin, *J. Biol. Chem. 276*, 7209-7217.
4. 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, *Proc. Natl. Acad. Sci. U.S.A. 99*, 12120-12125.
5. Cooke, H. A., Zhang, J., Griffin, M. A., Nonaka, K., Van Lanen, S. G., Shen, B., and Bruner, S. D. (2007) Characterization of NcsB2 as a promiscuous naphthoic acid/coenzyme A ligase integral to the biosynthesis of the enediyne antitumor antibiotic neocarzinostatin, *J. Am. Chem. Soc. 129*, 7728-7729.
6. Ji, N., Rosen, B. M., and Myers, A. G. (2004) Method for the rapid synthesis of highly functionalized 2-hydroxy-1-naphthoates. Syntheses of the naphthoic acid components of neocarzinostatin chromophore and N1999A2, *Org. Lett. 6*, 4551-4553.
7. Linne, U., and Marahiel, M. A. (2004) Reactions catalyzed by mature and recombinant nonribosomal peptide synthetases, *Methods Enzymol. 388*, 293-315.
8. Luo, Y., Lin, S., Zhang, J., Cooke, H. A., Bruner, S. D., and Shen, B. (2008) Regiospecific O-methylation of naphthoic acids catalyzed by NcsB1, an O-methyltransferase involved in the biosynthesis of the enediyne antitumor antibiotic neocarzinostatin, *J. Biol. Chem. 283*, 14694-14702.
9. Gulick, A. M. (2009) Conformational Dynamics in the Acyl-CoA Synthetases, Adenylation Domains of Non-ribosomal Peptide Synthetases, and Firefly Luciferase, *ACS Chem. Biol. 4*, 811-827.
10. Conti, E., Franks, N. P., and Brick, P. (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes, *Structure 4*, 287-298.
11. Gulick, A. M., Lu, X., and Dunaway-Mariano, D. (2004) Crystal structure of 4-chlorobenzoate:CoA ligase/synthetase in the unliganded and aryl substrate-bound states, *Biochemistry 43*, 8670-8679.
12. Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S, *EMBO J. 16*, 4174-4183.

13. Rhodes, G. (2006) *Crystallography made crystal clear : a guide for users of macromolecular models*, 3rd ed., Elsevier/Academic Press, Amsterdam ; Boston.
14. Franks, N. P., Jenkins, A., Conti, E., Lieb, W. R., and Brick, P. (1998) Structural basis for the inhibition of firefly luciferase by a general anesthetic, *Biophys. J. 75*, 2205-2211.
15. Miethke, M., Bisseret, P., Beckering, C. L., Vignard, D., Eustache, J., and Marahiel, M. A. (2006) Inhibition of aryl acid adenylation domains involved in bacterial siderophore synthesis, *FEBS J. 273*, 409-419.
16. Finking, R., Neumuller, A., Solsbacher, J., Konz, D., Kretzschmar, G., Schweitzer, M., Krumm, T., and Marahiel, M. A. (2003) Aminoacyl adenylate substrate analogues for the inhibition of adenylation domains of nonribosomal peptide synthetases, *Chembiochem 4*, 903-906.
17. Liu, Y., and Bruner, S. D. (2007) Rational manipulation of carrier-domain geometry in nonribosomal peptide synthetases, *Chembiochem 8*, 617-621.
18. Liu, W., Christenson, S. D., Standage, S., and Shen, B. (2002) Biosynthesis of the enediyne antitumor antibiotic C-1027, *Science 297*, 1170-1173.
19. Van Lanen, S. G., Dorrestein, P. C., Christenson, S. D., Liu, W., Ju, J., Kelleher, N. L., and Shen, B. (2005) Biosynthesis of the beta-amino acid moiety of the enediyne antitumor antibiotic C-1027 featuring beta-amino acyl-S-carrier protein intermediates, *J. Am. Chem. Soc. 127*, 11594-11595.
20. Zhang, C., Griffith, B. R., Fu, Q., Albermann, C., Fu, X., Lee, I. K., Li, L., and Thorson, J. S. (2006) Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions, *Science 313*, 1291-1294.
21. Kennedy, D. R., Gawron, L. S., Ju, J., Liu, W., Shen, B., and Beerman, T. A. (2007) Single chemical modifications of the C-1027 enediyne core, a radiomimetic antitumor drug, affect both drug potency and the role of ataxia-telangiectasia mutated in cellular responses to DNA double-strand breaks, *Cancer Res. 67*, 773-781.
22. Maeda, H., Edo, K., and Ishida, N. (1997) *Neocarzinostatin: The Past, Present, and Future of an Anticancer Drug*, Springer, Tokyo.
23. Shen, B., Liu, W., and Nonaka, K. (2003) Enediyne natural products: biosynthesis and prospect towards engineering novel antitumor agents, *Curr. Med. Chem. 10*, 2317-2325.
24. Bernier, S., Akochy, P. M., Lapointe, J., and Chenevert, R. (2005) Synthesis and aminoacyl-tRNA synthetase inhibitory activity of aspartyl adenylate analogs, *Bioorg. Med. Chem. 13*, 69-75.

Chapter 4

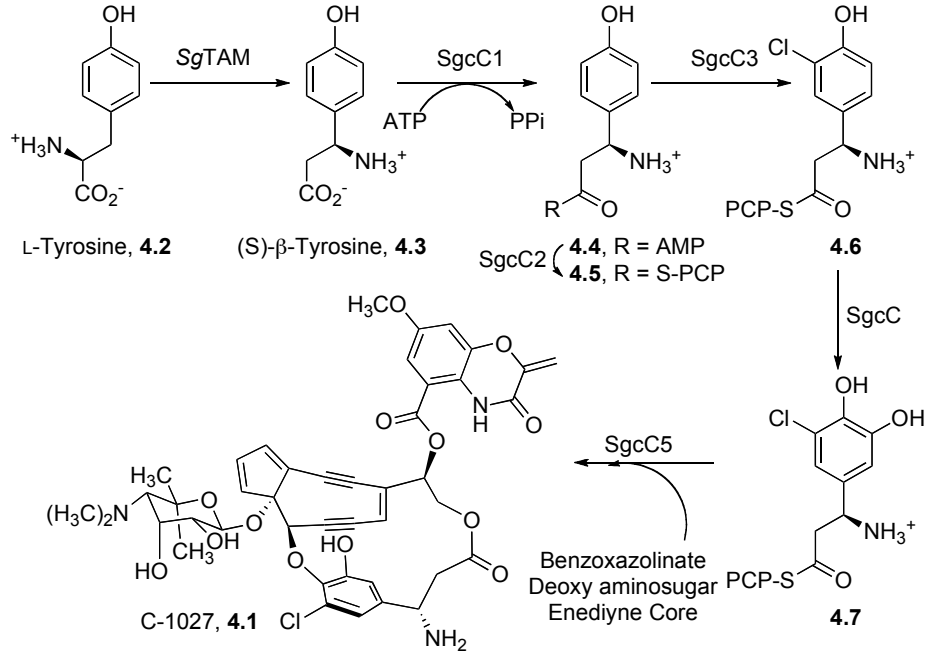
Biochemical and Structural Investigations into the Mechanism of the Tryosine Aminomutase *Sg*TAM

Includes data from an article published in *J. Am. Chem. Soc.* **2007**, *129*, 15744-5 in collaboration with C.V. Christianson, T.J. Montavon, G.M. Festin, B. Shen, and S.D. Bruner.

4.1 Introduction

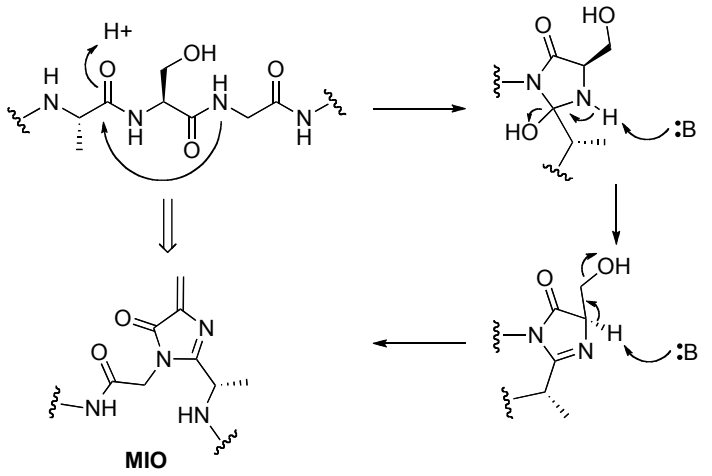
C-1027 (**4.1**) is an enediyne antitumor-antibiotic that contains a functionalized β-tyrosine moiety, (*S*)-3-chloro-4,5-dihydroxy-β-phenylalanine (*1, 2*). This component of C-1027 is biosynthesized and attached to the enediyne core via six enzymatic transformations (Scheme 4.1) (*3*). *Sg*TAM, originally named SgcC4, is the first of these enzymes to act, catalyzing a 2,3-amino shift on L-α-tyrosine (**4.2**) to form (*S*)-β-tyrosine (**4.3**) (*4-6*). Next, SgcC1 activates **4.3** as an (*S*)-β-tyrosyl adenylate (**4.4**) and is subsequently loaded onto the phosphopantheinyl arm of the peptidyl carrier protein (S-PCP) SgcC2, forming **4.5** (*7, 8*). The conjugated β-amino acid is then chlorinated and hydroxylated to generate **4.7** by SgcC3 and SgcC, respectively (*7, 9, 10*). Finally, the condensation enzyme SgcC5 attaches the functionalized (*S*)-β-tyrosine to the enediyne core of C-1027 *(11)*.

Scheme 4.1: Biosynthesis of β-tyrosine moiety and attachment to enediyne core of C-1027



*Sg*TAM is a tyrosine aminomutase (TAM) that contains the unique prosthetic group 4-methylideneimidazole-5-one (MIO, Scheme 4.2) (*4, 5*), which is formed by the self-condensation of a conserved Ala-Ser-Gly motif within the enzyme. MIO-containing

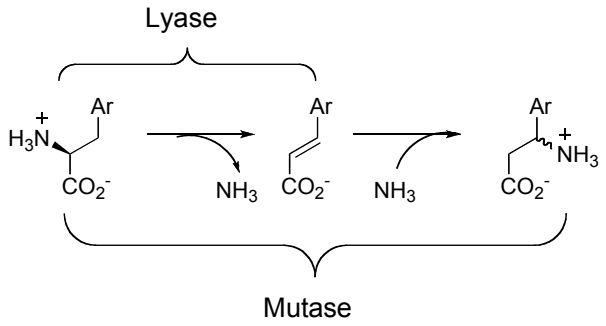
Scheme 4.2: Post-translational formation of MIO prosthetic group from Ala-Ser-Gly motif



enzymes are known to catalyze two types of transformations involving aromatic amino acids, including conversion of histidine, phenylalanine, and tyrosine to their corresponding deaminated, unsaturated products, urocanate, cinnamate, and *p*-hydroxycinnamate (HCA), respectively (Scheme 4.3) (*12-14*). Enzymes that catalyze these transformations are called ammonia lyases. Lyases accomplish this elimination via electrophilic catalysis involving the MIO, prosthetic group. A recently discovered transformation catalyzed by MIO-containing enzymes is a 1,2-amino shift. Examples are known where tyrosine and phenylalanine are converted to their corresponding β-amino acids involving an additional step to lyase activity, which involves the readdition of ammonia into an α,β-unsaturated intermediate (*4, 15-17*). *Sg*TAM belongs to the aminomutase class of MIO-dependent enzymes and preferentially converts L-tyrosine to

(*S*)-β-tyrosine. This recent discovery reignited interest in the unusual chemistry of MIO-containing enzymes. The mechanism by which *Sg*TAM and other MIO-containing enzymes catalyzes the conversion of L-amino acids to their lyase or mutase products is intensely debated (*18, 19*). For a comprehensive review of MIO chemistry, see Appendix 2.

Scheme 4.3: Reactions catalyzed by MIO-dependent enzymes with L-tyrosine, L-phenylalanine, and L-histidine. Ar = aromatic ring.



There are two very different proposed mechanisms by which lyases and mutases act, though both mechanisms involve electrophilic catalysis driven by the MIO. The first mechanism to be proposed is the amino-MIO adduct mechanism (Figure 4.1A), which was uncontested for 35 years (*18*). This mechanism involves the electrophilic addition of the α-amino group on the substrate into the exo-methylene carbon of the MIO. Then, deprotonation at the β-position occurs, followed by formation of a carbanion intermediate and elimination via an E1-cB mechanism. It is likely that binding of the MIO to the amine on the substrate puts it in proper alignment for antiperplanar elimination to occur, facilitating elimination of the leaving group. One issue with this mechanism is the high pKa of the benzylic proton; formation of the MIO adduct alone only lowers the pKa of this proton by about 4 units, from a value of ~40 (*20*). Because of this issue, the

amino-MIO adduct mechanism fell out of favor for some time upon the suggestion of a new mechanism by Rétey and coworkers—a Friedel-Craft type mechanism (Figure 4.1B) (*19*). This mechanism involves electrophilic activation of the substrate's aromatic side chain, resulting in aromatization of the MIO and in a carbocation intermediate. This carbocation contributes to a lowering in pKa of the β-proton. Reformation of the MIO leads to ammonia elimination from the substrate. For mutases, free ammonia adds back into the unsaturated substrate via a 1,4-addition.

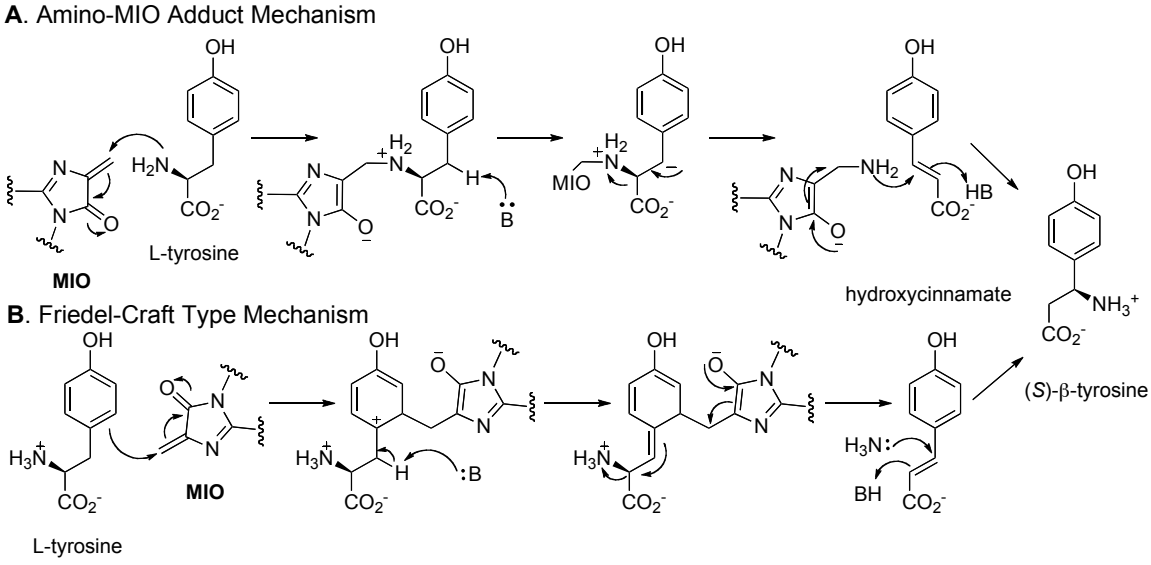


Figure 4.1: Two proposed mechanisms for MIO-containing enzymes with L-tyrosine as the substrate. Either hydroxycinnamate is produced or an additional step may occur for mutase activity involving the further addition of ammonia to produce (*S*)-β-tyrosine.

There are a few issues associated with a Friedel-Craft type mechanism, the first being the prediction of the carbocation intermediate. Based on the structures of a variety of MIO-containing enzymes, the active site is located at the positive end of a number of α-helices (*20*), a situation that would very much disfavor a positively charged

intermediate. On the other hand, the positive nature of the active site could play a role in lowering the pKa of the β-proton for the amino-MIO adduct mechanism by stabilizing the incipient carbanion on the α-carbon. Additionally, it has been demonstrated that the lyase (as well as mutase) reaction is reversible; feeding lyases exogenous ammonia and the corresponding olefin results in formation of the α-amino acid (*21, 22*). The reverse of the Friedel-Craft mechanism is not chemically reasonable, which adds additional doubt to the likelihood of this mechanism. Rétey originally postulated that a tyrosine ammonia lyase (TAL) could not exist because the *p*-hydroxyl group would make the formation of the carbocation intermediate very unlikely (*23*). This hypothesis has been subsequently disproven as a number of TALs and TAMs are now known. Therefore, based on current structural information as well as past biochemical studies, we favor the earlier of the two mechanisms: the amino-MIO adduct mechanism.

The mechanism of MIO-dependent lyases has been thoroughly studied and the discovery of MIO aminomutases has revealed even more insight into their chemistry (*4*). Specifically, it has been demonstrated that *Sg*TAM not only produces β-tyrosine, but also it produces hydroxycinnamate (HCA), the concentration of which increases over time to eventually make up 100% of total substrate/product ratio (Figure 4.2) (*5*). The overall time course for the reaction with L-tyrosine revealed that β-tyrosine concentration never exceeds 60% and is quickly outpaced by HCA production, which is produced from both β-tyrosine and L-tyrosine (Figure 4.2). Either mechanism would support HCA as an intermediate in the catalytic pathway of MIO aminomutases, and an increasing production of it over time suggests that it and ammonia diffuse out of the active site and

are unlikely to rebind. Interestingly, it has been shown that addition of high concentrations of ammonium chloride (10 mM) and HCA (10 mM) to *Sg*TAM results in the production of 13 μM β-tyrosine after 24 h (*5*). Additionally, *Sg*TAM can use L-3-chlorotyrosine as an amino source for the reverse reaction. When *Sg*TAM was incubated with 10 mM of L-3-chlorotyrosine and 10 mM HCA, 27 μM β-tyrosine was produced after 24 h. These observations suggest that ammonia remains bound in the active site longer than the unsaturated acid, perhaps due to a covalent bond with the MIO. Since L-3-chlorotyrosine is not the preferred substrate of *Sg*TAM, following deamination its unsaturated product leaves the active site, the ammonia molecule remains bound and HCA diffuses in. This differing rate of diffusion has been observed previously (*18*) and lends further support for the amino-MIO adduct mechanism. Over time and under standard reaction conditions, the concentration of free ammonia is too low to promote further β-tyrosine formation and all tyrosine is converted to HCA.

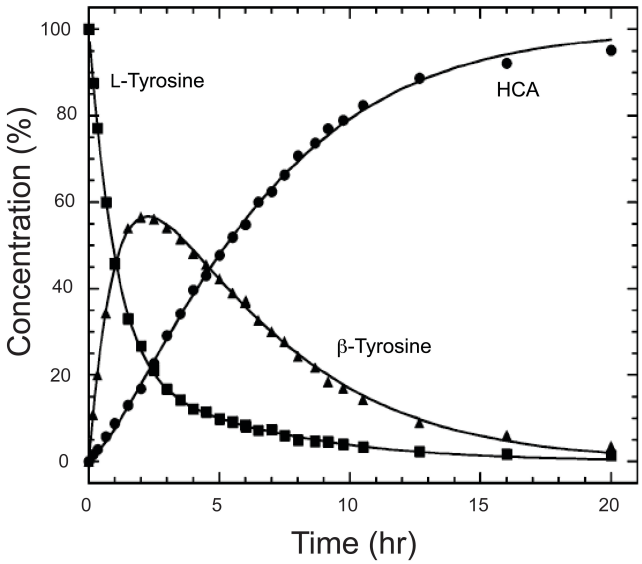


Figure 4.2: Time course of *Sg*TAM-catalyzed conversion of L-tyrosine (■) to β-tyrosine (▲) and HCA (●).

The fact that mutases also have lyase activity begs the question: What differences in these enzymes are responsible for different reaction pathways? A mutase needs to catalyze the additional step of a 1,4-addition of ammonia into the α,β-unsaturated acid to

form a β-amino acid. The fact that the reverse reaction of lyases is also observed complicates the issue of differing between lyases and mutases, as one would expect a significantly higher dissociation constant for α,β-unsaturated products with lyases that would disallow further reaction with unsaturated acids (though it is significant that a much higher concentration of ammonia is required to observe this activity). Furthermore, if the reverse reaction can be observed in lyases, why can't these enzymes form the corresponding β-amino acid? These questions need to be answered to truly differentiate the chemistry of lyases and mutates.

We have previously solved the first structure of an MIO aminomutase (*Sg*TAM) (Figures 4.3) and have determined pertinent active site residues (Figure 4.4) (*6*). With L-tyrosine modeled into the active site, key hydrogen bonding interactions are apparent between the substrate carboxylate and Arg311 as well as between the substrate phenol and His93 and Tyr415. Tyr308 and Asn438 are in close proximity to the MIO. Tyr63, positioned directly above the α-protons, is believed to be the enzymatic base and is rigidly conserved among MIO-containing enzymes. The biologically active unit is a tetramer and the observed asymmetric unit is a dimer. Residues within the active site originate from three monomers, though the majority is on the same chain from which the MIO originates. We compared the structure *Sg*TAM to that of a TAL and active site residues are well conserved that, except for Tyr415. Additionally, residues Glu71 and Tyr303 close the tunnel to the active site of *Sg*TAM, while the tunnel of the TAL is open. These residues that block the opening to the active site are believed to sequester ammonia and/or L-tyrosine in the active site so that the complete transformation to β-tyrosine can

occur, instead of immediately releasing HCA. We have probed Tyr63 to prove its function as the enzymatic base via biochemical (*24*) and structural studies. This chapter will include discussions into insights gained through these studies mentioned above as well as additional with mutant constructs of *Sg*TAM (a full list can be found in the Experimentals section).

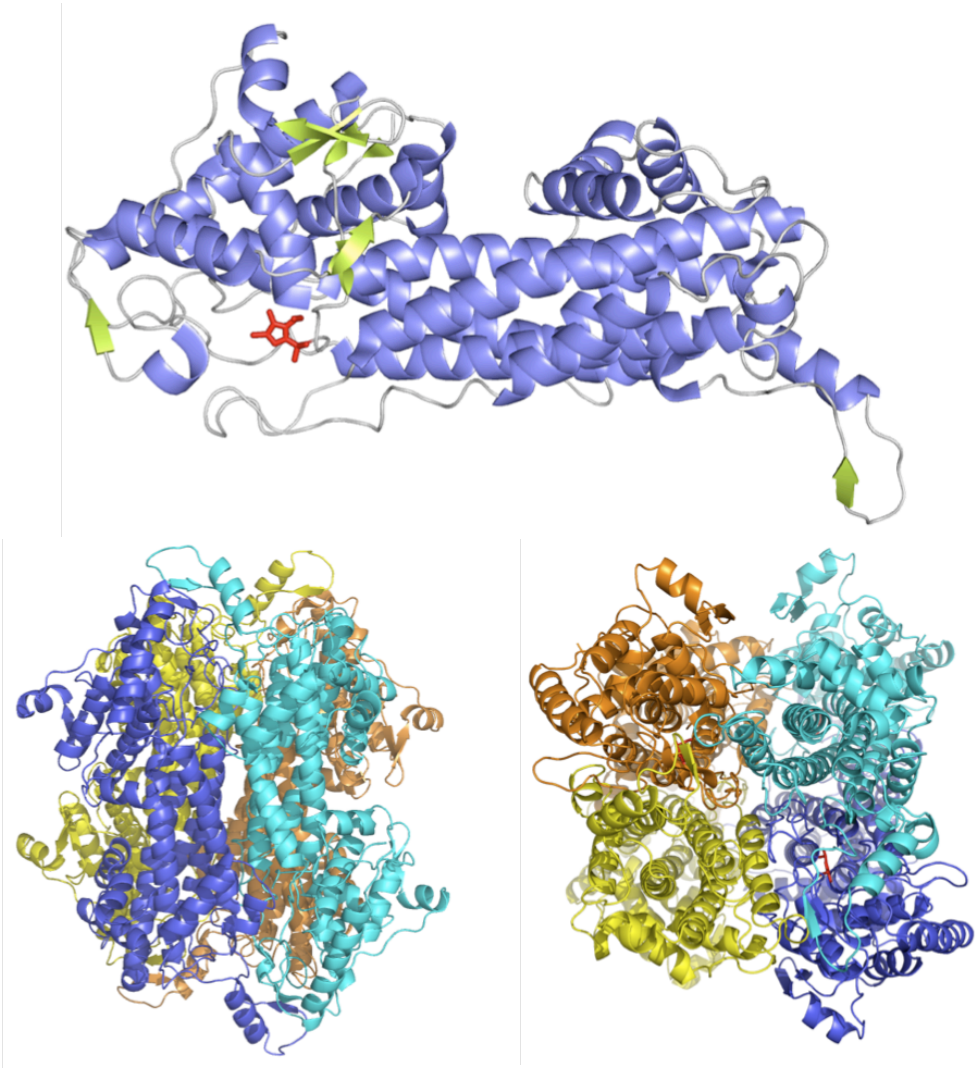


Figure 4.3: Cartoon representations of *Sg*TAM X-ray structure. (Top) Monomer is made up primarily of α-helices with one β-sheet. Two additional β-sheets form at the dimer interface. (Bottom) Two tetramers, the biologically active unit, are shown down two axes of symmetry. Each monomer is indicated by a different color.

4.2 Activity assays of mutants

All of the mutant constructs produced were tested for activity using a previously published procedure (*4*). As will be discussed in the following sections, no turnover was observed for any mutant constructs involving changes in the immediate active site (Figure 4.4). A variety of binding studies were conducted to gain further information into the effect of these mutations on substrate binding. Mutant constructs that did show turnover include residues located with in the tunnel leading up to the active side. These constructs were further probed via time course assays (as in Figure 4.2).

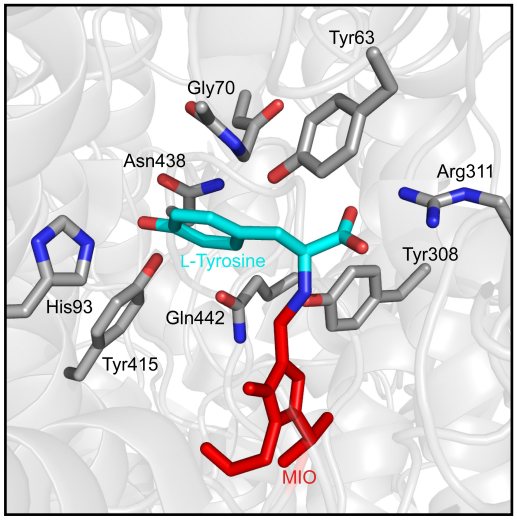


Figure 4.4: View of *Sg*TAM active site. L-Tyrosine is modeled in (cyan) and bound to MIO (red) with active site residues (gray).

4.3 Selectivity switch, His93

Residue His93 was proposed to be the selectivity switch for *Sg*TAM, based on similar studies with lyases. When mutated to phenylalanine, it was predicted that it would favor L-phenylalanine as a substrate and produce β-phenylalanine, converting *Sg*TAM into a phenylalanine aminomutase (PAM). This assumption was based on prior research done on *Rs*TAL, a tyrosine ammonia lyase from *R. sphaeroides*, which was shown to change substrate specificity from L-tyrosine to L-phenylalanine with mutation of His89 of this enzyme to Phe (for sequence alignment, see Figure 4.5) (*25, 26*). Unfortunately, in the His93Phe mutant construct of *Sg*TAM this change in specificity did

not occur; in fact, no activity for either substrate was detected even after 24 h of incubation with enzyme. The lack of activity may be explained by improper folding or by simply poor binding of both L-tyrosine and L-phenylalanine. Furthermore, wild-type mutase activity is quite slow (*5*) in comparison to *Rs*TAL lyase activity (*25*). *Rs*TAL has a k_{cat} of 27.7 s^{-1} and a k_{cat}/K_M of $1.1x10^5$ $M^{-1}s^{-1}$, while *Sg*TAM has a k_{cat} of 0.010 s^{-1} and a k_{cat}/K_M of 360 $M^{-1}s^{-1}$. Lyase activity for *Sg*TAM is substantially slower as well; HCA is a side-product of the mutase reaction, with a k_{cat} of $1.2x10^{-3}$ s^{-1} and k_{cat}/K_M of 50 $M^{-1}s^{-1}$. Mutase activity of *Sg*TAM is 10^8 times less efficient than *Rs*TAL lyase activity and perturbations to the active site could easily extinguish activity overall. Additionally, it was predicted based on UV-Vis spectroscopy that *Sg*TAM only has about 18% of possible MIO formed. Minor changes in the active site might perturb the structure enough to eliminate nearly all MIO formation. Finally, wild-type *Sg*TAM shows very little activity towards L-phenylalanine after 24 h (*27*), while *Rs*TAL does. Perhaps it is the combination of a sluggish enzyme and other factors that make conversion of this TAM into a PAM difficult.

4.4 A potential MIO modulator, Tyr308

Tyr308 appears to play a crucial role either in mediating the formation of the MIO or by maintaining the MIO in a non-aromatic, active state. As shown in Figure 4.4, Tyr308 is in close proximity to the MIO, and the hydroxyl group of the residue appears to form a hydrogen bond with the NH of the MIO and with the side chain of Gln442. Previously, the function of this residue was probed in the histidine ammonia lyase (HAL) *Pp*HAL via mutagenesis and biochemical and structural analysis (*28*). Tyr280 (the same

residue based on alignment with *Sg*TAM) was mutated to Phe (Figure 4.5). The structure of this mutant construct was compared to that of a high-resolution structure of the wild-type HAL; both structures were solved with free cysteine bound in the active site. In the wild-type structure (the MIO is formed at residues 142-144), the nitrogen from the Gly residue of the MIO (144-N) appeared to be in its pyramidal sp^3 hybridized state. Contrarily, the Tyr280Phe mutant X-ray structure revealed a planar sp^2 hybridized 144-N, suggesting that the MIO existed in its intermediate aromatic state. The effect of this mutation seemed to be due to the lack of hydrogen bonding of Tyr280 to Glu414 (analogous to Gln442 in *Sg*TAM), which caused a displacement of the Glu414 side chain. This hydrogen bond may present enough negative charge at the nitrogen from the Ser residue of the MIO (143-N) to cause sp^3 hybridization of 144-N in an unliganded state. Without this hydrogen bond, the MIO may not be able to remain in its active, catalytic state. Mutation of Tyr308 to Phe in *Sg*TAM led to a loss of catalytic activity, as expected (the Tyr280Phe mutant of *Pp*HAL had less than 1% activity compared to wild-type (*28*) and the Tyr300Phe mutant of *Rs*TAL had no activity (*29*)). This mutant was further investigated in binding studies, which will be discussed in section 4.8.

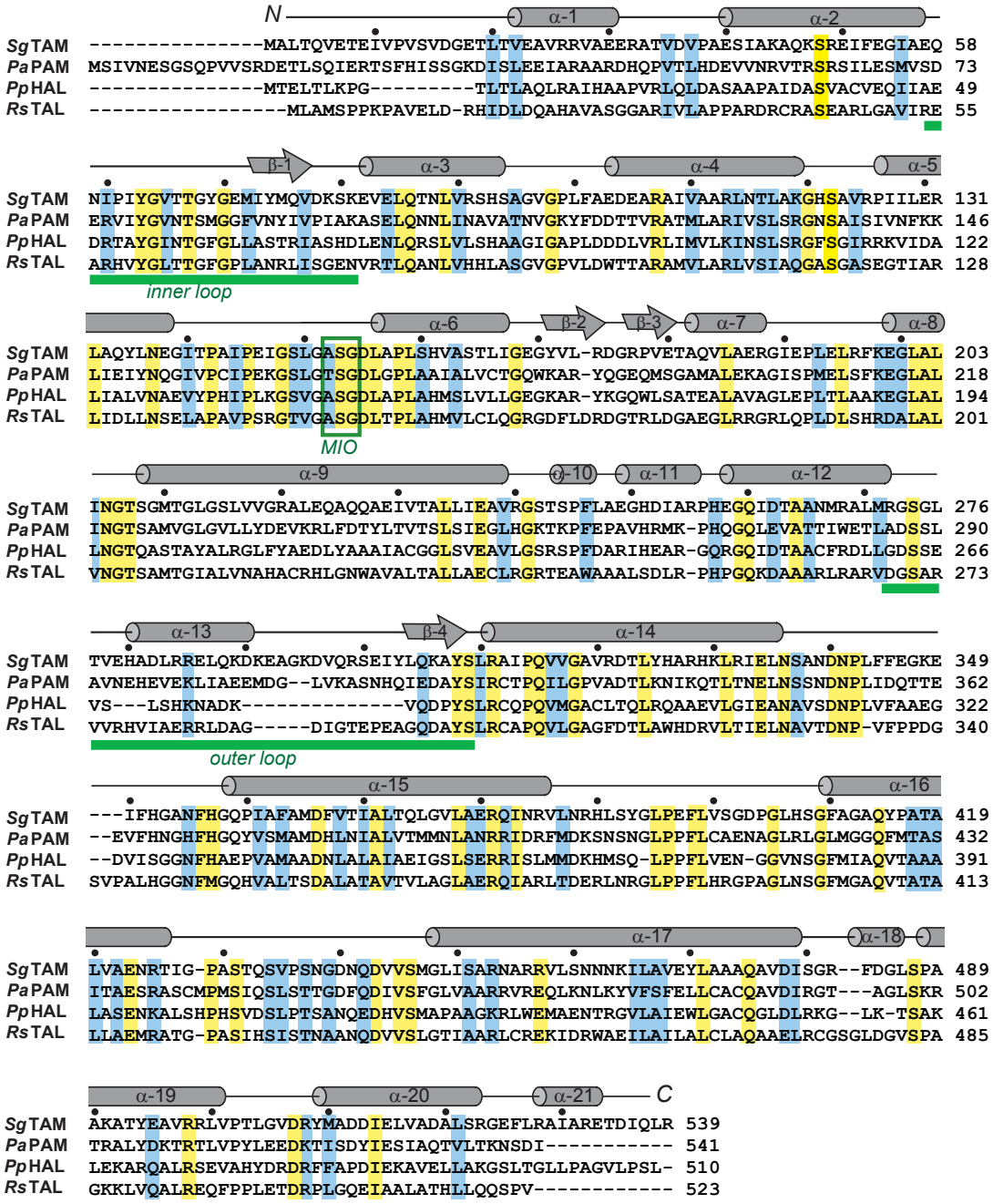


Figure 4.5: Sequence alignment of representative MIO-dependent ammonia lyases and aminomutases. The secondary structure of *Sg*TAM is indicated above the sequence and inner and outer loop regions are indicated by green lines below sequence. MIO forming residues are indicated by green box. Aligned sequences include *Sg*TAM (*S. globistorus* tyrosine aminomutase), *Pa*PAM (*P. agglomerans* phenylalanine aminomutase), *Pp*HAL (*P. putida* histidine ammonia lyase), and *Rs*TAL (*R. sphaeroides* tyrosine ammonia lyase).

4.5 A possible mutase/lyase switch, Tyr415

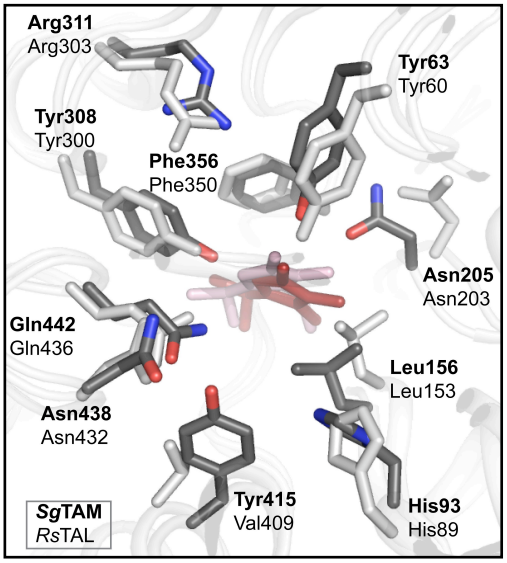


Figure 4.6: Overlay of active sites of *Sg*TAM (dark gray) and *Rs*TAL (light gray, pdb #2O6Y).

An overlay of the active sites of *Rs*TAL and *Sg*TAM shows very similar active sites where the major difference is at residue Tyr415, where a valine is located in *Rs*TAL (Figure 4.6) (*6, 26*). Since *Rs*TAL is a lyase, it was expected that the Tyr415Val mutant construct of *Sg*TAM would lack mutase activity and would produce only HCA. Unfortunately, this mutation did not convert *Sg*TAM into a lyase; no mutase or lyase activity was observed even after 24 h. Prior to any folding studies, we postulated that this change in polarity of the active site was affecting binding of L-tyrosine, perhaps to the point where it would prefer L-phenylalanine. This change in substrate specificity was also not detected, perhaps because additional mutations were required.

The double mutant of Tyr415Phe and His93Phe was made and tested for activity against the alternate substrate. Again, no activity toward L-phenylalanine was observed. It seems very likely that the single mutation in the active site is leading to problems in folding or tetramer assembly as attempts to crystallize the His93Phe mutant construct failed. In our experiments, mutants typically crystallized similarly to wild-type enzyme, with and without ligand (see Glu71Ala and Tyr63Phe mutants). Further biochemical studies of the Tyr415Val and His93Phe mutant constructs might give insight into why no lyase or mutase activity is observed.

4.6 The predicted enzymatic base, Tyr63

Tyr63 was predicted to be the enzymatic base of *Sg*TAM; it likely exists in the phenolate form and is situated directly above the α-carbon of substrate L-tyrosine (Figure 4.4) (*24*). Tyrosine is absolutely conserved at this position among all sequence homologs (Figure 4.5). The deprotonated form of tyrosine would be favored, as the active site is located at the positive end of seven α-helices, which would stabilize the negative charge. Additionally, the phenolate hydrogen-bonds to the backbone of Gly70, lowering the pKa of the sidechain. If our assumption is correct, mutation of the tyrosine residue to a phenylalanine would remove the phenolate base, rendering the enzyme inactive. As expected, the Tyr63Phe mutant showed no turnover of substrate. Mutation of *Rs*TAL Tyr60 to Phe also resulted in a complete loss of activity (*29*).

To further prove that Tyr63 is the enzymatic base and that mutation results in the loss of this base and is not a result of perturbation of the active site, we solved the cocrystal structure of the Tyr63Phe mutant construct with L-tyrosine (Figure 4.7). The

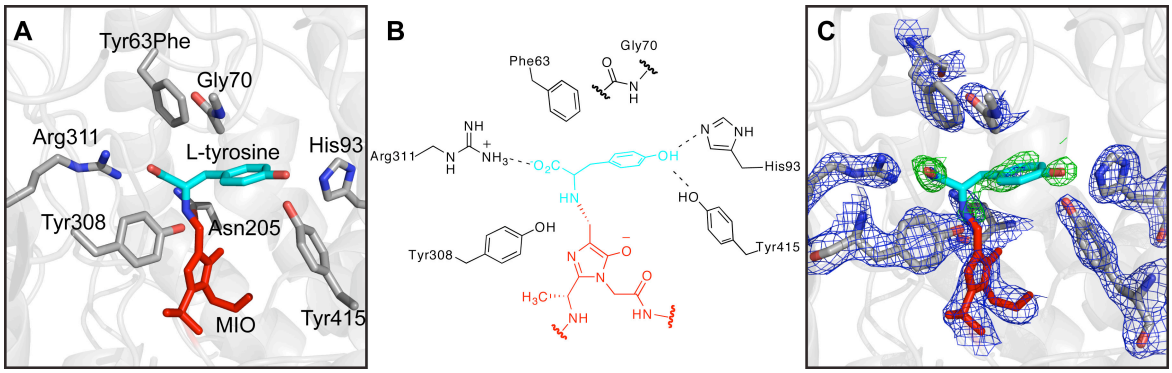


Figure 4.7: Tyr63Phe mutant structure. A) Active site of mutant with L-tyrosine bound to the MIO. B) Representation of major interactions in the active site. C) Electron density maps of active site. The $2F_o–F_c$ map is controured to 1.5 σ and the $F_o–F_c$ to map 3.0 σ. Maps were generated without L-tyrosine modeled into active site.

crystal diffracted to a resolution of 2.3 Å and the F_o-F_c map of the structure minus L-tyrosine in the active site showed density for the majority of the ligand bound to the MIO. Electron density corresponding to the substrate's phenyl ring is partially visible and the phenol has significant density, appearing to be hydrogen bonded to residues His93 and Tyr415, as seen in the α,α-difluoro-β-tyrosine cocrystal structure (*24*). Though density for the α-carbon of L-tyrosine is weak, the MIO is most certainly bound to a ligand, as density extends from the exo-cyclic methylidene carbon of the MIO. Electron density for this bond is apparent in both the F_o–F_c and $2F_o$–F_c maps. The weak density may be attributed either to substoichiometric binding of the ligand or ligand in multiple conformations in the active site. If it were the case that less than a quarter of the active sites have MIO in its active form and the rest are either not fully formed or in their inactive, aromatic state, then it would make sense for fewer active sites to be occupied by ligand or for ligand to be in an alternate positions. It may also be random as to which active site in a tetramer has bound L-tyrosine, which would result in electron density for ligand in both active sites in the asymmetric unit.

4.7 Tunnel mutants

Two main differences exist between TAL and TAM; *Sg*TAM has a tyrosine residue in the active site where *Rs*TAL has a valine and the tunnel to the active site of *Sg*TAM is closed and much smaller compared to that of *Rs*TAL. Since the initial attempt failed to convert *Sg*TAM into an ammonia lyase by making the Tyr415Val mutation, mutations to the tunnel leading to the active site were made. The open active site of *Rs*TAL exposes the MIO to outside solvent. It was postulated that an exposed active site decreases the ability of the enzyme to maintain HCA and ammonia in the active site; creating a more accessible active site in *Sg*TAM might preclude mutase activity. Glu71 and Tyr303 in the tunnel of *Sg*TAM are hydrogen bonded to each other and may act to block the MIO from direct contact with solvent (Figure 4.8). We mutated these residues to sterically less bulky amino acids and analyzed their activity over time. The tunnel

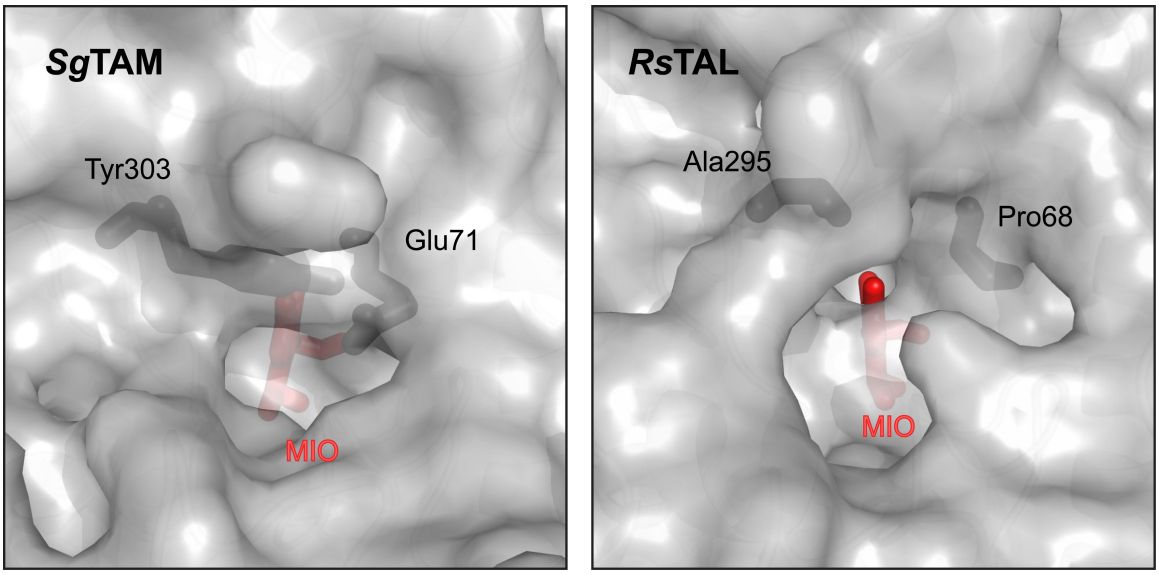


Figure 4.8: Surface views of MIO. The MIO of *Sg*TAM is not solvent accessible due to a hydrogen bond between Tyr303 and Glu71 (left). The MIO of *Rs*TAL is solvent accessible due to a less sterically crowded tunnel (right).

mutant constructs included Glu71Ala, Tyr303Ala, and the double mutant of the two. Despite a decrease in activity, all three showed lyase activity while still producing some amount of β-tyrosine (Figure 4.9). Glu71Ala had the biggest decrease in activity of the two single mutants; the percent of HCA and β-tyrosine both never surpassed 15% even after 24 hrs, with β-tyrosine forming in excess of HCA the entire time. The effect of the Tyr303Ala mutation was limited as a similar time curve was observed to the wild-type enzyme (Figure 4.2), though wild-type converted as much as 60% of tyrosine to β-tyrosine, while the mutant peaked just under 40%. The double mutant resulted in roughly the same time course as the single Glu71Ala mutant.

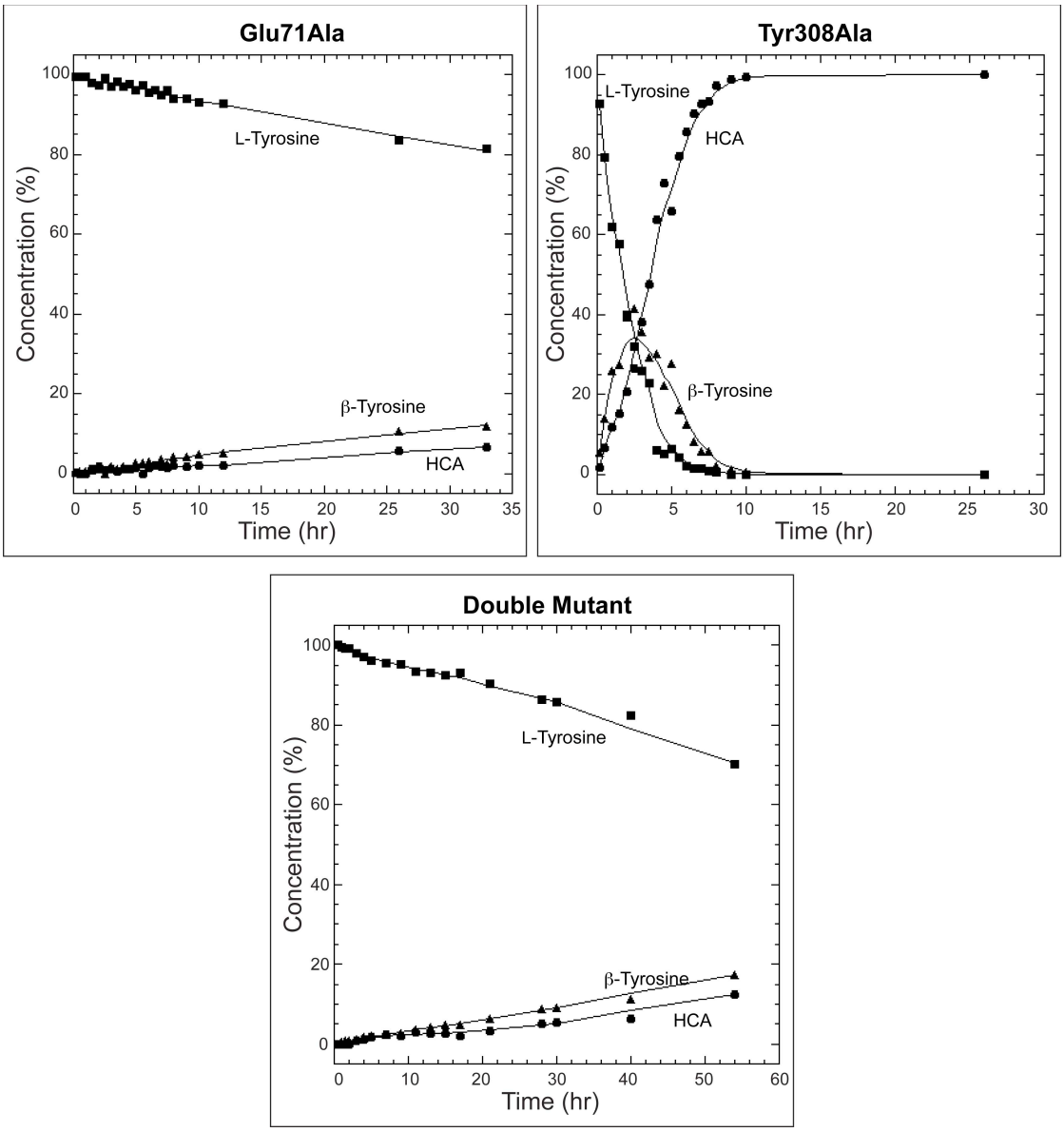


Figure 4.9: Time courses catalyzed converstion of L-tyrosine (■) to β-tyrosine (▲) and HCA (●) by tunnel mutants.

To further investigate the Glu71Ala mutant, we solved its structure to a resolution of 2.3 Å. Figure 4.10A shows a cutaway view of the wild-type *Sg*TAM active site, including the closed tunnel. When Glu71 is mutated to an alanine, the tunnel is open and the active site is solvent exposed (Figure 4.10B). It was expected that opening this tunnel so that it more resembled the tunnel from *Rs*TAL (Figure 4.10C) would result in the increased rate of release of ammonia or HCA before the ammonia could add back into HCA to form β-tyrosine. This activity was not observed for the Glu71Ala mutant or for the double mutant. It appears that *Sg*TAM still acts as a mutase, but its overall rate of turnover is decreased, perhaps as a result of improper formation of the active unit. The biological unit is composed of four monomers, three of which contribute residues to the active site; residues 71 and 303 within one tunnel are from two different monomers. Problems with the interface of the monomers may be to blame for the lowered activity.

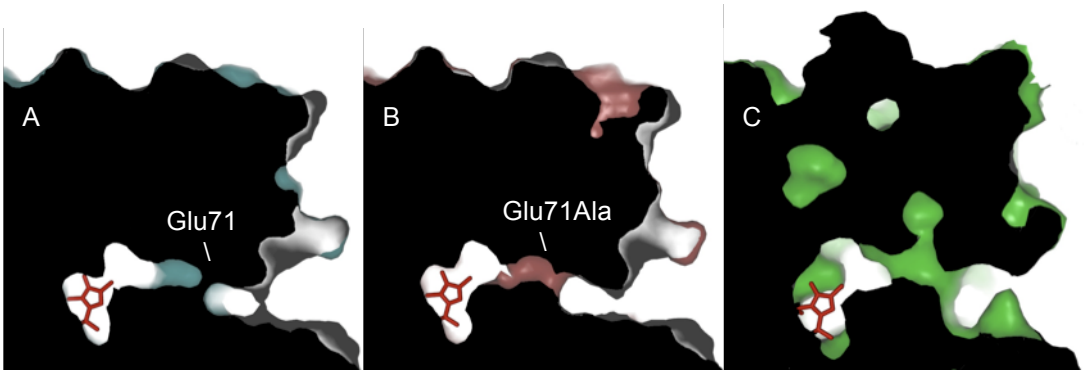


Figure 4.10: Cutaway views of active site and tunnel comparing solvent accessibility to MIO (probe sphere radius of 1.4 Å). A) Wild-type *Sg*TAM with a closed tunnel. B) Glu71Ala mutant of *Sg*TAM with an open tunnel. C) *Rs*TAL with an open tunnel and large active site.

4.8 Binding assays of inactive enzymes

In an effort to further characterize the inactive mutant constructs of *Sg*TAM, we set out to study the binding of L-tyrosine and, in some cases, L-phenylalanine. By doing so, we hoped to determine which mutant constructs were inactive due to lack of pertinent catalytic resides and which were simply unable to bind substrate.

4.7.1 Fluorescent quenching

Fluorescent quenching was the first technique attempted but was quickly abandoned, despite a small amount of observed quenching. A low signal-to-noise ratio was observed, caused by the complete lack of tryptophan residues in *Sg*TAM. The error associated with each measurement was simply too large to be considered accurate and the values were within the error range of the blanks (despite previous success with *Rs*TAL (*29*)).

4.7.2 Equilibrium dialysis

We next attempted to determine the binding constant of *Sg*TAM mutants for L-tyrosine using equilibrium dialysis. We initially focused on the Tyr63Phe mutant construct as we knew binding was occurring because electron density for L-tyrosine was present in the active site of the crystal structure. It was apparent that binding was occurring with the Tyr63Phe mutant construct, though values obtained from multiple experiments were imprecise and the error was significant. No specific binding was observed for the His93Phe/Tyr415Val double mutant, though aberrant binding of tyrosine was seen (that is, an excessive amount of radioactivity was measured on the protein side

of the dialyzer, but did not correlate to tyrosine concentration). Despite the minor success with the Tyr63Phe mutant, this technique was abandoned due to the inaccuracy of binding constants among experiments, likely a result of the difficult loading of the dialyzers. Air bubbles at the membrane of the dialyzers more often than not obscured dialysis across the membrane.

4.7.3 Isothermal Titration Calorimetry

Finally, we tried isothermal titration calorimetry (ITC), as it is a highly sensitive biophysical technique commonly used to study the binding of small molecules to macromolecules. It was hoped that its sensitivity would allow us to measure the minute differences in binding that we expected to observe among the various inactive mutants (and one tunnel mutant), which the other techniques did not allow. A number of protein/ligand combinations were tested, all of which showed no binding except for Tyr63Phe and Glu71Ala to L-tyrosine (Figure 4.10). Pairs that were tested and showed

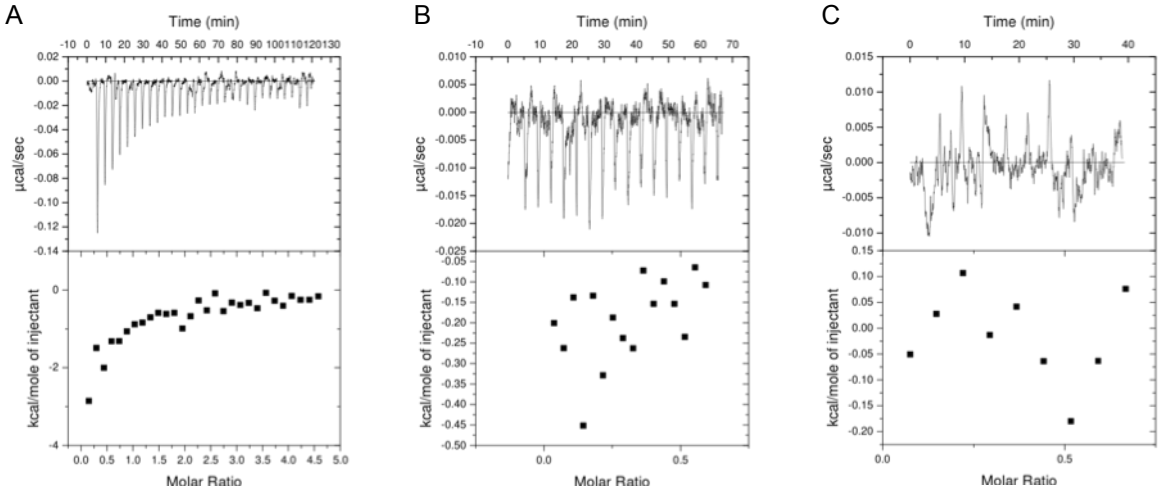


Figure 4.11: ITC curves of *Sg*TAM mutants with L-tyrosine. A)Tyr63Phe mutant, B)Tyr308Phe mutant, C) Tyr415Val mutant.

no binding include Tyr415Phe/L-tyrosine, Tyr415Phe/L-Phenylalanine, His93Phe + Tyr415Phe/L-Phenylalanine, and Tyr308/Phe/L-tyrosine. The latter experiment showed binding but the change in enthalpy was very small and constant for the entire run. It appears that all mutations made to the active site of *Sg*TAM have a significant effect on ligand binding and, for the most part, extinguish activity for this reason.

In the case of the Tyr63Phe mutant, it is not surprising that binding does occur, as this residue acts as the enzymatic base and does not interact with the substrate phenol and it is hypothesized that the mutation does not affect folding or tetramer assembly as we were able to grow crystals and solve the structure. These observations further support the role of the residue in catalysis. Nevertheless, a binding constant was difficult to calculate for this mutant because the full sigmoidal titration curve could not be obtained. Typically, the first few additions of titrant (a solution of ligand) result in roughly the same enthalpy change, while subsequent aliquots result in a decreasing change in enthalpy. In the case of this mutant, after the initial titration the decrease in enthalpy was immediate. Attempts to obtain the anticipated curve were not successful. These attempts include lowering the amount of ligand being titrated in or raising the concentration of the protein in solution. It appears as though all active sites are not available to bind ligand, which would result in a smaller enthalpy change per aliquot added. This result may be caused by incomplete MIO formation. Also, active sites with formed, active MIO may covalently bind L-tyrosine particularly strongly, which would result in smaller changes in enthalpy upon further addition of ligand. ITC is best used for protein/ligand pairs with dissociation constants (K_d) greater than 1 μM, a minimum that can be lowered with more

favorable binding conditions, specifically those that affect a large enthalpic change. The calculated K_d from the partial curve obtained for Tyr63Phe/L-tyrosine was 6.85 μM ± 0.74 (while the apparent binding constant, K_m, was originally measured to be 28 μM (*5*)). Whether or not this value can be considered accurate is debatable, as the appropriate ITC curve was not obtained. It is possible that the dissociation constant is actually lower. The combination of a less than 1:1 binding stoichiometry and a low K_d makes quantifying binding difficult. Nevertheless, we have confirmed that all mutations in the active site, aside from Tyr63Phe, impede ligand binding. Tyr308Phe shows an enthalpy change upon ligand titration, but a typical sigmoidal curve did not appear. Perhaps this result is due to the state of the MIO; as discussed above, mutations to this residue cause the MIO to exist primarily in its aromatic state. Clearly ligand binding is partially due to binding to MIO, hence the lack of the typical sigmoidal titration curve. The Glu71Ala mutant construct still turns over the substrate and hence shows ligand binding, as expected.

4.9 Folding as determined by circular dichroism

The lack of binding of ligands to the His93Phe+Tyr415Phe double mutant (as well as the single mutants of the two) may be a result of improper folding. In order to test this hypothesis, a simple CD experiment was set up to compare the folding of the double mutant to the Tyr63Phe mutant (Figure 4.12). It appears that there is a difference in folding between the two mutant constructs, as seen by the significant reduction in α-helical character of the protein (signified by large dip between 210 and 330 nm). Addition of salt to the double mutant seems to improve folding slightly but the difference between the two mutant constructs is still significant. Furthermore, the increased α-

helical character of the Tyr63Phe mutant compared to the double mutant supports the fact that the latter mutant is well folded, maintains normal ligand binding and has no catalytic ability due to lack of the enzymatic base.

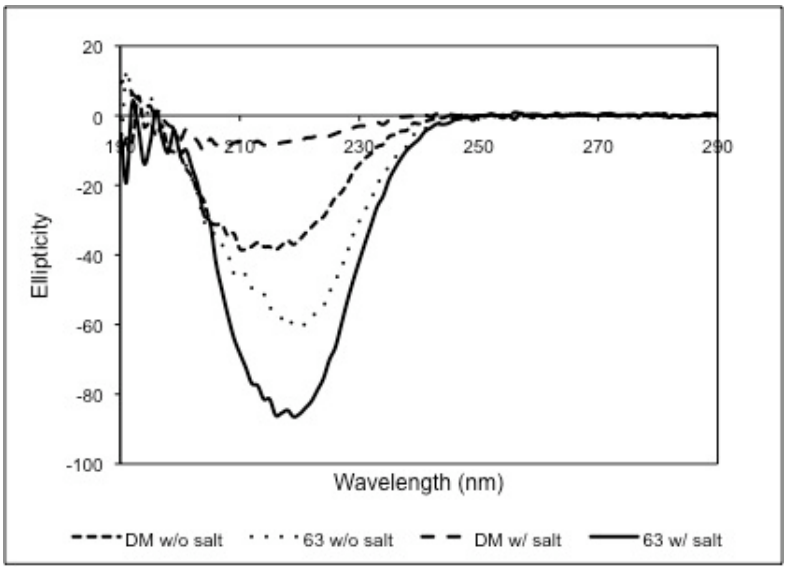


Figure 4.12: Circular dichroism spectra of Tyr63Phe and His93Phe/Tyr415Val mutants with and without salt.

4.10 Conclusion

From these observations, it is quite clear that *Sg*TAM is sensitive to perturbations to the active site. Futhermore, from studies on the Tyr63Phe mutant, which maintains substrate binding, we can conclude that an enzymatic base is required for catalysis and conversion to β-tyrosine cannot occur without deprotonation of the β-carbon first. Simple proximity to MIO and correct geometric arrangement of L-tyrosine is not enough for catalysis. These observations suggest that the active site of *Sg*TAM is incredibly sensitive to mutations. It is possible that certain changes in active site residues affect

either the formation of MIO, maintenance of the MIO in its active, non-aromatic state, or the makeup of the active site due to incorrect folding. To determine which of these problems are responsible for the lack of activity of the majority of mutants, we would need further biochemical experiments, including more in depth folding experiments, dynamic light scattering to determine size, and UV experiments to measure the absorbance of the MIO to verify its state.

4.11 Experimentals

General

PfuUltra, DpnI and other reagents for site-directed mutagenesis were ordered from Stratagene (La Jolla, CA). Oligonucleotide primers were ordered from IDT DNA (Coralville, IA). T4 DNA ligase and restriction enzymes were orderd from New England Biolabs (Ipswich, MA). Kits for uses for DNA purification were purchased from QIAGEN (Valencia, CA). DNA sequencing was done by Genewiz (S. Plainfield, NJ). Thrombin, IPTG, LB media, antibiotics and other reagents were from Sigma-Aldrich (St. Louis, MO) or VWR (West Chester). NovaBlues and BL21 (DE3) were from VWR. HiTrap-Q column and HiLoad 16/60 SuperDex 200 gel filtration column were from GE Healthcare (Pittsburgh, PA).

DNA manipulation

Mutant constructs of *Sg*TAM from *Streptomyces globisporus* (Locus AY048670) were produced from the plasmid expression construct pBS1022 (in pET-28a in *Nde*I-*Hind*III sites) using a modified version of the Quikchange site-directed mutagenesis

protocol as described previously (*5*). Mutants were produced using their corresponding primer pairs as shown in Table 2.1.

Table 2.1: Oligionucleotide primers used for producing mutant constructs of *Sg*TAM with mutant codon underlined.

Mutation	Forward Primer	Reverse Primer
Y63F	5'-cagaacatccccatcttcggggtgaccaccggg-3'	5'-cccggtggtcaccccgaagatggggatgttctg-3'
E71A	5'-ccgggtacggcgcggatgatctacatgc-3'	5'-gcatgtagatcatcgcgccgtacccgg-3'
E71A, DM	5'-ccgggtacggcgctatgatctacatgc-3'	5'-gcatgtagatcatagcgccgtacccgg-3'
H93F	5'-ccaatctcgtccgtagcttcagcgcgggagtcggtcc-3'	5'-ggaccgactcccgcgctgaagctacggacgagattgg-3'
Y303A	5'-gcgctcggagatcgcccctgcagaaggcc-3'	5'-ggccttctgcagggcgatctccgagcgc-3'
Y308F	5'-cctgcagaaggccttctcgctgcgggcc-3'	5'-ggcccgcagcgagaaggccttctgcagg-3'
Y415V	5'-cgccggcgcccaggtccggccaccgc-3'	5'-gcggtggccgggacctgggcgccggcg-3'

The Tyr303Ala mutant was produced using the procedure from the Quikchange handbook. Modifications to the published protocol for the remaining mutants were as follows: the Tyr63Phe and Tyr308Phe mutants were produced using 1x reaction buffer, 100 ng plasmid, 500 ng forward and reverse primer, 200 μM dNTPs, 5% DMSO, and 2.5 U *PfuUltra*, with the following thermal cycler parameters: 95 ºC for 5 min; 35x (95 ºC for 30 sec, 62 ºC for 45 sec, 72 ºC for 10 min); 72 ºC for 10 min; 4 ºC. The remaining mutants had slightly altered procedures: Tyr415Val had an extenstion time of 8 min, Glu71Ala and Tyr93Phe were cycled 20x to lower the chance of random annealing and had an extension time of 10 min. To produce the Glu71Ala/Tyr303Ala double mutant, new primers were used with the Tyr303Ala mutant (E71A, DM) because the original could not again successfully produce the mutation despite repeated attempts. The reaction was cycled 20x with an extension time of 10 min, producing the double mutant. Following mutagenesis, 2 μL of DpnI was added to the reaction mixture and incubated for 2 h at 37 ºC. The QIAquick PCR purification kit was used to removed enzyme and

purify plasmid DNA, which was eluted with 30 μL H_2O. To close nicks resulting from mutagenesis, a ligation was performed overnight with 1 μL T4 DNA ligase, 1x T4 DNA ligase reaction buffer at 16 ºC. Typically, 1 μL of the ligation mixture was transformed into NovaBlues, which were grown in 5 mL LB broth overnight supplemented with kanamycin. Plasmid was isolated using the QIAprep Spin MiniPrep Kit and mutations were confirmed by sequencing.

To produce the Tyr93Phe/Tyr415Val mutant, the Tyr93Phe and Tyr415Val plasmids were digested with NdeI and SacI. The digests were run on an 1% agarose gel and the smaller fragment from Tyr93Phe containing the 93 mutation and the larger fragment from Tyr415Val containing the 415 mutation were excised from the gel. The DNA was purified using the QIAquick Gel Extraction Kit. The two pieces were ligated together with T4 DNA ligase and transformed as above.

Enzyme production and purification

Mutant *Sg*TAM constructs were transformed into BL21(DE3) *E. coli* cells. One colony was used to inoculate 5 mL of LB media and was grown for 16 h at 37 ºC. The overnight culture was subsequently used to inoculate 1 L of LB media, which was grown to O.D. = 0.6 with shaking (37 ºC at 150 RPM). Overexpression was induced by addition of 50 μM IPTG and the culture was incubated overnight at 18 ºC. Cells were collected by centrifugation (20 min, 2000 g), resuspended in 25 mL of lysis buffer containing 20 mM Tris, pH 7.5 and 500 mM NaCl, and frozen (-78 ºC.) The cell pellet was thawed and lysed by two passes through a French Press at 1000 PSI. The lysed cells were then centrifuged (20 min, 10,000 g) to remove cell debris and the protein was purified batch-

wise using Ni-NTA resin. For use in activity and binding assays, the purified protein was dialyzed into a low salt buffer (50 mM CHES pH 9.0, 50 mM KCl) and was either used immediately or stored at -78 ºC in 20% glycerol to be used for assays within one week. Protein for crystallography was dialyzed into low salt buffer (20 mM Tris pH 7.5, 100 mM NaCl), concentrated to less than 1 mL, and the His_6-tag was cleaved using 5 U of thrombin in cleavage buffer (20 mM Tris pH 7.5, 100 mM NaCl, 2 mM $CaCl_2$) for 2 days at 4 ºC. The protein solution was then diluted to 5 mL and purified on a HiTrap-Q ion exchange column (0-1 M NaCl gradient in 50 mM Tris-HCl pH 7.5 and 1 mM β-mercaptoethanol) followed by a Superdex 200 gel filtration column (20 mM Tris-HCl pH 7.5, 100 mM NaCl). The purified protein was concentrated to ~10 mg/mL for crystallography (a total yield of 4 mg/L of cells).

Biochemical assay of Sg*TAM mutant constructs*

The activity of *Sg*TAM mutants was measured using an HPLC based assay (*5*). Reactions were performed at 25 ºC with 100 μM L-tyrosine, 50 mM CHES, 50 mM KCl, pH 9.0 and 0.5 mg/mL enzyme in a total volume of 1.5 mL. All mutant constructs were monitored for activity with time points at 1 h and 18 h. For Glu71Ala and Tyr303Ala mutants, a time course assay was run by quenching 50 μL aliquots of reaction mixture by addition to a tube containing 2 μL 0.5 M HCl. Time points were taken at 10 min, 30 min, then every half hour for 8 h, and then every hour until 12 h. Additional time points were taken after 24 h. Each aliquot was centrifuged for 1 min at 14,000 g to remove enzyme. The supernantant was subjected to *o*-phthaldialdehyde (OPA) derivation by increasing the pH to 10 with 2 μL 1M NaOH followed by mixing with 50 μL of fresh, degassed

OPA reagent (4 mg OPA, 4.25 mL 100 mM sodium borate buffer pH 10.7, 50 μL ethanol and 11 μL β-mercaptoethanol) for 30 sec, and then quenched with 1 μL glacial acetic acid. Derivatized samples were subjected to HPLC analysis on a VYDAC C18 Protein and Peptide column (eluent A: 5 mM sodium acetate pH 5.7 in with 5% THF; eluent B: 5 mM sodium acetate pH 5.7, 45:45:10 methanol/acetonitrile/water; gradient: 0-38 min at 10-50% B, 38-40 min at 50%, 40-52 min at 50-10% B, 52-57 min at 10% B at a flow rate of 1 mL/min).

Mutant constructs Tyr93Phe, Tyr415Val, and the double mutant of the two were tested for activity with L-phenylalanine under similar assay and analysis conditions.

Fluorescence quenching

The binding of *p*-hydroxycinnamic acid (HCA) to the Tyr63Phe mutant was monitored using a Jas.Co FP-6300 Spectrofluorimeter equipped with stirring. In a cuvette, 900 μL of 50 mM CHES pH 9.0, 50 mM KCl and 1.5 μM protein was equilibrated to 25 ºC by stirring for 3 minutes. A number of excitation/emission values were tested and the largest signal was observed for an excitation range between 220 nm and 300 nm, producing the highest emission signal at 345 nm. A 5 mM solution of HCA was titrated in 2 μL aliquots over time, to a concentration of 0.15 mM HCA (no more than 3% of total volume was exceeded with the HCA solution). The same experiment was used to measure HCA binding with the His93Phe mutant and His93Phe/Tyr415Val double mutant.

Equilibrium dialysis

DISPO equilibrium dialyzers (Harvard Apparatus, Holliston, MA) were used to run small-scale dialyses of L-tyrosine and SgTAM mutant constructs to measure the binding constant. One side of the dialyzer contained 50 µl of 0.5 mg/mL protein, 50 mM CHES pH 9.0 and 1 µM ^{3}H-tyr. The other side contained 50 µL of various concentrations of L-tyrosine, ranging from 0 to 100 µM. The dialyzers were shaken at 25 ºC for 16 h and three 3 µL aliquots of solution were removed from each side of the dialyzer and were added to 5 mL aliquots of scintillation fluid in 20 mL vials and subjected to scintillation counting.

Isothermal titration calorimetry

A VP-ITC (MicroCal Inc.) at the Biophysical Instrumentation Facility at MIT was used for all measurements. The basic experiment was set up with a 1.4 mL of protein at a concentration of 30-60 µM in the cell. Ligand was titrated in, typically in 10 µL aliquots at a concentration of 100-300 µM. All experiments were run at 25 ºC, with a DP of 15 µcal/sec. Concentrations of protein and ligand were varied to optimize the curve obtained. Protein/ligand pairs tested include Tyr63Phe with L-tyrosine, Tyr415Val with L-tyrosine and L-phenylalanine, His93Phe+Tyr415Val/L-phenylalanine, Gln71Ala/L-tyrosine and Tyr308Phe/L-tyrosine. Data was interpreted using the program provided by MicroCal.

Circular dichroism

Protein was prepared as for assays and concentrated to 0.5 mM. All runs contained 10 mM phosphate pH 9.0 buffer and sometimes contained 10 mM KCl. A cuvette blank was run at 190-300 nm, with a 2 sec average time, for 1 scan at a 1 nm step. A buffer blank was run at 178-300 nm, 0.5 sec average time, for 3 scans and a 1 nm step (with and without KCl). Protein solutions were made with a final concentration of protein at 5 μM with buffer, with and without KCl, and were run with the same parameters as the buffer blank.

Crystallization and X-ray data collection

*Sg*TAM mutant constructs Glu71Ala and Tyr63Phe (both 58 kDA, 539 amino acids) were crystallized as described previously with some minor modifications (*6*). The Glu71Ala mutant was crystallized in 100 mM trimethylamine N-oxide (TMAO), 4.6 M sodium formate, at 4 ºC. The Tyr63Phe mutant crystallized in 95 mM TMAO, 4.6 M sodium formate with 2 mM L-tyrosine (incubated with protein for two days at 4 ºC), at 20 ºC. All stock solutions of crystallization conditions were homemade with reagents purchased from Aldrich. Solutions were sterile filtered prior to use. Crystals typically formed within two days and reached full size within a few weeks. Data was collected at the National Synchrotron Light Source at Brookhaven National Laboratories on beamlines X12C and X25, respectively (Table 2.2).

Table 2.2: Data collection and refinement statistics

	Tyr63Phe	**Glu71Ala**
Space group	$P2_12_12$	$P2_12_12$
Unit cell parameters (Å)	a=92.7,b=145.9,c=74.8	a=92.6,b=145.9,c=74.9
Data collection		
Resolution (Å)	50.0-2.20 (2.28-2.20)	50-2.20 (2.28-2.20)
Wavelength (Å)	0.9793	1
No. of reflections (measured/unique)	751714/52517	226063/50047
% Completeness*	99.9 (100.0)	95.2 (87.1)
R_{sym}*	0.087 (0.653)	0.086 (0.440)
I/σ*	30.0 (4.3)	11.0 (1.8)
Redundancy*	14.3 (14.2)	4.6 (4.0)
Refinement		
No. of reflections (total/test)	50153/5053	47693/4814
R_{work}	0.204	0.238
R_{test}	0.249	0.265
No. amino acids (chains A/B)	529/529	529/529
No. atoms		
Protein (chains A/B)	4012/4012	4010/4010
Ligands (L-tyrosine)	26	-
Water Molecules	216	468
Mean B (protein/Tyr/water) ($Å^2$)	42.9/30.8/37.8	39.3/-/38.3
RMSD from ideal geometry		
RMSD bonds (Å)	0.008	0.005
RMSD angles (^{o})	1.4	0.8
Ramachandran plot by PROCHECK (%)		
Core region	89.8	87.0
Allowed region	10.2	12.7
Generously allowed	0.0	0.2
Disallowed	0.0	0.0

*overall (highest resolution shell)

4.10 References

1. Hu, J. L., Xue, Y. C., Xie, M. Y., Zhang, R., Otani, T., Minami, Y., Yamada, Y., and Marunaka, T. (1988) A new macromolecular antitumor antibiotic, C-1027. I. Discovery, taxonomy of producing organism, fermentation and biological activity, *J. Antibiot. (Tokyo) 41*, 1575-1579.
2. Otani, T., Minami, Y., Marunaka, T., Zhang, R., and Xie, M. Y. (1988) A new macromolecular antitumor antibiotic, C-1027. II. Isolation and physico-chemical properties, *J. Antibiot. (Tokyo) 41*, 1580-1585.
3. Liu, W., Christenson, S. D., Standage, S., and Shen, B. (2002) Biosynthesis of the enediyne antitumor antibiotic C-1027, *Science 297*, 1170-1173.
4. Christenson, S. D., Liu, W., Toney, M. D., and Shen, B. (2003) A novel 4-methylideneimidazole-5-one-containing tyrosine aminomutase in enediyne antitumor antibiotic C-1027 biosynthesis, *J. Am. Chem. Soc. 125*, 6062-6063.
5. Christenson, S. D., Wu, W., Spies, M. A., Shen, B., and Toney, M. D. (2003) Kinetic analysis of the 4-methylideneimidazole-5-one-containing tyrosine aminomutase in enediyne antitumor antibiotic C-1027 biosynthesis, *Biochemistry 42*, 12708-12718.
6. Christianson, C. V., Montavon, T. J., Van Lanen, S. G., Shen, B., and Bruner, S. D. (2007) The structure of L-tyrosine 2,3-aminomutase from the C-1027 enediyne antitumor antibiotic biosynthetic pathway, *Biochemistry 46*, 7205-7214.
7. Van Lanen, S. G., Dorrestein, P. C., Christenson, S. D., Liu, W., Ju, J., Kelleher, N. L., and Shen, B. (2005) Biosynthesis of the beta-amino acid moiety of the enediyne antitumor antibiotic C-1027 featuring beta-amino acyl-S-carrier protein intermediates, *J. Am. Chem. Soc. 127*, 11594-11595.
8. Van Lanen, S. G., Lin, S., Dorrestein, P. C., Kelleher, N. L., and Shen, B. (2006) Substrate specificity of the adenylation enzyme SgcC1 involved in the biosynthesis of the enediyne antitumor antibiotic C-1027, *J. Biol. Chem. 281*, 29633-29640.
9. Lin, S., Van Lanen, S. G., and Shen, B. (2007) Regiospecific chlorination of (S)-beta-tyrosyl-S-carrier protein catalyzed by SgcC3 in the biosynthesis of the enediyne antitumor antibiotic C-1027, *J. Am. Chem. Soc. 129*, 12432-12438.
10. Lin, S., Van Lanen, S. G., and Shen, B. (2008) Characterization of the two-component, FAD-dependent monooxygenase SgcC that requires carrier protein-tethered substrates for the biosynthesis of the enediyne antitumor antibiotic C-1027, *J. Am. Chem. Soc. 130*, 6616-6623.
11. Lin, S., Van Lanen, S. G., and Shen, B. (2009) A free-standing condensation enzyme catalyzing ester bond formation in C-1027 biosynthesis, *Proc. Natl. Acad. Sci. U.S.A. 106*, 4183-4188.
12. Tabor, H. M., A. H.; Hayaishi, O.; White, J. (1952) Urocanic acid as an intermediate in the enzymatic conversion of histidine to glutamic and formic acids., *J. Biol. Chem. 196*, 121-128.

13. Hanson, K. R., and Havir, E. A. (1970) L-phenylalanine ammonia-lyase. IV. Evidence that the prosthetic group contains a dehydroalanyl residue and mechanism of action, *Arch. Biochem. Biophys. 141*, 1-17.
14. Kyndt, J. A., Meyer, T. E., Cusanovich, M. A., and Van Beeumen, J. J. (2002) Characterization of a bacterial tyrosine ammonia lyase, a biosynthetic enzyme for the photoactive yellow protein, *FEBS Lett. 512*, 240-244.
15. Walker, K. D., Klettke, K., Akiyama, T., and Croteau, R. (2004) Cloning, heterologous expression, and characterization of a phenylalanine aminomutase involved in Taxol biosynthesis, *J. Biol. Chem. 279*, 53947-53954.
16. Krug, D., and Muller, R. (2009) Discovery of additional members of the tyrosine aminomutase enzyme family and the mutational analysis of CmdF, *Chembiochem 10*, 741-750.
17. Magarvey, N. A., Fortin, P. D., Thomas, P. M., Kelleher, N. L., and Walsh, C. T. (2008) Gatekeeping versus promiscuity in the early stages of the andrimid biosynthetic assembly line, *ACS Chem. Biol. 3*, 542-554.
18. Peterkofsky, A. (1962) The mechanism of action of histidase: amino-enzyme formation and partial reactions, *J. Biol. Chem. 237*, 787-795.
19. Langer, M., Pauling, A., and Retey, J. (1995) The role of dehydroalanine in catalysis by Histidine Ammonia Lyase, *Angew Chem Int Ed Engl 34*, 1464-1465.
20. Calabrese, J. C., Jordan, D. B., Boodhoo, A., Sariaslani, S., and Vannelli, T. (2004) Crystal structure of phenylalanine ammonia lyase: multiple helix dipoles implicated in catalysis, *Biochemistry 43*, 11403-11416.
21. Williams, V. R., and Hiroms, J. M. (1967) Reversibility of the "irreversible" histidine ammonia-lyase reaction, *Biochim. Biophys. Acta 139*, 214-216.
22. Yamada, S., Nabe, K., Izuo, N., Nakamichi, K., and Chibata, I. (1981) Production of L-Phenylalanine from trans-Cinnamic Acid with Rhodotorula glutinis Containing L-Phenylalanine Ammonia-Lyase Activity, *Appl. Environ. Microbiol. 42*, 773-778.
23. Schuster, B., and Retey, J. (1995) The mechanism of action of phenylalanine ammonia-lyase: the role of prosthetic dehydroalanine, *Proc. Natl. Acad. Sci. U.S A. 92*, 8433-8437.
24. Christianson, C. V., Montavon, T. J., Festin, G. M., Cooke, H. A., Shen, B., and Bruner, S. D. (2007) The mechanism of MIO-based aminomutases in beta-amino acid biosynthesis, *J. Am. Chem. Soc. 129*, 15744-15745.
25. Watts, K. T., Mijts, B. N., Lee, P. C., Manning, A. J., and Schmidt-Dannert, C. (2006) Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family, *Chem. Biol. 13*, 1317-1326.
26. Louie, G. V., Bowman, M. E., Moffitt, M. C., Baiga, T. J., Moore, B. S., and Noel, J. P. (2006) Structural determinants and modulation of substrate specificity in phenylalanine-tyrosine ammonia-lyases, *Chem. Biol. 13*, 1327-1338.
27. Montavon, T. J. *Unpublished data*.
28. Baedeker, M., and Schulz, G. E. (2002) Structures of two histidine ammonia-lyase modifications and implications for the catalytic mechanism, *Eur. J. Biochem. 269*, 1790-1797.

29. Schroeder, A. C., Kumaran, S., Hicks, L. M., Cahoon, R. E., Halls, C., Yu, O., and Jez, J. M. (2008) Contributions of conserved serine and tyrosine residues to catalysis, ligand binding, and cofactor processing in the active site of tyrosine ammonia lyase, *Phytochemistry 69*, 1496-1506.

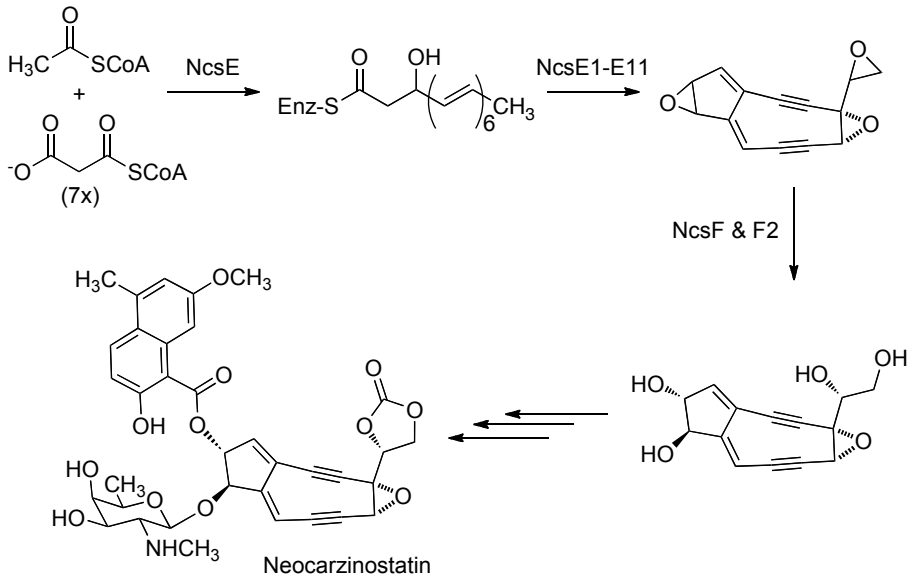
Appendix 1

Discussion and Experimentals for the NcsEs and for the Synthesis of the Benzoxazolinate Moiety of C-1027

A1. 1 Biosynthesis of the neocarzinostatin enediyne core

The carbon scaffold of the enediyne core is produced by the polyketide synthase (PKS) NcsE, which iteratively catalyzes seven Claisen condensations as well as ketoreductions and dehydrations to yield the enzyme bound hexadecahexaene product (Scheme A1.1) (*1*). The thioesterase (TE) NcsE10 hydrolyzes the linear polyene, which is then decarboxylated and dehydrated (*1*). Subsequent transformations, likely catalyzed by all or some of the remaining NcsEs, result in the an enediyne core intermediate (*2*). The activity of these enzymes has yet to be determined as many of the NcsEs are homologous only to proteins of unknown function. We sought to elucidate the function of some of these unknown proteins using X-ray crystallography and comparing their structures to structures of enzymes of known function.

Scheme A1.1: Biosynthesis of enediyne core of neocarzinostatin by the NcsEs



A1.2 Experimentals for NcsEs

Cloning of ncsE2 *and* ncsE3

Both *ncsE2* and *ncsE3* were cloned from genomic DNA into pET-30b by T. Montavon. A 2 bp frameshift required them to be recloned into pET-30a. This was done by a sequential digest of the vector using BamHI and HindIII, purification on a 1% agarose gel, and excision of the DNA from the gel, which was purified using the QIAquick DNA purification kit (QIAGEN, Valencia, CA). The DNA was ligated into the new plasmid using T4 DNA ligase (NEB, Ipswich, MA) and the mixture was transformed into TOP10 cells (Invitrogen, Carlsbad, CA). Successful clones were confirmed by double digest using NcoI and HindII, resulting in the correct digest pattern (Figure A1.1), and by sequencing (GENEWIZ, S. Plainfield, NJ).

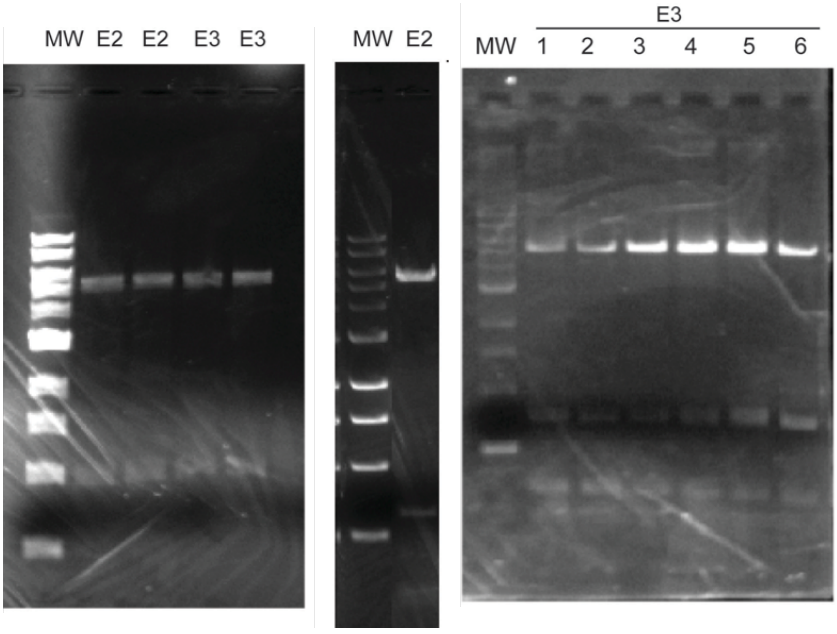


Figure A1.1: Cloning of *ncsE2* and *ncsE3*. Double digest of inserts from pET-30b (left) and double digest of inserts in new expression vector, pET-30a (middle and left).

NcsE2 expression tests

The *ncsE2*-pET-30a plasmid was transformed into BL21(DE3) cells (Invitrogen) and one colony was used to inoculate 5 mL of LB broth (EMD) supplemented with 30 μg/mL kanamycin (Aldrich, St. Louis, MO.). The culture was shaken at 37 ºC at 250 rpm to OD_{600} = 0.60 and 2 mL of culture was transferred to a new tube containing 100 μM IPTG (Aldrich). Both tubes were shaken for 1 h at 37 ºC. Various amounts of each culture were run on an SDS-PAGE gel to confirm overexpression of induced culture (Figure A1.2). A 1 L culture was grown at 37 ºC and 150 rpm to OD_{600} = 0.70 and overexpression was induced at 18 ºC with 100 μM IPTG. Cells were shaken at this temperature for 16 h. Following centrifugation at 2,000 g for 20 min, the cells were resuspended in 20 mL lysis buffer (20 mM Tris-HCl and 500 mM NaCl, pH 7.5), frozen to -70 ºC. The cells were lysed and the protein was purified using same procedure for NcsB2 (see section 3.9 Experimentals). Gel showed a significant amount of overexpression of NcsE2 in inclusion bodies (Figure A1.3). Overexpression at 16 ºC did not have a positive effect on the solubilization of NcsE2 and, in fact, appeared to result in a decreased amount of expression (Figure A1.3).

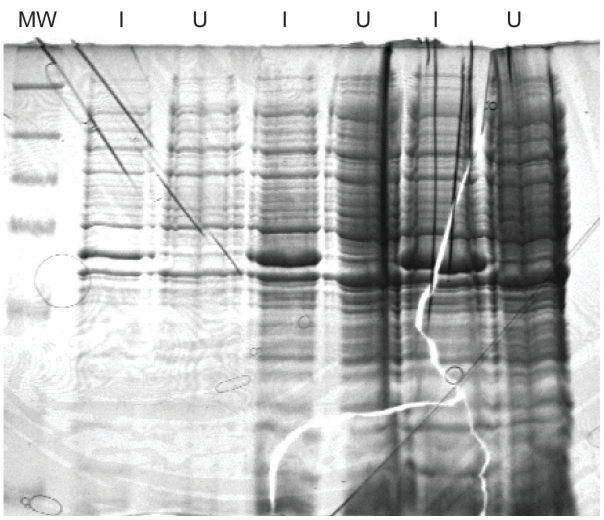


Figure A1.2: Small-scale over-expression test of NcsE2 showing induced (I) and uninduced (U) cultures.

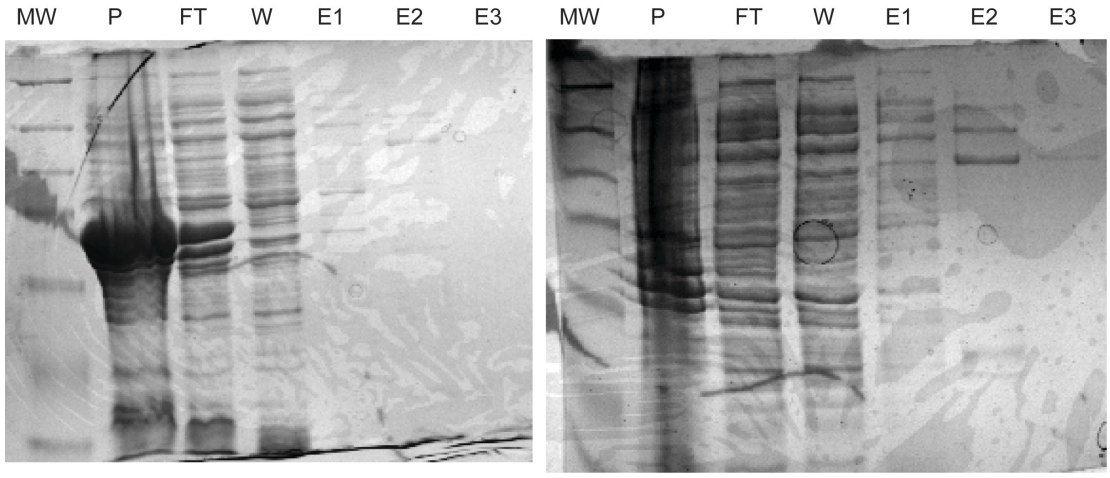


Figure A1.3: SDS-PAGE gels of lysed NcsE2 cell pellets subjected to Ni-NTA purification. Gels are of NcsE2 over-expression at 18 °C (left) and at 16 °C (right). Each gel shows aliquots taken from pellet (P), flow-through (FT), wash (W) and elutions 1-3 (E1-E3) during purification.

We attempted to purify NcsE2 under denaturing conditions. The pellet was resuspended in 15 mL buffer A (6 M Gu-HCl, 0.1M NaH_2PO_4, 0.01 M Tris-HCl pH 8.0). The slurry was stirred at 25 ºC for 2 h. To pellet insolubilized portion, the slurry was centrifuged at 10,000 *g* for 30 min at 4 ºC. Ni-NTA resin was washed with buffer A and was incubated with supernatant for 1 h at 25 ºC. The resin was spun down at 2,000 *g* for 5 min. The supernatant was poured and the resin was resuspended in buffer A and washed 4 x 10 mL with the same buffer. In an attempt to renature the solubilized protein, a gradient maker was used to exchange buffer A for a non-denaturing buffer (20 mM Tris-HCl, 100 mM NaCl, 1 mM β-mercaptoethanol, pH 7.5). About 80 mL of each was placed in either side of the gradient maker, which went directly onto the column containing the Ni-NTA resin and protein. Following gradient exchange, a 10 mL elution off of the Ni-NTA column was collected (elution buffer: 20 mM Tris pH 7.5, 500 mM NaCl, 250 mM imidazole, 1 mM β-mercaptoethanol). No protein was obtained.

Finally, a freshly lysed pellet was resuspended in a denaturing buffer containing 4 M urea (Figure A1.4). Denatured protein was successfully obtained but renaturing by dialyzing into standard low salt buffer with Tween-20 (at the critical micelle concentration, 0.005%) resulted in nearly all of the protein precipitating out of solution. It seems unlikely that this protein could be purified under standard conditions, even with the help of refolding screens. Because its activity is unknown, determining if proper refolding into its native conformation has occurred would be quite difficult. At this point, we began to focus more on the isolation and purification of NcsE3.

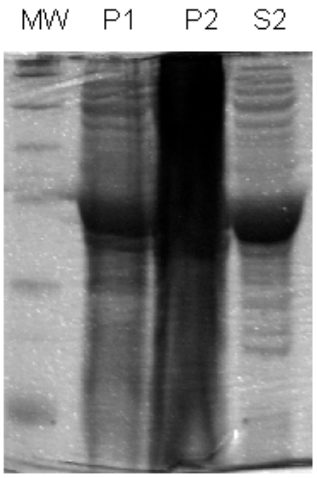


Figure A1.4: SDS-PAGE gel of solubilized, denatured NcsE2. Lanes are pellet prior to denaturing (P1), pellet following denaturing (P2), and supernatant after denaturing (S2).

NcsE3 expression tests

A 5 mL expression test was run as for NcsE2, but overexpression was not apparent (Figure A1.5). Therefore, a large-scale 1 L prep was set up and overexpression was induced with 100 μM IPTG at 18 ºC. Using standard purification conditions, soluble protein was obtained in good yield (Figure A1.5). NcsE3 is a 328 amino acid protein with a molecular weight of 35711.3 Da and a calculated pI of 9.19 (E_c = 59025 M^{-1}). With the His_6-tag, it is 377 amino acids long with a molecular weight of 41012.0 and a pI of 7.88 (E_c = 59025 M^{-1}). Even with the affinity tag, NcsE3 has a very high pI. In general, soluble proteins have a pI below 7, and often around 5. It is not surprising that upon dialysis into low salt buffer at a neutral pH (50 mM NaCl, 20 mM Tris-HCl, 1 mM

β-mercaptoethanol, pH 7.5), protein began to precipitate out. It became apparent that a high concentration of salt was needed to stabilize the protein in a soluble form. A new protein preparation was run and the protein was purified under standard conditions and dialyzed into an ammonium sulfate buffer (1 mM NH_4SO_4, 50 mM NaH_2PO_4, 1 mM β-mercaptoethanol, pH 7.0). Some protein precipitated but most remained soluble. Dialysis studies to determine optimal salt concentration were run. First, the purified protein was dialyzed for 2 h into 1 L of high salt buffer (500 mM NaCl, 20 mM Tris, 1 mM β-mercaptoethanol, pH 7.5). It was then transferred to 300 mM NaCl dialysis buffer for 2 h, then to 200 mM NaCl dialysis buffer for 2 h, followed by 100 mM for 16 h. The bulk of protein precipitation occurred between 100 and 200 mM NaCl. It seemed unlikely that standard purification methods could be used (i.e. ion-exchange on a Mono-Q column).

Therefore, following dialysis into the ammonium sulfate buffer, NcsE3 was subjected to further purification on a hydrophobic-interaction column (HIC) (phenyl sepharose column from GE Healthcare) using the following buffer system: A- 1 M NH_4SO_4, 50 mM NaH_2PO_4, 1 mM β-mercaptoethanol, pH 7.0; B- 50 mM NH_4SO_4, 50 mM NaH_2PO_4, 1 mM β-mercaptoethanol, pH 7.0) at 2 mL a minute for 40 mL total, 0-100 % B. Very pure protein resulted (Figure A1.5).

The HIC method was further improved by dialyzing into a higher pH buffer (50 mM Tris-HCl, 1 mM NH_4SO_4, 1 mM β-mercaptoethanol, pH 8.8 with 10% glycerol). The HIC purification was run under pH 8.8, with a 30 mL 0-100% B gradient followed by a 10 mL wash at 100% B, which resulted in better elution of protein off the column.

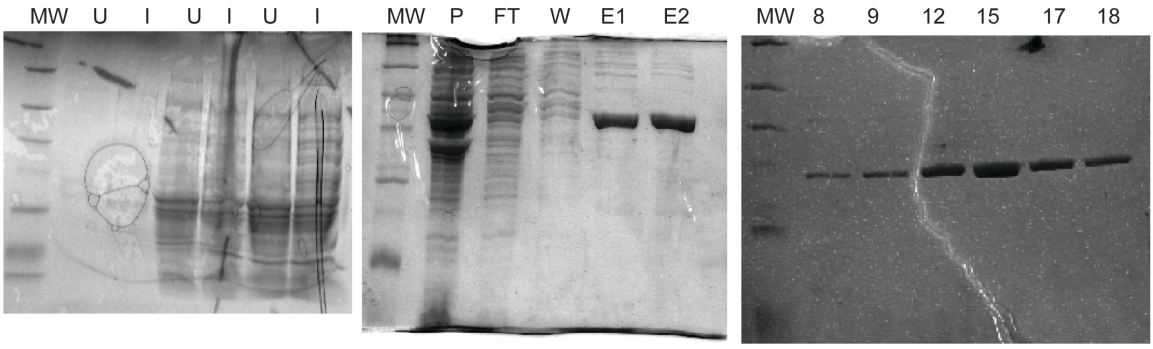


Figure A1.5: SDS-PAGE gels of overexpression and purification of NcsE3. A gel of the small-scale over-expression test at 37 ^{o}C is shown (left) as well as a gel of lysis and Ni-NTA purification of large-scale over-expression (middle). HIC purification results in very pure protein, as observed in gel (right).

NcsE2 was concentrated to under 20 mL and dialyzed into a 250 mM NaCl buffer, followed by concentration to 10 mg/mL. Hampton crystal screens were set up at both 4 and 20 ºC at 10 mg/mL, but most resulted in a large amount of precipitated protein, which could not be lessened with addition of 0.1% Triton x-100. Many false crystal hits resulted, as ammonium sulfate is known to form salt crystals in the presence of divalent cations, in this case in the presence of Ca^{2+} and Mg^{2+}. Attempts to remove ammonium sulfate were made by dialyzing into high-salt gel filtration buffer and by purifying on a size-exclusion column prior to screening using a Superdex 200 gel filtration column (buffer- 20 mM Tris-HCl, 1 M NaCl, 1 mM β-mercaptoethanol, 10% glycerol, pH 8.8). Typical protein yields from all of the above preps were about 5-6 mg protein per liter prep.

A final attempt was made at improving the solubility of NcsE3 without using exceedingly high salt concentrations by dialyzing into another common detergent solution: 1.5 mM deoxycholic acid, 20 mM Tris-HCl, 50 mM NaCl, 1 mM β-

mercaptoethanol, pH 7.5. Nearly 100% of previously soluble protein precipitated out of solution.

Finally, NcsE3 was purified using only Ni-NTA resin and size exclusion chromatography (buffer contained 250 mM NaCl, 20 mM Tris pH 7.5, 1 mM β-mercaptoethanol). Crystal screens were set up at 4 and 20 ºC. Crystal screens did not result in 100% crashed out protein but higher salt concentrations maintained solubility. Trays containing PEG 400 and sodium citrate were set up (30% PEG 400 and 0.2 M sodium citrate were identified as conditions that allowed good solubility) as well as conditions containing a variety of precipitants (MPD, PEG 1000 and 4000), all at pH 7.5. No crystal hits resulted.

An NcsE3 solubility study with 100 mM metal solutions was run because a blast search reveled some remote homology to metal utilizing enzymes. Metals included: nickel(II) sulfate, ferrous(II) sulfate, copper(II) sulfate, ferric(III) chloride, cobalt(II) chloride. All solutions resulted in precipitated protein after 30 min at room temperature.

Attempts to improve solubility of NcsE2 and NcsE3 during expression

Maltose-binding protien (MBP) constructs of misfolded proteins have been shown to improve overall solubility. Therefore, both *ncsE2* and *ncsE3* were cloned into the pMAL-C2X expression vector at the BamHI and HindIII site (Figure A1.6).

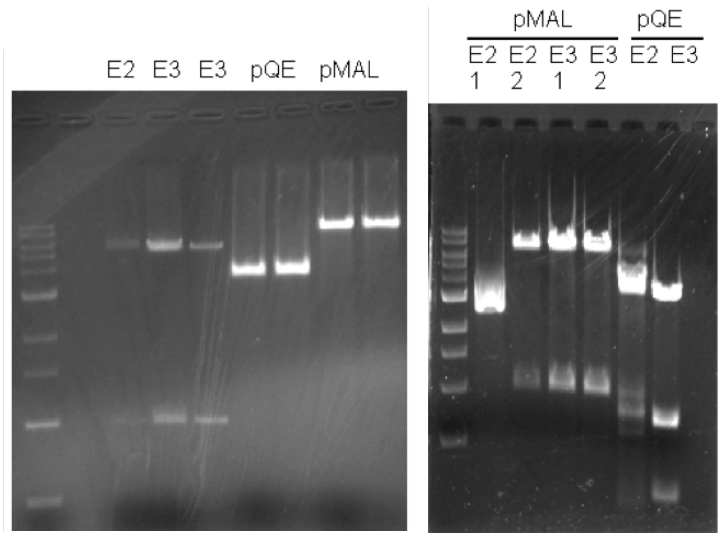


Figure A1.6: Cloning of NcsE2 and NcsE3 into various expression vectors. (Left) Double digest of *ncsE2* and *ncsE3* and pQE and pMAL plasmids and (right) double digest of DNA extracted from colonies following transformation of ligation mixture into *E coli.* The last five lanes show successful ligations.

A small-scale over-expression test was run (as previously) and it was apparent that both pMAL constructs over-expressed (Figure A1.7).

Also, because NcsE3 is significantly more soluble than NcsE2, which is expressed in inclusion bodies, and because NcsE3 and NcsE2 are 41% identical and 57% similar, we thought that perhaps the two enzymes are coexpressed *in vivo*. Coexpression of the two constructs in *E. coli* might result in an overall soluble protein heterodimer. Therefore, *ncsE2* and *ncsE3* were cloned into the pQE expression vector, which contains no affinity tag and is ampicillin resistant (pET-30a is kanamycin resistant) (Figure A1.6). Cotransformation of the two plasmids and double selection using both antibiotics would result in colonies containing both expression plasmids.

An initial small-scale overexpression test was run, which included not only coexpression of the His_6-tagged constructs with the untagged pQE contructs, but also

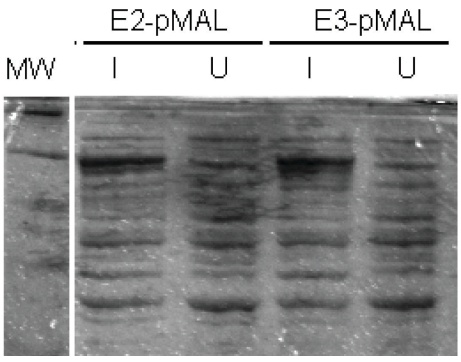


Figure A1.7: Small scale overexpression test of pMAL constructs of NcsE2 and NcsE3 showing induced (I) and uninduced (U) cultures.

coexpression of the His_6-tagged constructs with the pMAL constructs (Figure A1.8). There was a difference between the induced and uninduced cultures but it was unclear from this test if coexpression was advantageous to solubility. Large-scale preps were set up with coexpression of the His_6-tagged constructs and the pQE constructs. Unfortunately, not only did coexpression of the two proteins not increase the solubility of NcsE2, it appeared that the solubility (and perhaps overexpression) of NcsE3 suffered remarkably.

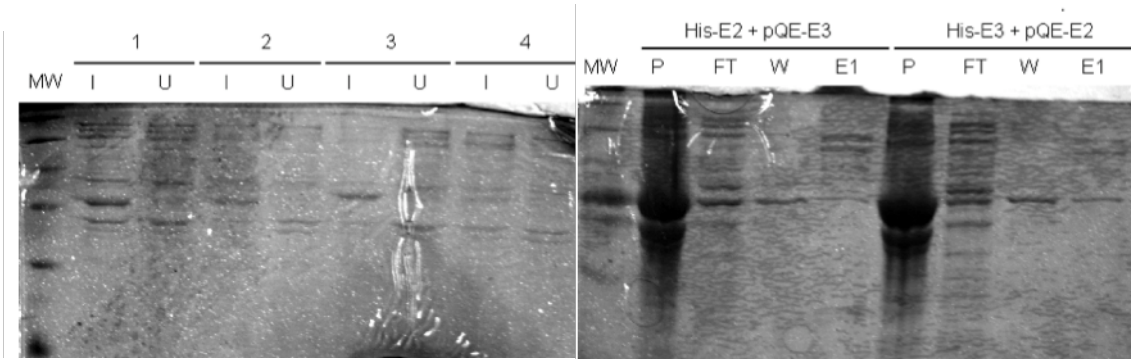


Figure A1.8: SDS-PAGE gels of coexpression of plasmids. Small scale (left), including pET-30a-*ncsE2* and pQE-60-*ncsE3* (1), pET-30a-*ncsE2* and pMAL-c2x-*ncsE3* (2), pET-30a-*ncsE3* and pQE-60-*ncsE2* (3), pET-30a-*ncsE3* + pMAL-c2x-*ncsE2* (4) induced (I) and uninduced (U). Gel of large scale prep (right) shows that solubility is not improved by coexpression, which is apparent from a large amount of accumulated protein in pellet.

Cloning and overexpression of ncsE4

We next sought to clone *ncsE4* from genomic DNA as it is also highly conserved among enediyne producers and many other organisms, yet has an unknown function. Prior to cloning, we analyzed it using the Dense Alignment Surface Method (DAS) (*3*) to predict any transmembrane domains that would lead to issues with soluble overexpression (Figure A1.9). The first 25 or so amino acid residues appear to be embedded in a membrane and we predicted that truncating *ncsE4* would give us soluble protein. Therefore, we cloned both the entire gene and a truncated construct (lacking the first 25 amino acids). The full NcsE4 construct would express as a 636 amino acid protein, with a molecular weight of 68852.2 Da, with a predicted pI of 4.71, and an E_c of

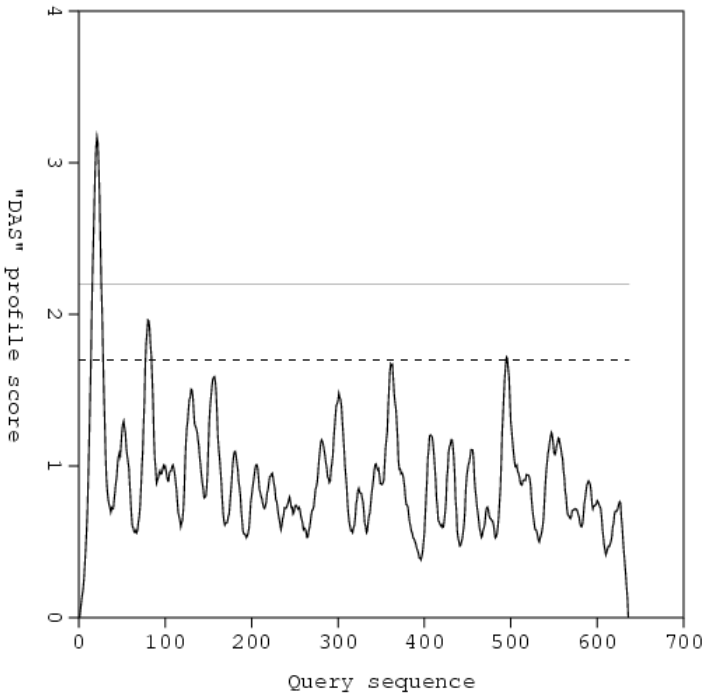


Figure A1.9: DAS transmembrane segment prediction of NcsE4 with a loose cutoff (-----) and a strict cutoff (——).

100,505 M^{-1}. The truncated construct would be 611 amino acids in length, with a molecular weight of 66093.9 Da, a pI of 4.64 and an E_c of 95, 005 M^{-1}.

Two sets of primers (from IDT DNA, Coralville, IA) were used to clone *ncsE4* from genomic DNA into the expression vector pET-28a. The first set of primers was forward 5'-CCG TTG CCT CCC GGA ATG-3' and reverse 5'-ACG GAG ATC GAG ATC GCG-3' and were used to clone a 2100 bp piece of DNA from genomic DNA previously isolated from *S. carzinostaticus*. Three dilutions of genomic DNA were made up by diluting the stock by 1/10, 1/100 and 1/1000 in water. PCR solutions were 100 μl and contained 10% DMSO, 1x Deep Vent Polymerase buffer, 2 mM $MgSO_4$, 20 mM dNTPS, 0.1 nmol of each primer, 1 μl of DNA dilution, 2 U of Deep Vent R Polymerase (all reagents were from NEB, Ipswich, MA) and water. The time program used was as follows: 94 ºC for 5 min, then 30 cycles of 94 ºC for 1 min, 1 min at 51.4, 54.3 or 57.5 ºC, and 1 min at 72 ºC, with a final extension time of 10 min at 72 ºC. Each genomic DNA was set up at each annealing temperature. Following PCR, the reactions were run on a 1% agarose gel (Figure A1.10). DNA from lane three of the 1/100 genomic DNA dilution was excised from the gel and purified using the QIAquick DNA purification kit (QIAGEN, Valencia, CA).

Another PCR was set up with the same reagents as above but with nested primers containing the proper restriction sites for cloning. Two different forward primers were used to clone either the entire gene of *ncsE4* or the truncated construct of it. The primers were forward (full length) 5'-GCA GCT AGC ACC ATG GCG AAG AAC-3', forward (truncated) 5'-GAT GCT AGC TAC GTC GTT CGC CTC-3' and reverse 5'-CGC CTC

GAG TCA CTT CTC CTT GAC CTG-3'. The underlined codons represent restriction sites for NheI (forward) and XhoI (reverse). The time program was the same as for the first set of primers but with a variety of different annealing times. The successful annealing time for *ncsE4* was 56.2 ºC and for *ncsE4t* (truncated construct) was 60.2 ºC. The DNA was purified as above and double digested using the two restriction enzymes per standard NEB protocol along with pET-28a. Successful cloning of both constructs was confirmed by double digest and agarose gel electrophoresis (Figure A1.10). Plasmids were transformed into BL21(DE3).

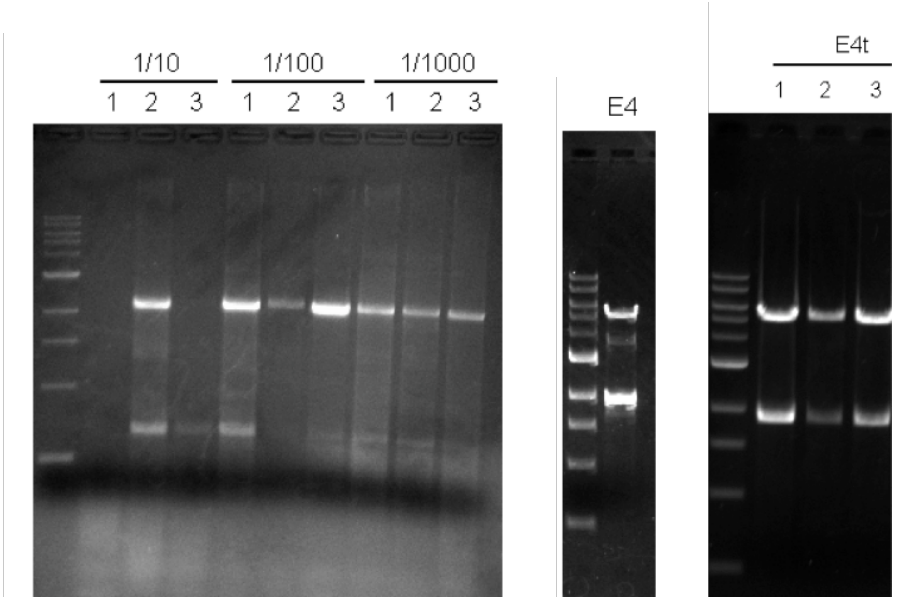


Figure A.10: Cloning of *ncsE4* from genomic DNA. First round of PCR produced 2100 bp DNA fragment (left). Each column is a different annealing temperature (1=51.4 ^{o}C, 2=54.3 ^{o}C, 3=57.5 ^{o}C) at three dilutions of genomic DNA. Second round of PCR using nested primers for full length and truncated constructs, followed by purification, double digest and ligation into pET-28a, which were confirmed by 1% agarose gel (middle, *ncsE4* and right, *ncsE4t*).

A small-scale expression test was run on both constructs of NcsE4 (Figure A1.11). From this simple test, it was clear that the truncated construct (NcsE4t) overexpressed better than the full-length construct. Therefore, large-scale preps of NcsE4t were set up. The first prep included growth of cells at 37 ºC and induction of overexpression at 16 ºC, which resulted in little to no soluble protein following purification on Ni-NTA resin (Figure A1.12). Next, an attempt to improve the solubility of NcsE4t was made by growing cells at 15 ºC and inducing overexpression at the same temperature after three days. The cells were shaken for an additional 16 h. No soluble protein resulted as all was present in inclusion bodies.

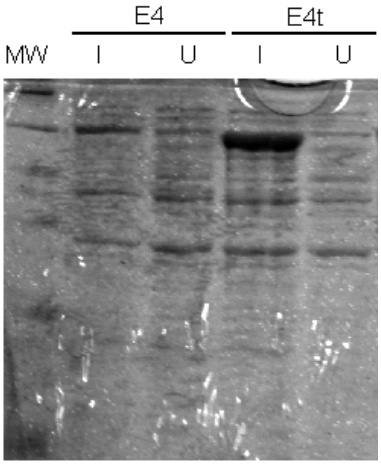


Figure A1.11: Overexpression of full length (E4) and truncated (E4t) constructs of NcsE4.

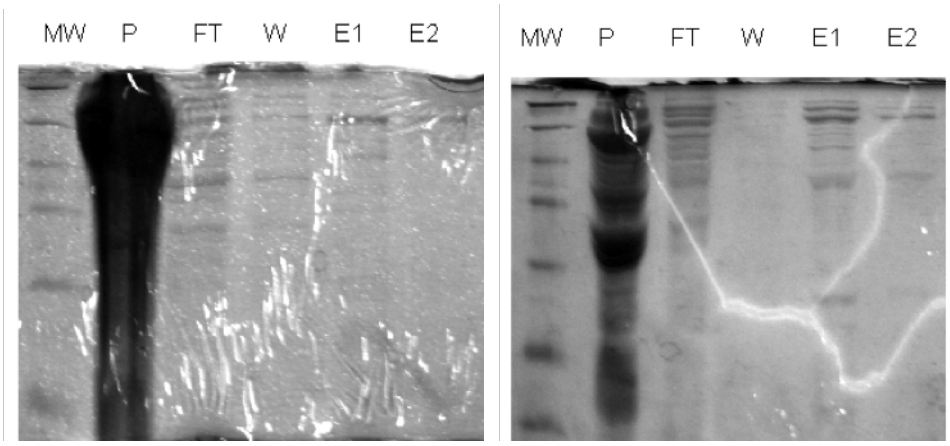


Figure A1.12: Large-scale overexpression of NcsE4t. Growth at 37 ^{o}C with 16 ^{o}C induction (left) and growth and induction at 15 ^{o}C (right). Both purified using Ni-NTA resin.

A1.3 Benzoxazolinate moiety

In an effort to study the convergent biosynthesis of C-1027, we sought to synthesize benzoxazolinate (**A1.1**) from C-1027. The benzoxazolinate methyl-ester (**A1.2**) has previously been synthesized by Shibuya and coworkers (*4*). We intended to use benzoxazolinate in studies with the predicted aryl-transferase SgcD6. Unfortunately, overexpression of SgcD6 was never realized, as it is a transmembrane protein consisting of six transmembrane regions. Whether or not the gene encoding for SgcD6 is directly involved in the biosynthesis of C-1027 is debatable. Nevertheless, benzoxazolinate was synthesized using the published procedure as a guideline, with some minor modifications and additions. Benzoxazolinate was tested as a potential substrate for NcsB2.

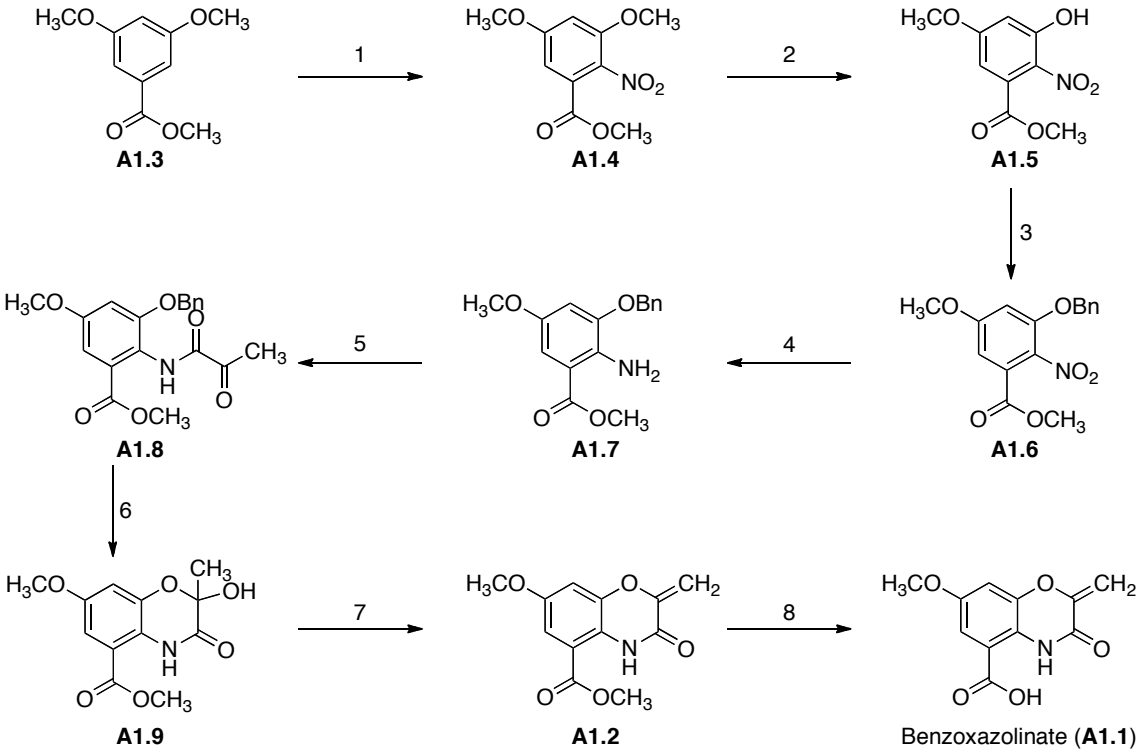


Figure A1.13: Synthesis of benzoxazolinate (**A1.1**) (1) $Cu(NO_3)_2 \cdot H_2O$, Ac_2O, 79%, (2) BCl_3, CH_2Cl_2, -78 °C to 4 °C, 95%, (3) BnBr, NaH, reflux, 93%, (4) $SnCl_2 \cdot H_2O$ (5 eq), EtOAc, 78 °C, 61%, (5) DCC (4 eq), HOBt, DMAP, (5 mol %), Pyruvic acid, NMP, 66%, (6) H_2, Pd/C (10 mol%), CH_3OH, 89%, (7) MsCl (3 eq), Et_3N (5 eq), DMAP (4 mol %), 58%, (8) LiOH, H_2O/THF, 71%.

Experimentals for benzoxazolinate synthesis

General. Chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, MO) or VWR (West Chester, PA). Solvents were dried using stills by Glass Contour (Santa Monica, CA). Deuterated solvents were purchased from Cambridge Isotopes (Andover, MA). Silica gel used for all flash chromatography purifications was purchased from VWR, EMD brand Geduran. Mass Spectrometry (MS) was performed on a Micromass LCT TOF. ^{1}H NMR spectra were recorded on a Varian 400 or 300 MHz spectrometer. Proton chemical shifts are reported in ppm (δ) relative to internal tetramethylsilane (TMS, δ 0.0 ppm). Data are reported as follows: chemical shift (multiplicity [singlet (s), doublet (d) or multiplet (m)], coupling constants [Hz], integration).

Methyl 3,5-dimethoxy-2-nitrobenzoate (A1.4). Methyl 3,5-dimethoxy benzoate (5.81 g, 26.6 mmol) was added to a round-bottom flask containing 50 mL acetic anhydride. The clear solution was cooled to 0 ºC in an ice bath. Cupric nitrate (7.97 g, 33.0 mmol) was added to the solution over a 1 h. The clear blue solution was slowly warmed to 25 ºC and stirred for 3 h. The reaction mixture was poured over crushed ice to quench and was extracted with ethyl acetate (3 x 100 mL). The organic layer was washed with a saturated bicarbonate solution (2 x 250 mL) and brine (2 x 10 mL), dried with magnesium sulfate, filtered using a Buchner funnel and then concentrated. The resulting residue was purified via flash column chromatography (6 cm column, 1:1 hexanes-ethyl acetate), yielding the nitrated product (**A1.4**) (5.61 g, 78.6% yield). **Rf** = 0.61. **^{1}H NMR**

(400 MHz, $CDCl_3$) δ: 7.00 (d, *J* = 2.8 Hz, 1H), 6.71 (d, *J* = 2.8 Hz, 1H), 3.89 (s, 3H), 3.88 (s, 6H).

Methyl 3-hydroxy-5-methoxy-2-nitrobenzoate (A1.5). A1.4 (5.61 g, 23.3 mmol) was dissolved in 300 mL dry DCM in a round-bottom flask. The solution was cooled to -78 ºC and a 1 M solution of boron trichloride (47 mL, 46.6 mmol) was added dropwise. The clear, orange solution was warmed to 0 ºC and stirred overnight. The brown solution was slowly quenched with 10 mL methanol, followed by 10 mL water and was concentrated. The brown residue was dissolved in 100 mL DCM and was washed with 1 M HCl (2 x 100 mL) and with brine (1 x 100 mL). The solution was dried with magnesium sulfate, filtered thru a silica plug, and concentrated under reduced pressure, yielding the monodemethylated product (**A1.5**) (4.97 g, 95% yield). **^{1}H NMR (400 MHz, DMSO) δ:** 11.39 (s, 1H), 6.84 (d, *J* = 2.4 Hz, 1H), 6.75 (d, *J* = 2.4 Hz, 1H), 3.82 (s, 3H), 3.80 (s, 3H).

Methyl 3-(benzyloxy)-5-methoxy-2-nitrobenzoate (A1.6). A1.5 (4.97 g, 22.0 mmol) was dissolved in 200 mL dry DMF in a round-bottom flask, which was cooled to 0 ºC in an ice bath under an inert atmosphere. Sodium hydride (0.97 g, 24.2 mmol) was added to solution and was stirred for 5 min. Benzyl bromide (287 mL, 24.2 mmol) was added to the solution, which was slowly warmed to 25 ºC and allowed to stir overnight. The reaction was quenched first with a small amount of methanol, followed by 10 mL brine. The solution was extracted with DCM (4 x 150 mL), which was washed with brine (1 x

100 mL). The solution was concentrated under reduced pressure with slight heating to remove the DMF. Mineral oil was removed by mixing with hexanes followed by decanting. The final product was recrystallized in ethyl acetate (15 mL). Crystals were washed with hexanes to yield benzyl-protected product (**A1.6**) (6.46 g, 92.5% yield). **^{1}H NMR (400 MHz, DMSO) δ:** 7.42 (m, 5H), 7.20 (d, *J* = 2.4 Hz, 1H), 7.00 (d, *J* = 2.4 Hz, 1H), 5.33 (s, 2H), 3.88 (s, 3H), 3.83 (s, 3H).

Methyl 2-amino-3-(benzyloxy)-5-methoxybenzoate (A1.7). A1.6 (6.46 g, 20.4 mmol) was added to a round-bottom flask containing 100 mL ethyl acetate. Tin (II) chloride (23.0, 0.102 mol) was added to the clear solution, which was refluxed at 78 ºC overnight. The reaction was cooled to room temperature and poured over ice. The pH was adjusted to 7 with saturated sodium bicarbonate solution. The solution was concentrated under reduced pressure to remove the ethyl acetate. To remove the tin, the aqueous solution was mixed with 100 mL DCM and the biphasic solution was filtered through a Buchner funnel. The organic layer was separated from the aqueous layer, the latter of which was extracted with DCM (2 x 50 mL). The organic extracts were combined, dried with magnesium sulfate, and concentrated under reduced pressure. The resulting residue was purified via flash column chromatography (5 cm column, 3:1 hexanes-ethyl acetate), yielding the reduced product (**A1.7**) (3.59 g, 61.3% yield). **^{1}H NMR (400 MHz, $CDCl_3$) δ:** 7.40 (m, 5H), 6.94 (d, *J* = 2.8 Hz, 1H), 6.64 (d, *J* = 2.8 Hz, 1H), 5.77 (s, b, 1H), 5.07 (s, 2H), 3.88 (s, 3H), 3.75 (s, 3H).

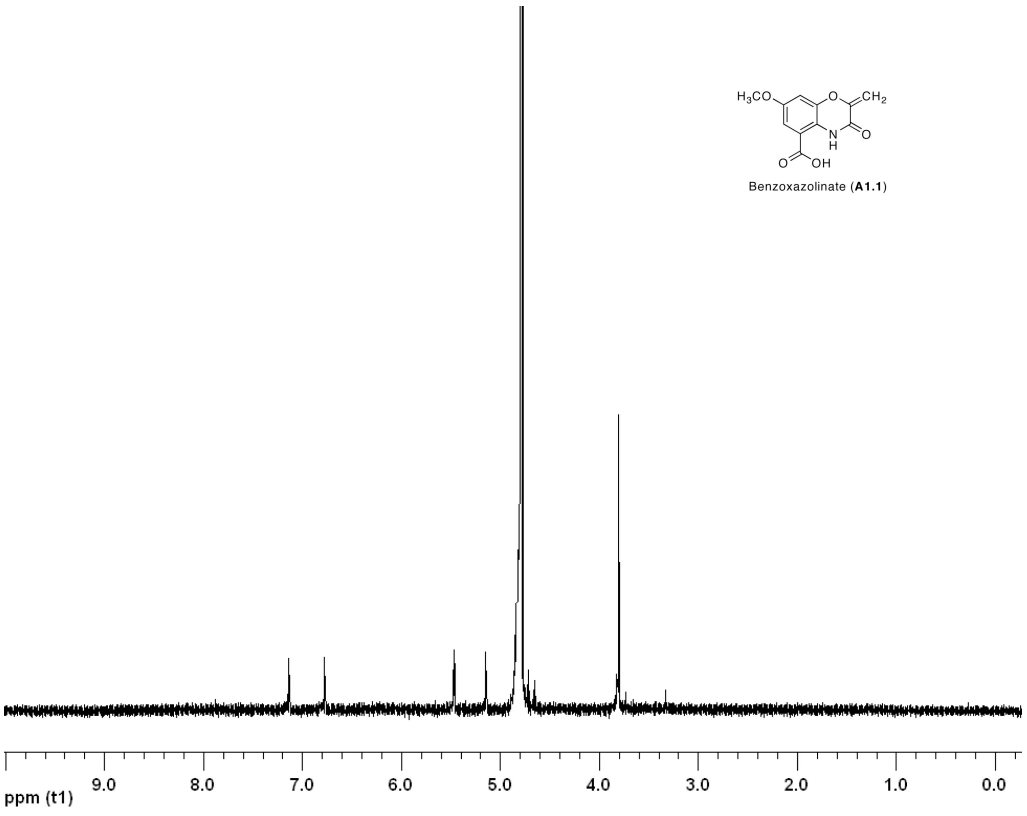
Pyruvyl aminobenzoate (A1.8). A1.7 (1.92 g, 6.67 mmol) was added to a round-bottom flask containing *N*,*N*-dicyclohexylcarbodiimide (5.51 g, 26.7 mmol), HOBt (3.61 g, 26.7 mmol) and DMAP (18 mg, 10 mol %). 30 mL NMP was added to flask, which was cooled to 0 ºC. Pyruvic acid (1.86 mL, 26.7) was added to stirring, orange slurry. The solution was warmed to 25 ºC and was stirred overnight. The reaction mixture was diluted with 50 mL of ethyl acetate, washed with saturate sodium bicarbonate solution (2 x 50 mL) and brine (1 x 50 mL). The organic layer was filtered to remove the insoluble dicyclohexylurea and was concentrated under reduced pressure. The crude product was purified via flash column chromatography (7:3 hexanes-ethyl acetate), to yield **A1.8** (1.58 g, 66.3% yield). **^{1}H NMR (400 MHz, $CDCl_3$) δ:** 9.24 (s, 1H), 7.40 (m, 5H), 6.94 (d, *J* = 2.8 Hz, 1H), 6.64 (d, *J* = 2.8 Hz, 1H), 5.77 (s, b, 1H), 5.07 (s, 2H), 3.88 (s, 3H), 3.75 (s, 3H).

Hydroxy benzoxazolinate (A1.9). A1.8 (1.56 g, 4.36 mmol) was dissolved in 30 mL methanol in a round-bottom flask. Pd(OH)/C (0.067 g) was added to flask. The flask was flushed with hydrogen three times and stirred under a hydrogen atomosphere overnight. The slurry was filtered through a pad of celite and washed with ethyl acetate and toluene. The solution was concentrated under reduced pressure, yielding **A.19** (1.03 g, 88.5%). **^{1}H NMR (300 MHz, DMSO) δ:** 9.95 (s, 1H), 7.74 (s, 1H), 7.08 (d, *J* = 2.4 Hz, 1H), 6.94 (d, *J* = 2.4 Hz, 1H), 3.89 (s, 3H), 3.76 (s, 3H), 1.65 (s, 3H).

Methyl benzoxazolinate (A1.2). A1.9 (0.686 g, 2.57 mmol) was dissolved in 30 mL dry DCM in a 50 mL round-bottom flask. Diisopropylethylamine (2.24 mL, 12.9 mmol) and DMAP (10 mg) were added to the yellow solution, which was stirred under an inert atmosphere in an ice bath. Mesyl chloride (0.598 mL, 7.70 mmol) was added dropwise, turning the solution brown. The solution was warmed to 25 °C and was stirred overnight. The reaction mixture was washed with saturate sodium bicarbonate solution (3 x 50 mL) and brine (1 x 50 mL). The organic layer was dried using magnesium sulfate, filtered and concentrated under reduced pressure. The crude product was purified via flash column chromatography (7:3 hexanes-ethyl acetate), to yield **A1.2** (0.371 g, 58.0% yield). **^{1}H NMR (400 MHz, $CDCl_3$) δ:** 10.41 (s, 1H), 7.18 (d, *J* = 2.8 Hz, 1H), 6.79 (d, *J* = 2.8 Hz, 1H), 5.64 (d, *J* = 2, 1H), 5.08 (d, *J* = 2, 1H), 3.94 (s, 3H), 3.80 (s, 3H).

Benzoxazolinate (A1.1). A1.2 (0.373 g, 1.50 mmol) was dissolved in 12 mL THF and 12 mL 2 M sodium hydroxide solution in a 50 mL round-bottom flask. The yellow solution was refluxed overnight. The reaction mixture was concentrated under reduced pressure to remove the THF and was extracted with DCM (2 x 15 mL) to remove unreacted starting material. The aqueous layer was acidified to pH less than 2 and was extracted with DCM (3 x 15 mL). Insoluble material was filtered and saved. The organic layer was dried using magnesium sulfate, filtered, concentrated under reduced pressure, and combined with insoluble material (which was confirmed to be product by NMR) yielding **A1.1** (0.249 g, 70.5% yield). **^{1}H NMR (400 MHz, D_2O) δ:** 7.14 (d, *J* = 4.0 Hz,

1H), 6.78 (d, J = 4.0 Hz, 1H), 5.47 (d, J = 2.8 Hz, 1H), 5.15 (d, J = 2.8 Hz, 1H), 3.810 (s, 3H). **LRMS (ESI$^-$):** m/z calculated for $(C_{12}H_9O_4)^-$ 234.04, found 234.07.



A1.4 References

1. Zhang, J., Van Lanen, S. G., Ju, J., Liu, W., Dorrestein, P. C., Li, W., Kelleher, N. L., and Shen, B. (2008) A phosphopantetheinylating polyketide synthase producing a linear polyene to initiate enediyne antitumor antibiotic biosynthesis, *Proc. Natl. Acad. Sci. U.S.A. 105*, 1460-1465.
2. Liu, W., Nonaka, K., Nie, L., Zhang, J., Christenson, S. D., Bae, J., Van Lanen, S. G., Zazopoulos, E., Farnet, C. M., Yang, C. F., and Shen, B. (2005) The neocarzinostatin biosynthetic gene cluster from Streptomyces carzinostaticus ATCC 15944 involving two iterative type I polyketide synthases, *Chem. Biol. 12*, 293-302.
3. Cserzo, M., Wallin, E., Simon, I., von Heijne, G., and Elofsson, A. (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method, *Protein Eng. 10*, 673-676.
4. Shibuya, M., Sakurai, H., Maeda, T., Nishiwaki, E., and Saito, M. (1986) Synthesis of the degradation product of auromomycin chromophore and DNA-cleaving activities of its derivatives, *Tet. Lett. 27*, 1351-1354.

Appendix 2

Structure and Chemistry of 4-Methylideneimidazole-5-one Containing Enzymes

Adapted from a review published in *Curr*. *Opin*. *Chem*. *Biol*. (**2009**), *13*, 453-461 with C.V. Christianson and S.D. Bruner

A2.1 Summary of recent advances

The prosthetic group 4-methylideneimidazole-5-one (MIO) is the catalytic component of the ammonia lyase class of enzymes. This family is responsible for the processing of amino acids in a variety of metabolic pathways through the elimination of ammonia to form unsaturated products. Recently, new chemistry has been attributed to this family with the discovery of MIO-based aminomutases. The mechanism of electrophilic chemistry catalyzed by MIO-based enzymes has been investigated for several decades. Recent X-ray crystal structures of members of the family have provided novel insight into the molecular basis for catalysis and substrate recognition. In addition, the inclusion of aminomutases in natural product biosynthetic pathways has spurned recent advances toward rational engineering and chemoenzymatic applications.

A2.2 Introduction

Electrophilic catalysis by enzymes often requires small molecule cofactors or metal ions as the functional group repertoire of proteins is limited. Flavins, pyridoxal phosphate, and chelated metals are representative of electrophilic cofactors employed in diverse, well-characterized enzymes. The ammonia lyase class of enzymes are unique members of this family, carrying out chemistry without a metal or common cofactor (Figure A2.1A) *(1, 2, 3)*. The family catalyzes the elimination of ammonia from L-amino acids to generate unsaturated products, functioning in a variety of biological pathways from amino acid catabolism to the generation of intermediates in the biosynthesis of secondary metabolites. In 2003, an enzyme with aminomutase activity, highly homologous to ammonia lyases, was described *(4)*. This expanding sub-family is

responsible for the biosynthesis of β-amino acids in structurally diverse natural products (Figure A2.1C). In this review, we will discuss recent efforts toward elucidating the mechanism of these enzymes and summarize their use in chemoenzymatic applications.

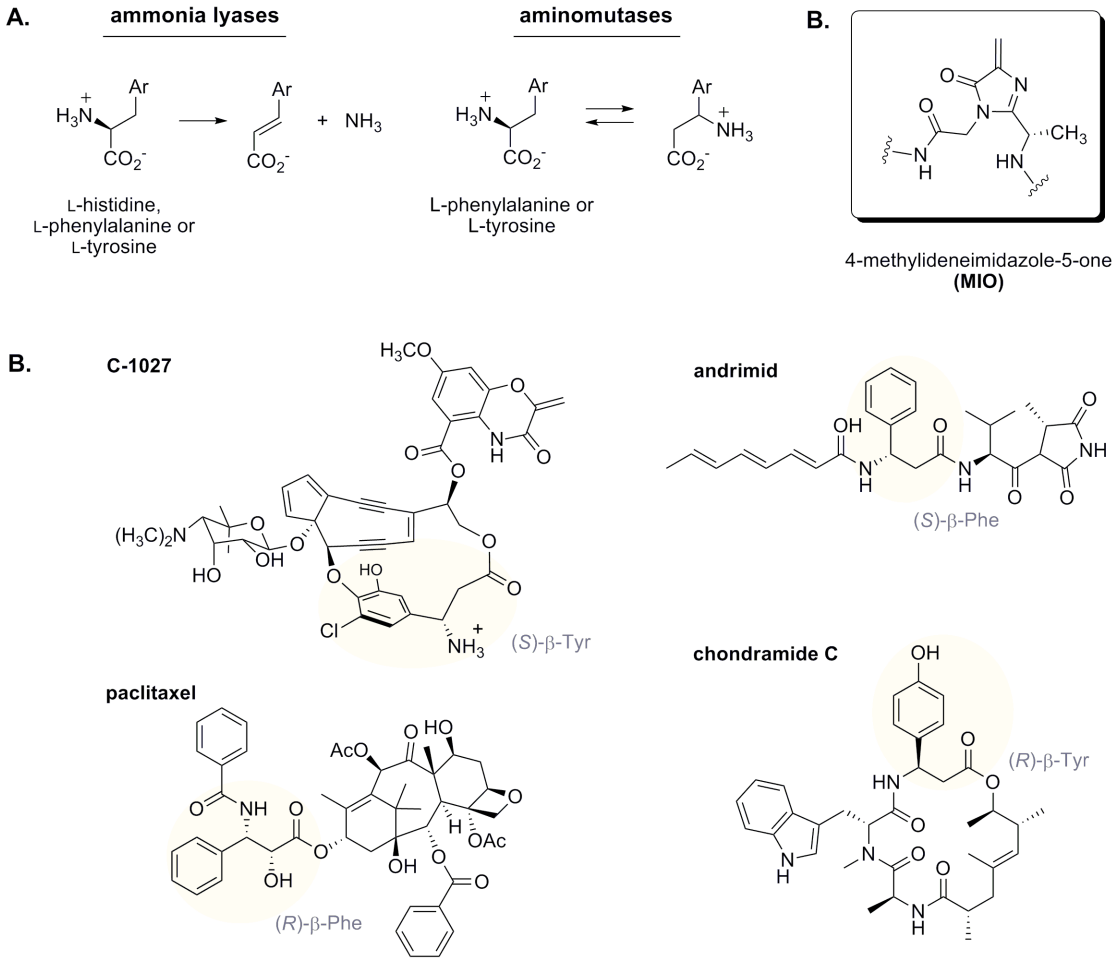


Figure A2.1: The chemistry of MIO-based enzymes. A) Summary of chemical reactions catalyzed by MIO-based ammonia lyases and aminomutases. The currently known amino acid substrates for each class are listed. B) The structure of 4-methylideneimidazole-5-one (MIO) generated by the autocatalytic condensation of an Ala-Ser-Gly motif. C) Examples of natural products containing β-amino acid moieties (highlighted) derived from MIO catalysis.

A2.3 MIO-based ammonia lyases

The chemical mechanism of ammonia lyases has been under investigation since the middle part of the last century *(5)*. It was originally postulated that electrophilic catalysis was promoted through an active site dehydroalanine moiety *(6)*. Early biochemical investigations supported a mechanism where the α-amine of the substrate added into the dehydroalanine and the elimination of bound ammonia occurred through an E1-cB mechanism (Figure A2.2A) *(7, 8)*. In 1999, the first X-ray crystal structure of an ammonia lyase revealed that the active site moiety was not dehydroalanine, but was the novel prosthetic group 4-methylideneimidazole-5-one (MIO, Figure A2.1B) *(9)*. The MIO is derived from auto-catalytic condensation of an Ala-Ser-Gly motif and is structurally similar to the well-characterized green fluorescence protein (GFP) chromophore *(10)*. The mechanistic conclusions proposed for dehydroalanine are largely transferable to MIO as both are non-aromatic electrophilic species. The discovery of this novel chemical entity produced renewed interest in the study of this enzyme class and resulted in additional investigations into an alternate mechanism that proceeds via electrophilic activation (Friedel-Crafts-type) of the substrate aryl side chain (Figure A2.2B) *(2, 11)*. Among the data used to support this mechanism are isotope effects consistent with hydrogens on the aromatic ring being directly involved in the reaction pathway *(12)*, the activity of alternate substrates with substitutions on the aromatic ring *(11, 13)* and designed small molecule model systems *(14)*.

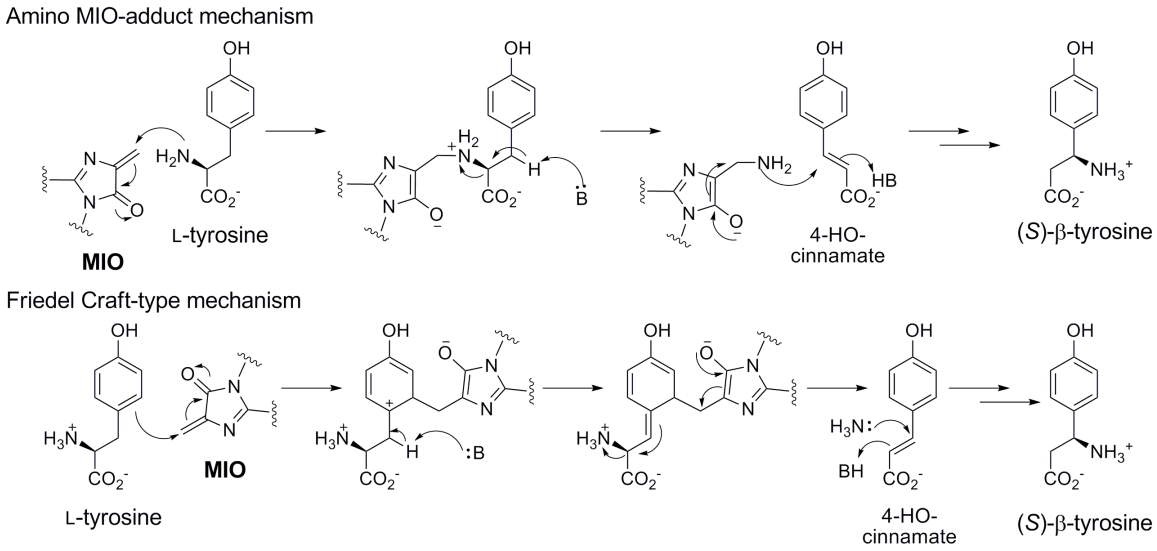


Figure A2.2: Proposed chemical mechanisms for MIO utilizing enzymes. Two mechanisms are illustrated with L-tyrosine as an example substrate. In a lyase enzyme, the cinnamate product will be released while a mutase will catalyze the conjugate addition as illustrated.

A2.4 MIO-based aminomutases

In 2003, new chemistry for the MIO-based ammonia lyase family was described with the characterization of a novel class of aminomutases *(4, 15)*. β-Amino acids are typically prepared enzymatically from the corresponding L-amino acid via a 2,3-amine shift *(16, 17)*. Members of the enediyne antitumor/antibiotic class of natural products often contain a functionalized β-tyrosine moiety (see Figure A2.1C) *(18)*. Analysis of the biosynthetic gene cluster for the enediyne C-1027, produced by *Streptomyces globisporus*, did not reveal a gene homologous to previously characterized 2,3-aminomutases. There was, however, a gene with high sequence homology to MIO-dependent ammonia lyases and, indeed, the purified enzyme, *Sg*TAM, was shown to convert L-tyrosine to (*S*)-β-tyrosine *(4)*. The enzyme exhibits biochemical characteristics consistent with an active site MIO and, after extended reaction times, the lyase product (4-hydroxycinnamate) predominates. *Sg*TAM can catalyze the reverse reaction (β-amino

acids to L-amino acids) as well as the addition of ammonia into cinnamates. These findings, along with the high sequence homology, suggests that the chemical mechanism of MIO-based aminomutases is similar to that of lyases and that the former may have evolved from the latter.

In addition to the C-1027 enediyne, the biosynthetic pathway to the anticancer therapeutic paclitaxel, isolated from the yew tree, also utilizes an MIO-based aminomutase *(19)*. The first biosynthetic step to the paclitaxel-sidechain is the preparation of (*R*)-β-phenylalanine by an L-phenylalanine specific aminomutase. Characterization of this enzyme suggested an MIO-based mechanism similar to *Sg*TAM except *Sg*TAM generated the opposite stereochemistry. Unlike *Sg*TAM, there is no scrambling of the stereochemistry of β-amino acid products over extended reaction times. Recently, additional MIO-based aminomutases have been described from biosynthetic gene clusters including the enediyne maduropeptin *(20)*, antibiotic andrimid *(21, 22)*, and the chondramides *(23)*.

The chemical mechanism for aminomutases was originally suggested to be an extension of the Friedel-Crafts-type mechanism proposed for ammonia lyases, but with an additional step involving conjugative addition of ammonia to yield the β-amino acid (see Figure A2.2B) *(4)*. However, with this mechanism the role of the MIO in the conjugate addition was not clear and the observed reversibility of the reaction was difficult to rationalize.

A2.5 Structural characterization of MIO-based ammonia lyases

The first X-ray crystal structure of an MIO-containing enzyme was histidine ammonia lyase from *Pseudomonas putida*, which revealed the chemical nature of the active site electrophile *(9)*. The overall structure is composed of multiple parallel α-helices forming a tetramer arranged as a dimer of head-to-tail dimers. The active site is located at the monomer interface with amino acid side chains from three subunits comprising the active site region around each MIO prosthetic group (Figure A2.3). Subsequent crystal structures of phenylalanine and tyrosine ammonia lyases provided a structural basis for additional substrate specificity *(24, 25)*. The combined work demonstrated that the class has significant structural homology, particularly around the active site. One common structural feature is that a large number (typically six) of the protein α-helices direct the positive end of their helix dipoles toward the active site *(24)*. The culmination of these dipoles results in a significant positive charge in the active site, which would favor a reaction pathway with negatively charged intermediates as in the E1-cB elimination. In order to probe the nature of substrate recognition, previous structures of the histidine and phenylalanine ammonia lyases were used as a guide to rationally switch the substrate specificity in a tyrosine ammonia lyase from *Rhodobacter sphaeroides* (*Rs*TAL) to favor L-phenylalanine *(25, 26)*. A single point mutation in the active site was made (His89Phe) at a position that forms a hydrogen bond with the substrate phenol. Modeling the substrate into the active site showed that an amine adduct with the MIO would place the phenol of tyrosine 2.4 Å away from His89, supporting an amine activation mechanism *(25)*. Additional support for the amino-MIO adduct

mechanism comes from recent structures of the phenylalanine ammonia lyases from *Anabaena variabilis* and *Nostoc punctiforme (27-29)*. A double mutant of *Av*PAL was used to generate crystals with ordered electron density around the active site, including loops that are typically disordered in other lyase structures, resulting in an improved view of the active site. Catalytic residues can also be identified by the proximity of inherently bound cinnamic acid, including Tyr78, which is positioned appropriately to act as the enzymatic base. Docking studies were also used to show that binding of substrate and product is consistent with a carbanion elimination mechanism.

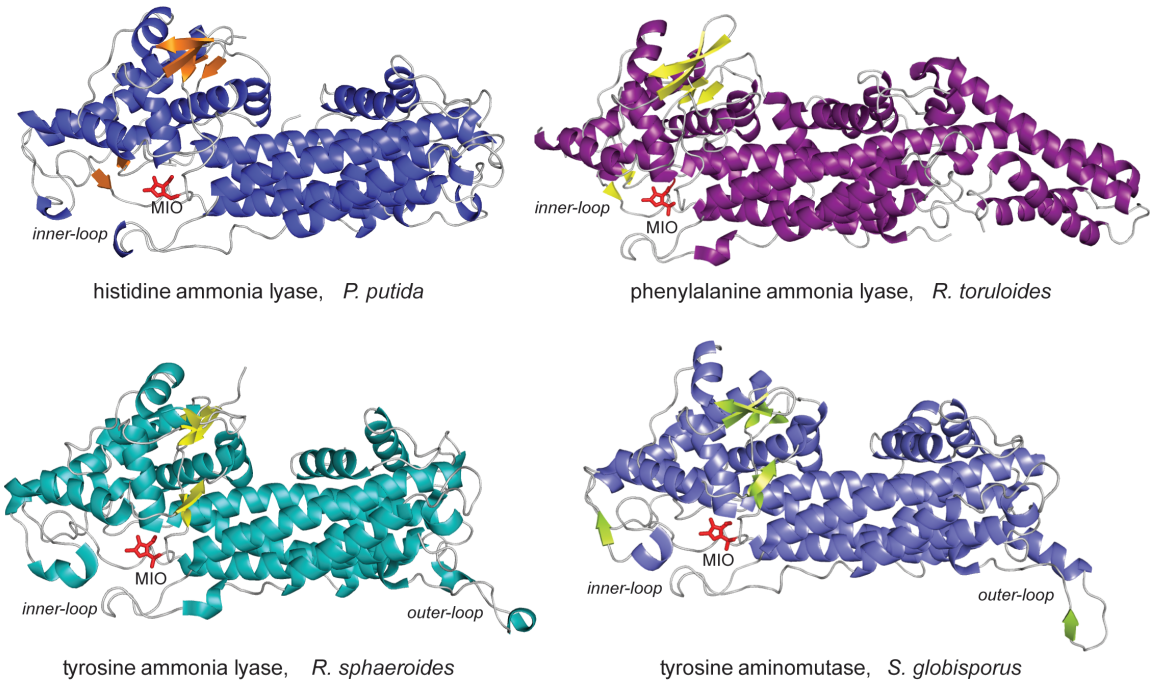


Figure A2.3: Example X-ray crystal structures of MIO-ammonia lyases and aminomutases. The protein monomers (shown) form biologically relevant tetramers where the active site is protected by inner and outer loop regions. The PDB codes are 1B8F, *P. putida*; 1T6J, *R. toruloides*; 2O6Y, *R. sphaeroides* and 2OHY, *S. globisporus*.

A2.6 Structural basis for MIO aminomutase activity

In 2007, the first structure of an MIO-based aminomutase was determined with the solution of the tyrosine aminomutase, *Sg*TAM *(30)*. The structure is highly homologous to ammonia lyases, in particular to the tyrosine ammonia lyase, *Rs*TAL *(9, 25)*. One feature shared between the two structures is that the active site loops were in an ordered, 'closed' conformation around the MIO prosthetic group. A conserved tyrosine residue (Tyr63 in *Sg*TAM) resides on the inner loop in close proximity to the MIO. Modeling of a substrate into the active site placed this residue in the proper position to act as a general enzymatic base. Indeed, mutation of this residue to a phenylalanine results in complete loss of activity, an observation also seen in *Rs*TAL *(31, 32)*.

With available structures of both an L-tyrosine specific ammonia lyase and aminomutase, the basis for the different chemistries can be explored. Comparison of the active sites of *Sg*TAM and *Rs*TAL showed similar positions and identity of side chains around the MIO *(30)*. Only one obvious difference is observed at a site adjacent to the selectivity switch (His93). At this position, *Rs*TAL has a valine residue (Val409), and *Sg*TAM has a tyrosine (Tyr415). A difference of solvent accessibility to the MIO in the active site was also noted. In the ammonia lyase, an open channel leads to the active site and no such channel was present in *Sg*TAM. This could enable mutases to sequester the substrate and ammonia in the active site, favoring the conjugate addition to form the β-amino acid.

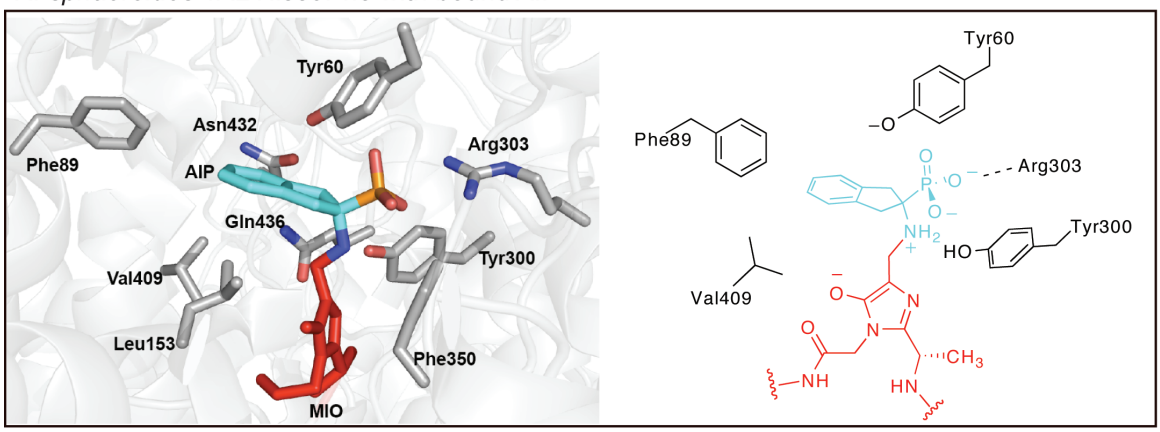
A2.7 Small molecule probes into the mechanism of MIO-enzymes

Earlier work characterized the mode of irreversible inhibition of histidine ammonia lyase by the amino acid L-cysteine in the presence of molecular oxygen. Tanner and coworkers isolated and structurally characterized the adduct formed showing that L-cysteine is bound to the MIO through the α-amine *(33)*. Heteroaryl-acrylate and alanine analogs were demostrated to be inhibitors of both histidine and phenylalanine ammonia lyases, a result that was proposed to support a Friedel-Crafts-type mechanism *(34, 35)*. An additional synthetic inhibitor is a conformationally restrained substrate analog, 2-aminoindan-2-phosphonate (AIP) *(36)*. The specificity mutant (His89Phe) of *Rs*TAL, as described above, was cocrystallized with AIP to form an approximation of the substrate complex *(25)*. In this structure, the α-amine of AIP formed a covalent adduct with the MIO (Figure A2.4). A variety of cocomplex crystal structures were obtained with the wild-type enzyme, the His89Phe mutant and products 4-hydroxycinnamate or cinnamate. The *trans*-alkene of the products is positioned directly over the MIO, maintaining the interactions between residue 89 and the aryl ring as predicted. In addition, in the absence of substrate/product analogs, the electron density maps show the MIO covalently modified by a small molecule, possibly ammonia. This work provides further evidence for the ordered release of products from the active site of lyases, as biochemically demonstrated *(37)*.

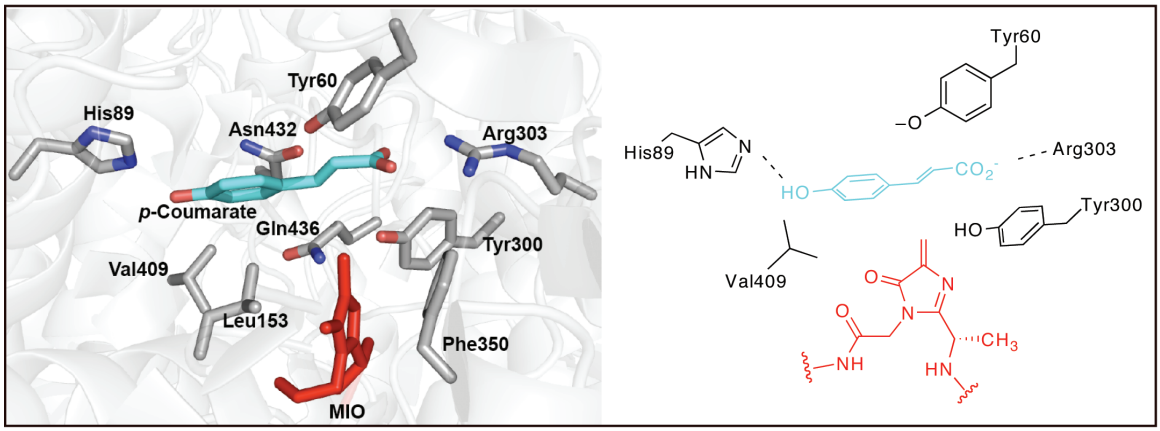
An analogous approach using cocrystallization with small molecule probes was applied to the tyrosine aminomutase, *Sg*TAM. Exploiting the reversibility of the chemistry, a stable analog of the product, α,α-difluoro-(*S*)-β-tyrosine, was used to assign

key residues (Figure A2.4) *(32)*. As with the lyase/AIP structure, the amine of the product analog is covalently attached to the MIO. The general base Tyr63 is well positioned to deprotonate at the α- and β-positions and eliminate MIO-bound ammonia through an anti-periplanar mechanism. In addition, cinnamate epoxides were designed and synthesized to mimic the coumarate intermediate in the reaction, while providing a reactive moiety that could act as a mechanistic 'trap' *(38)*. When incubated with *Sg*TAM, the epoxide was converted to the 2,3-diol and the cocrystal structures showed the analog bound to the MIO through the 2-hydroxyl group, forming a complex that mimics an enzyme-substrate complex.

R. sphaeroides TAL His89Phe with bound AIP



R. sphaeroides TAL with bound 4-hydroxycinnamate



S globisporus TAM with bound α,α-difluoro-β-tyrosine

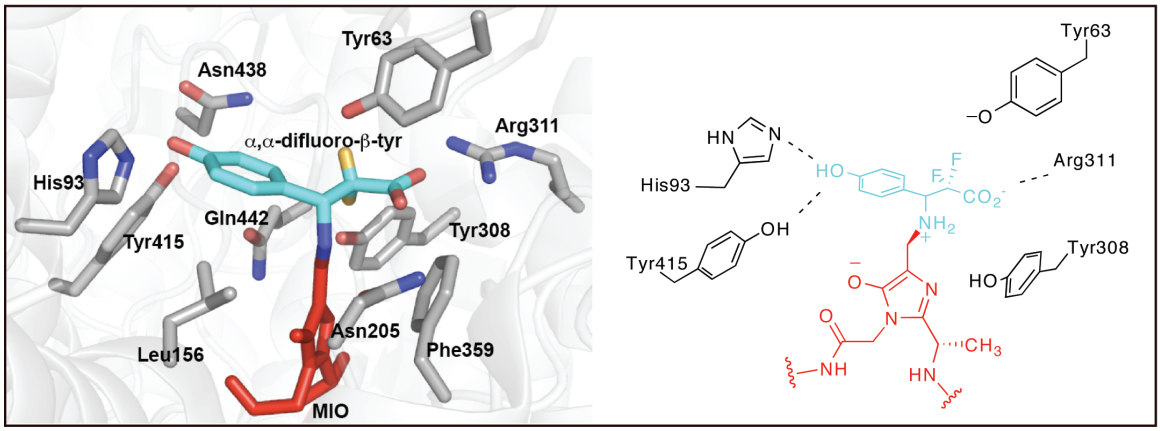


Figure A2.4: Active site structures of MIO-ammonia lyases and aminomutases with bound small molecule probes. Enzyme residues (gray) in proximity to the MIO (red) are illustrated along with the observed small molecule (teal). The PDB codes are 2O7E, *R. sphaeroides:* AIP; 2O7B, *R. sphaeroides*: HO-cinnamate and 2QVE, *S. globisporus:* α,α-diF-β-tyrosine.

A2.8 Unified mechanism for MIO-based enzymes

The recent structural characterization of multiple MIO-enzymes supports an elimination/conjugate addition mechanism (Figure A2.5). In this cycle, the α-amine of the substrate amino acid forms a covalent adduct with the exocyclic alkene of the MIO prosthetic group. The bound amino acid is then deprotonated at the benzylic position by an enzymatic base, a tyrosine side chain possibly in the phenolate ionization state. Upon loss of ammonia from the substrate, the cinnamic acid is sequestered in the active site in the case of aminomutases, and released in the case of lyases. For aminomutases, the ammonia remains bound to the MIO, and, by simple bond rotation, the amine can be repositioned for addition into the β-position. The β-amino acid is then released, regenerating the MIO. The amino-MIO adduct mechanism accounts for the observed kinetic isotope effects *(7)*, the ordered release of products *(12)*, and the characterization of intermediates and products of the reaction *(15)*.

Recent work by Walker and coworkers has provided biochemical insights into the mechanism**.** Deuterium labeling studies were used to demonstrate that the amine and the benzyllic hydrogen atom trade positions with facial selectivity in paclitaxel PAM *(39)*. They proposed a syn-configuration for the unsaturated intermediate, which would be the only way for the amine and *proS* hydrogen to maintain facial selectivity without the substrate rotating during the reaction sequence. This general mechanistic proposal involving retention of the α- and β-nitrogen was also supported by studies on the tyrosine aminomutase from andrimid biosynthesis *(40)*.

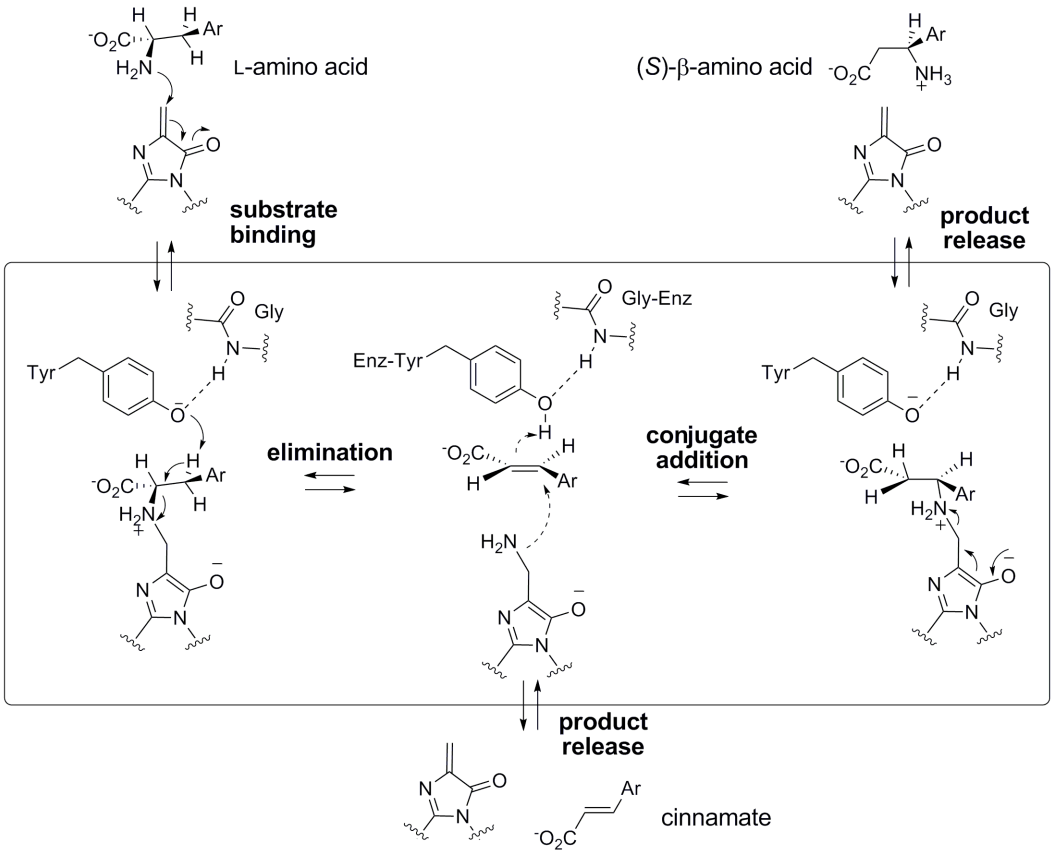


Figure A2.5: Unified mechanistic proposal for MIO-based ammonia lyase and aminomutase enzymes. The substrate 'Ar' corresponds to imidazole, phenyl, or 4-hydroxyphenyl.

An amine-adduct mechanism does not, by itself, fully resolve the difficultly of deprotonating a benzylic hydrogen. However, the extensively electropositive active site formed by α-helical dipoles can stabilize the benzylic anion generated through an E1-cB mechanism, effectively lowering the pKa. In addition, cocomplex structures indicate that the enzyme binds substrate in a conformation favoring orbital alignment for a periplanar elimination. A recent publication by Bartsch and Bornscheuer suggests that perhaps different mechanisms may occur depending on the nature of the substrate *(41)*. For PAL, the substrate would undergo a Friedel-Crafts-like mechanism, while for TAL the

substrate would go through the amino-MIO adduct mechanism. Their reasoning is based on a conserved glutamate in the active site for PAL and PAM enzymes, which acts by blocking binding of the amino group to the MIO; an asparagine is conserved at this position for TAL and TAM enzymes, which would allow the MIO to bind to the amino group on the substrate. These assumptions are supported via mutations of a single residue (Glu484Asn) and the resulting activity towards tyrosine and phenylalanine. It is uncommon for homologous members of the same enzyme class to utilize drastically different chemical mechanisms based on the nature of the substrate.

A2.9 Utilization of MIO-aminomutases in chemoenzymatic synthesis applications

The chemistry of MIO-dependent aminomutases has led to attempts to exploit substrate promiscuity for the production of novel β-amino acids. Walker and coworkers have shown that paclitaxel PAM catalyzes the isomerization of diverse arylalanine substrates to the corresponding (*R*)-β-arylalanines *(42)*. Both fluorinated phenylalanines and phenylalanines substituted with electron-donating groups were turned over with increased efficiency over the natural substrate. Interestingly, (*S*)-styrylalanine was converted to (*R*)-β-styrylalanine, although at a significantly lower reaction rate. Janssen and coworkers illustrated another chemoenzymatic application by demonstrating that PAM can aminate functionalized (*E*)-cinnamic acids to produce both α- and β-amino acids with high enantiomeric excess *(43)*. It was found that electronic effects influenced the regioselectivity of the reaction with electron-donating groups on the aryl ring favoring β-amino acids and, conversely, electron-withdrawing groups preferentially generating α-amino acids. For example, in the case of the (*E*)-methyl- and (*E*)-nitrocinnamates,

amination occurred almost exclusively at the β- and α-positions, respectively. As opposed to the high substrate tolerance shown with PAM, the tyrosine aminomutase *Sg*TAM has significantly tighter selectivity *(15)*.

A2.10 Rational engineering of lyase and aminomutase activity

While some enzymes from this class show a considerable amount of flexibility for substrates, the products are generally fixed in terms of lyase and mutase paths. Lyases produce elimination products (cinnamates) while mutases primarily produce one β-amino acid enantiomer. With the high degree of homology shared between lyases and mutases, determining the molecular basis for the two distinct activities poses a significant challenge. Theoretically, an ammonia lyase could be mutated into an aminomutase and vice versa, the latter of which would seem facile since mutases already exhibit lyase activity. Also, one could envision changing the stereoselectivity of mutases to produce different β-amino acids. Towards these ends, Krug and Müller recently reported the effects of a variety of mutations of CmdF, which is responsible for the production of (*R*)-β-tyrosine for incorporation into the chondramide family of natural products *(44)*. Attempts to engineer CmdF into an ammonia lyase by exchanging conserved sequence motifs from lyases resulted in no activity or substantially reduced activity with retained mutase activity. One successful mutant was an exchange of a sequence in CmdF for a homologous region in MdpC4 (the tyrosine aminomutase found in maduropeptin biosynthesis *(20)*), which surprisingly did exhibit a shift from aminomutase activity to ammonia lyase activity. Additionally, a mutation (Glu399Lys) led to an increase in enantioselectivity as compared to the wild-type enzyme. On the basis of these and other

studies, it appears that this family of enzymes is quite sensitive to changes in and around the active site and that engineering novel activity will not be straightforward. The MIO cofactor is a chemically reactive species and even conservative mutations to the active site structure results in the abolishment of activity. This could stem from exposure of the active site to external nucleophiles or possible disruption of the quaternary structure by interfering with the tetramer interface.

A2.11 Conclusions

Recent structural biology efforts have provided insight into the chemistry of enzymes dependent on the MIO prosthetic group. The use of small molecule probes, along with biochemical studies, is consistent with a mechanism where the electrophilic MIO reacts with the amino acid and the enzyme orients the bound complex for an E1-cB elimination. The recently described aminomutase subfamily diverts the lyase intermediate toward β-amino acid products, resulting in an overall 2,3-amino shift. These investigations have provided a structural basis for substrate specificity and defined key residues involved in catalysis. Further mechanistic work is necessary to resolve ambiguities in the chemical mechanism, particularly in defining the precise role the MIO prosthetic group plays in facilitating the catalytic cycle. Members of this enzyme family are important catalysts in the production of secondary metabolites and are rapidly expanding in both number and scope. β-amino acids are biosynthetic building blocks in a variety of natural products. The unique chemistry of MIO-based enzymes has the potential to be used in the preparation of chiral building blocks and in biosynthetic engineering approaches toward novel therapeutics.

A2.12 References

1. Poppe, L. (2001) Methylidene-imidazolone: a novel electrophile for substrate activation, *Curr. Opin. Chem. Biol. 5*, 512-524.
2. Poppe, L., and Retey, J. (2005) Friedel-Crafts-type mechanism for the enzymatic elimination of ammonia from histidine and phenylalanine, *Angew. Chem. Int. Ed. Engl. 44*, 3668-3688.
3. MacDonald, M. J., and D'Cunha, G. B. (2007) A modern view of phenylalanine ammonia lyase, *Biochem. Cell. Biol. 85*, 273-282.
4. Christenson, S. D., Liu, W., Toney, M. D., and Shen, B. (2003) A novel 4-methylideneimidazole-5-one-containing tyrosine aminomutase in enediyne antitumor antibiotic C-1027 biosynthesis, *J. Am. Chem. Soc. 125*, 6062-6063.
5. Tabor, H. M., A. H.; Hayaishi, O.; White, J. (1952) Urocanic acid as an intermediate in the enzymatic conversion of histidine to glutamic and formic acids., *J. Biol. Chem. 196*, 121-128.
6. Wickner, R. B. (1969) Dehydroalanine in histidine ammonia lyase, *J. Biol. Chem. 244*, 6550-6552.
7. Hermes, J. D., Weiss, P. M., and Cleland, W. W. (1985) Use of nitrogen-15 and deuterium isotope effects to determine the chemical mechanism of phenylalanine ammonia-lyase, *Biochemistry 24*, 2959-2967.
8. Peterkofsky, A. (1962) The mechanism of action of histidase: amino-enzyme formation and partial reactions, *J. Biol. Chem. 237*, 787-795.
9. Schwede, T. F., Retey, J., and Schulz, G. E. (1999) Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile, *Biochemistry 38*, 5355-5361.
10. Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) Crystal structure of the Aequorea victoria green fluorescent protein, *Science 273*, 1392-1395.
11. Langer, M., Pauling, A., and Retey, J. (1995) The role of dehydroalanine in catalysis by Histidine Ammonia Lyase, *Angew. Chem. Int. Ed. Engl. 34*, 1464-1465.
12. Furuta, T., Takahashi, H., Shibasaki, H., and Kasuya, Y. (1992) Reversible stepwise mechanism involving a carbanion intermediate in the elimination of ammonia from L-histidine catalyzed by histidine ammonia-lyase, *J. Biol. Chem. 267*, 12600-12605.
13. Schuster, B., and Retey, J. (1995) The mechanism of action of phenylalanine ammonia-lyase: the role of prosthetic dehydroalanine, *Proc. Natl. Acad. Sci. U.S A. 92*, 8433-8437.
14. Rettig, M., Sigrist, A., and Retey, J. (2000) Mimicking the reaction of phenylalanine ammonia lyase by a synthetic model, *Helvetica Chimica Acta 83*, 2246-2265.
15. Christenson, S. D., Wu, W., Spies, M. A., Shen, B., and Toney, M. D. (2003) Kinetic analysis of the 4-methylideneimidazole-5-one-containing tyrosine

aminomutase in enediyne antitumor antibiotic C-1027 biosynthesis, *Biochemistry 42*, 12708-12718.
16. Frey, P. A., and Reed, G. H. (2000) Radical mechanisms in adenosylmethionine- and adenosylcobalamin-dependent enzymatic reactions, *Arch. Biochem. Biophys. 382*, 6-14.
17. Wetmore, S. D., Smith, D. M., and Radom, L. (2001) Enzyme catalysis of 1,2-amino shifts: the cooperative action of B6, B12, and aminomutases, *J. Am. Chem. Soc. 123*, 8678-8689.
18. Liu, W., Christenson, S. D., Standage, S., and Shen, B. (2002) Biosynthesis of the enediyne antitumor antibiotic C-1027, *Science 297*, 1170-1173.
19. Walker, K. D., Klettke, K., Akiyama, T., and Croteau, R. (2004) Cloning, heterologous expression, and characterization of a phenylalanine aminomutase involved in Taxol biosynthesis, *J. Biol. Chem. 279*, 53947-53954.
20. Van Lanen, S. G., Oh, T. J., Liu, W., Wendt-Pienkowski, E., and Shen, B. (2007) Characterization of the maduropeptin biosynthetic gene cluster from Actinomadura madurae ATCC 39144 supporting a unifying paradigm for enediyne biosynthesis, *J. Am. Chem. Soc. 129*, 13082-13094.
21. Jin, M., Fischbach, M. A., and Clardy, J. (2006) A biosynthetic gene cluster for the acetyl-CoA carboxylase inhibitor andrimid, *J. Am. Chem. Soc. 128*, 10660-10661.
22. Magarvey, N. A., Fortin, P. D., Thomas, P. M., Kelleher, N. L., and Walsh, C. T. (2008) Gatekeeping versus promiscuity in the early stages of the andrimid biosynthetic assembly line, *ACS Chem. Biol. 3*, 542-554.
23. Rachid, S., Krug, D., Weissman, K. J., and Muller, R. (2007) Biosynthesis of (*R*)-beta-tyrosine and its incorporation into the highly cytotoxic chondramides produced by Chondromyces crocatus, *J. Biol. Chem. 282*, 21810-21817.
24. Calabrese, J. C., Jordan, D. B., Boodhoo, A., Sariaslani, S., and Vannelli, T. (2004) Crystal structure of phenylalanine ammonia lyase: multiple helix dipoles implicated in catalysis, *Biochemistry 43*, 11403-11416.
25. Louie, G. V., Bowman, M. E., Moffitt, M. C., Baiga, T. J., Moore, B. S., and Noel, J. P. (2006) Structural determinants and modulation of substrate specificity in phenylalanine-tyrosine ammonia-lyases, *Chem. Biol. 13*, 1327-1338.
26. Watts, K. T., Mijts, B. N., Lee, P. C., Manning, A. J., and Schmidt-Dannert, C. (2006) Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family, *Chem. Biol. 13*, 1317-1326.
27. Wang, L., Gamez, A., Archer, H., Abola, E. E., Sarkissian, C. N., Fitzpatrick, P., Wendt, D., Zhang, Y., Vellard, M., Bliesath, J., Bell, S. M., Lemontt, J. F., Scriver, C. R., and Stevens, R. C. (2008) Structural and biochemical characterization of the therapeutic Anabaena variabilis phenylalanine ammonia lyase, *J. Mol. Biol. 380*, 623-635.
28. Xiang, L., and Moore, B. S. (2005) Biochemical characterization of a prokaryotic phenylalanine ammonia lyase, *J. Bacteriol. 187*, 4286-4289.

29. Moffitt, M. C., Louie, G. V., Bowman, M. E., Pence, J., Noel, J. P., and Moore, B. S. (2007) Discovery of two cyanobacterial phenylalanine ammonia lyases: kinetic and structural characterization, *Biochemistry 46*, 1004-1012.
30. Christianson, C. V., Montavon, T. J., Van Lanen, S. G., Shen, B., and Bruner, S. D. (2007) The structure of L-tyrosine 2,3-aminomutase from the C-1027 enediyne antitumor antibiotic biosynthetic pathway, *Biochemistry 46*, 7205-7214.
31. Schroeder, A. C., Kumaran, S., Hicks, L. M., Cahoon, R. E., Halls, C., Yu, O., and Jez, J. M. (2008) Contributions of conserved serine and tyrosine residues to catalysis, ligand binding, and cofactor processing in the active site of tyrosine ammonia lyase, *Phytochemistry 69*, 1496-1506.
32. Christianson, C. V., Montavon, T. J., Festin, G. M., Cooke, H. A., Shen, B., and Bruner, S. D. (2007) The mechanism of MIO-based aminomutases in beta-amino acid biosynthesis, *J. Am. Chem. Soc. 129*, 15744-15745.
33. Galpin, J. D., Ellis, B. E., and Tanner, M. E. (1999) The inactivation of histidine ammonia-lyase by L-cysteine and oxygen: Modification of the electrophilic center, *J. Am. Chem. Soc. 121*, 10840-10841.
34. Katona, A., Tosa, M. I., Paizs, C., and Retey, J. (2006) Inhibition of histidine ammonia lyase by heteroaryl-alanines and acrylates, *Chem. Biodivers. 3*, 502-508.
35. Paizs, C., Katona, A., and Retey, J. (2006) The interaction of heteroaryl-acrylates and alanines with phenylalanine ammonia-lyase from parsley, *Chemistry 12*, 2739-2744.
36. Zon, J., Miziak, P., Amrhein, N., and Gancarz, R. (2005) Inhibitors of phenylalanine ammonia-lyase (PAL): synthesis and biological evaluation of 5-substituted 2-aminoindane-2-phosphonic acids, *Chem. Biodivers. 2*, 1187-1194.
37. Givot, I. L., Smith, T. A., and Abeles, R. H. (1969) Studies on the mechanism of action and the structure of the electrophilic center of histidine ammonia lyase, *J. Biol. Chem. 244*, 6341-6353.
38. Montavon, T. J., Christianson, C. V., Festin, G. M., Shen, B., and Bruner, S. D. (2008) Design and characterization of mechanism-based inhibitors for the tyrosine aminomutase SgTAM, *Bioorg. Med. Chem. Lett. 18*, 3099-3102.
39. Mutatu, W., Klettke, K. L., Foster, C., and Walker, K. D. (2007) Unusual mechanism for an aminomutase rearrangement: retention of configuration at the migration termini, *Biochemistry 46*, 9785-9794.
40. Fortin, P. D., Walsh, C. T., and Magarvey, N. A. (2007) A transglutaminase homologue as a condensation catalyst in antibiotic assembly lines, *Nature 448*, 824-827.
41. Bartsch, S., and Bornscheuer, U. T. (2009) A single residue influences the reaction mechanism of ammonia lyases and mutases, *Angew. Chem. Int. Ed. Engl. 48*, 3362-3365.
42. Klettke, K. L., Sanyal, S., Mutatu, W., and Walker, K. D. (2007) Beta-styryl- and beta-aryl-beta-alanine products of phenylalanine aminomutase catalysis, *J. Am. Chem. Soc. 129*, 6988-6989.
43. Wu, B., Szymanski, W., Wietzes, P., de Wildeman, S., Poelarends, G. J., Feringa, B. L., and Janssen, D. B. (2009) Enzymatic synthesis of enantiopure alpha- and

beta-amino acids by phenylalanine aminomutase-catalysed amination of cinnamic acid derivatives, *Chembiochem 10*, 338-344.

44. Krug, D., and Muller, R. (2009) Discovery of additional members of the tyrosine aminomutase enzyme family and the mutational analysis of CmdF, *Chembiochem 10*, 741-750.