Structural investigation of the substrate specificity of offloading and sugar methylation in macrolide biosynthesis

by

Steffen Bernard

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2014

Doctoral Committee:

Professor Janet L. Smith, Chair
Associate Professor Patrick O'Brien
Professor David H. Sherman
Professor John J. G. Tesmer

Acknowledgements

The work presented in this dissertation would not have been possible without the guidance, advice and support of many. I am grateful for their contributions in helping me complete this work. I would like to thank my mentor, Professor Janet Smith, for supporting my ongoing development as a scientist. Janet allowed me to pursue the research directions I was most interested in, and provided me with guidance along the way. She never failed to recognize an opportunity for teaching. I am also grateful to the many current and former members of the Smith lab who provided support along the way. Professors David Sherman, John Tesmer and Patrick O'Brien served on my dissertation committee and provided feedback on experiments and also helped me place my work within a broader context.

Much of the work described here is a result of collaboration. Many members of David Sherman's group at the University of Michigan have been very helpful. In particular, I would like to thank Sung Ryeol Park for providing essential compounds and Ashootosh Tripathi for assistance with NMR experiments. Professors Fumio Kato and Yojiro Anzai of Toho University also supported this work by providing the mycinamicin compounds. Diffraction experiments were performed at GM/CA at the Advanced Photon Source and would not have been possible without the support of the beamline staff.

My family and friends have been instrumental in my success and have made my time at University of Michigan an outstanding experience. I would like to especially thank my partner, Katie Bernard, for supporting me in pursuing my doctorate and sharing in both the challenges and the successes along the way.

Table of Contents

List of Figures

List of Tables

Abstract

The increase in antibiotic resistance over the past century emphasizes the need for new antibiotic therapies. Macrolide antibiotics are commonly used to treat infections in humans and animals. Many macrolides are produced by type I polyketide synthases in actinobacteria. This thesis investigates the structural basis of substrate specificity of macrolactonization and sugar-*O*-methylation, two key steps in macrolide biosynthesis. These results are presented with the aim of supporting future efforts to design and produce new compounds through biocatalytic and bioengineering routes.

Methylation of sugar hydroxyl groups is common in biosynthetic pathways for macrolides and other natural products. Two distinct methyltransferases act sequentially in the late stages of mycinamicin biosynthesis to methylate the 2' and 3' hydroxyls of a 6-deoxyallose sugar. Structural and biochemical investigation of the 3'-*O*-methyltransferase, MycF, from the mycinamicin biosynthetic pathway provided insight into the mechanism and substrate selectivity of this family of methyltransferases. The series of structures presented herein illuminate the mechanisms that underlie sequential methylation by two families of sugar-*O*-methyltransferases common in natural product biosynthetic pathways. This foundation led to successful enzyme engineering which circumvented the natural order of the pathway to produce a new macrolide.

Macrolactonization is another common feature of natural product biosynthesis. The structure of the thioesterase (TE) domain from the tylosin biosynthetic pathway is the first structure of a 16-membered macrolactone forming TE. Similar to the pikromycin TE, the Tyl TE has a hydrophillic barrier between the active site and the exterior of the protein, which is proposed to aid in promoting cyclization. A model of the product complex provides insights into the substrate specificity of the Tyl TE. Biochemical experiments demonstrated the utility of fluorophosphonate affinity labels in future structural characterization of TE domains.

Chapter 1 : Introduction

Natural Product Antibiotics

The majority of clinically useful small molecules are derived from or inspired by natural compounds (1). Plants, fungi, bacteria and even animals produce these natural products as secondary metabolites. Clinically, natural products have been directly used, or provided the scaffold for virtually all types of pharmaceuticals including antibiotics, antivirals, immunosuppressants, and cancer chemotherapy agents. Natural product antibiotics provide a telling example of the impact of natural products on medical care. The discovery of the antibiotic penicillin, from a fungal source, significantly increased average lifespan and decreased the risk of infection associated with all medical procedures (2). Almost concomitant with the introduction of penicillin was the observation of antibiotic resistance (3) and by the end of the 1940s the majority of *Staphloccus aureaus* isolates in Europe were penicillin resistant (4). While several new classes of synthetic and natural product antibiotic agents have been discovered since their initial introduction, their ubiquitous administration has resulted in broadly resistant bacterial strains that challenge current treatment regimes (5, 6). In the 1960s the first methicilin resistant *S. aureus* (MRSA) strains were isolated. These strains were found almost exclusively in hospital settings and were resistant to almost all available antibiotic therapies. MRSA infections commonly occurred in elderly individuals or chronically ill patients. In the 1990s MRSA infections that originated outside of hospital settings were identified (7, 8). These community acquired (CA) MRSA infections frequently occur in children or otherwise healthy individuals (9). The spread of MRSA strains to

environments that lack the high selective pressure of the hospital settings has heightened the need for new therapies for the treatment of bacterial infection. Antibiotic resistance is not restricted to *S. aureaus* and many significant human pathogens have shown a similar trend of increasing antibiotic resistance including *Clostridium difficile (10), Neisseria gonorrhoeae (11), Klebsiella pneumonia (12), Haemophilus influenzae (13), Heliobacter pylori (14),* and *Mycobacterium tuberculosis (15).* This escalating antibiotic resistance crisis underscores the need for more judicious use of current antibiotics and for new antibiotic therapies.

Macrolide Antibiotics

Macrolides are a diverse group of natural products that feature a macrolactone ranging in size from 8-to 62-members (16). The term macrolide was originally used as shorthand for macrolactone glycosides, a group of macrolactone containing natural products that are glycosylated (17). The majority of macrolides are produced by actinomycetes, however other bacteria, fungi and plants also produce them (16). They have been used clinically as antibiotics, antifungals and immunosuppressants and also used agriculturally as growth promoters in livestock or as insecticides.

Macrolide antibiotics are commonly used to treat infections in humans and animals. They are composed of a 12-16 membered macrolactone ring with one or more covalently linked hexose sugars (17). The binding sites for several macrolide antibiotics including erthromycin, azithromycin and tylosin have been observed in co-crystal structures with ribosomes from several bacterial sources (18, 19). Macrolides bind to the 50S subunit of the ribosome in the peptide exit tunnel near the peptidyl transfer center. The primary interactions are stacking of the bases A2058, A2059, and U2611 (*E. coli* numbering) of the

23S rRNA against the hydrophobic macrolactone. In eukaryotic ribosomes there is a guanine at position 2058 (20). Water molecules involved in solvating the amino-group of the guanine base decrease the affinity of macrolides for eukaryotic ribosomes resulting in selectivity for bacterial ribosomes. In bacteria, the substitution A2058G, methylation of A2058, or macrolide specific efflux pumps confer macrolide resistance (2). The structures of macrolides (and other antibiotics) bound to bacterial ribosomes were expected to provide an avenue for structure based drug design of new antibiotics to overcome current resistance mechanisms. Unfortunately, a number of structures have been solved which suggest that antibiotics bind in subtly different ways to ribosomes of different species, underscoring the need for empirical evaluation of new antibiotic derivatives (21).

Tylosin and mycinamicin are 16-membered macrolide antibiotics of similar overall structure with slightly different macrolactone cores (Figure 1-1). Tylosin has a disaccharide moiety at the C5 carbon and a monosaccharide at the C14 position (22). Mycinamicin lacks a sugar at the C5 carbon but still has the mycaminose at C5 and a mycinose sugar at C14 (23). Tylosin is produced by the bacterium *Streptomyces fradiae (22)*. It is used as a veterinary antibiotic for the treatment of respiratory, skin and foot infections in cattle and pigs (24). Mycinamicin, which is produced by the bacterium *Micromonospora griseorubida*, is not currently used clinically but shows better activity against antibiotic resistant *S. aureus* strains than the clinically used antibiotics erythromycin and leucomycin (23). Skin irritation and hair loss were observed in early animal studies, which presumably prevented further clinical development (25, 26).

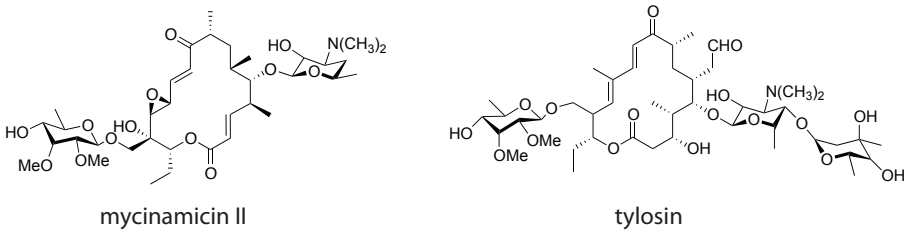


Figure 1-1 Chemical structures of the macrolide antibiotics mycinamicin II and tylosin.

Type 1 Polketide Synthases

The macrolactone core of a macrolide antibiotic is synthesized by type 1 polyketide synthase (PKS) pathways. Type 1 PKSs are homologous to type 1 fatty acid synthases (FAS) and use acyl building blocks, derived from acyl-Coenzyme A (CoA) found in primary metabolism, to produce complex products (27). Type 1 PKS pathways consist of several modules which each carry out sequential extension and sometimes modification of the pathway intermediate (Figure 1-2). The catalytic domains of each module occur on a single polypeptide. Each module minimally consists of an acyltransferase (AT) domain to select an acyl monomer, a ketosynthase (KS) domain to catalyze the decarboxylative condensation of the upstream intermediate and the newly selected building block, and an acyl carrier protein (ACP) domain that shuttles the substrate between catalytic domains (28). The pathway intermediates are covalently tethered to the ACP through a phosphopantetheine prosthetic group derived from CoA. Many modules also contain ketoreductase, dehydratase and enoyl reductase domains to further modify the acyl chain (29). At the end of the PKS assembly line a thioesterase domain catalyzes the release of the product from the ACP domain (30).

A type 1 PKS pathway carries out the biosynthesis of tylactone, the unglycosylated precursor to tylosin, with seven modules expressed on five polypeptides (Figure 1-2). The

PKS modules use one propionyl, four methylmalonyl, two malonyl and one ethylmalonyl CoAs to produce a final 17 carbon linear intermediate, which is cyclized to form a 16-membered macrolactone (22). Through the action of ketoreductase, dehydratease and enoylreductase domains present in the PKS pathway, tylactone contains 9 stereocenters, two hydroxyl groups and a conjugated dienone (Figure 1-2). The large number of stereocenters and similarly reactive functional groups make PKS products, like tylactone, challenging (but nonetheless desirable) synthetic targets. These challenges to traditional chemical synthesis and the modular nature of PKS assembly lines have inspired bioenginnering of both optimized and novel PKS pathways and additionally hybrid chemo-enzymatic routes to produce natural products and new compounds (31-33).

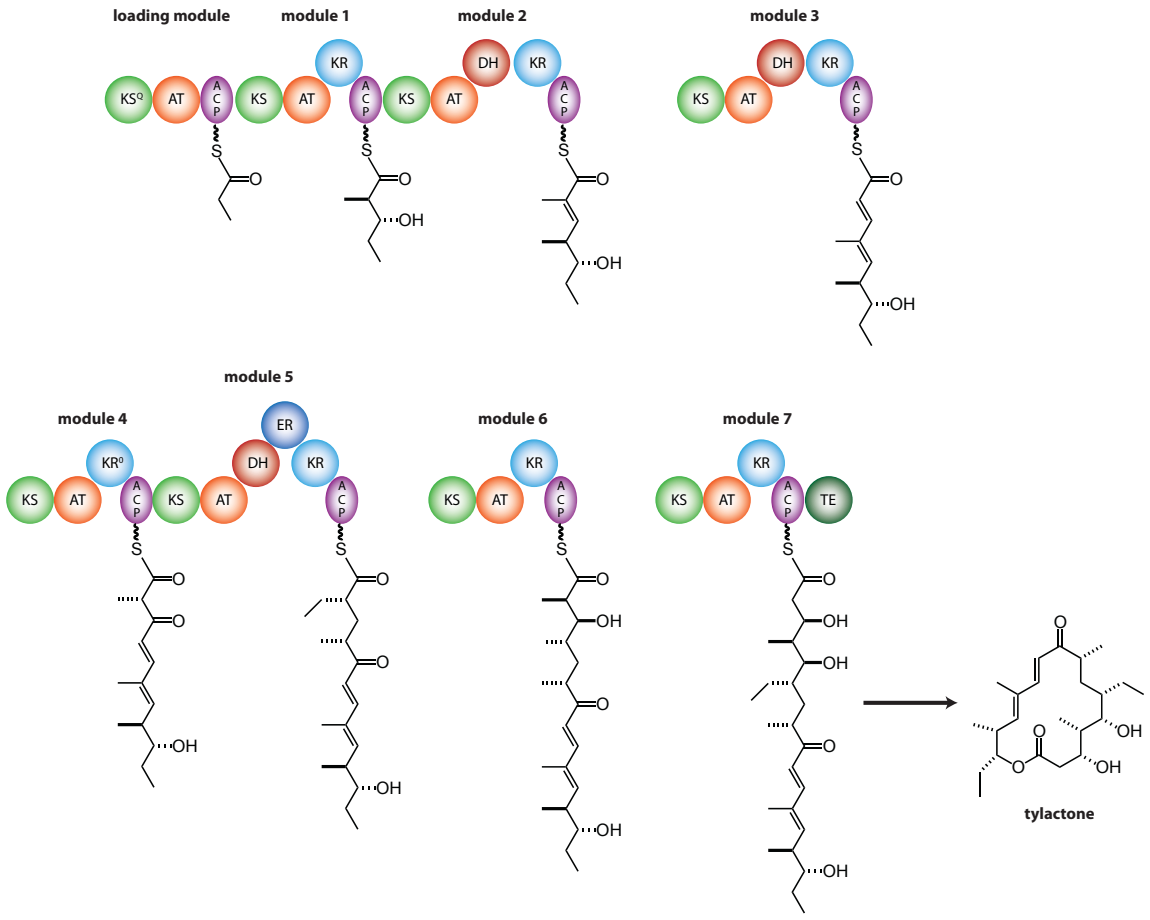


Figure -1-2 Tylosin PKS pathway (22). The five polypeptides contain seven modules that produce the 16-memberd macrolactone precursor to tylosin. Individual catalytic domains are represented as circles. Abbreviatiosn are KS^{Q} = malonyl-ACP decarboxylase, KS = ketosynthase, AT = acyltransferase, ACP = acyl carrier protein, KR = ketoreductase, DH = dehydratase, ER = enoylreductase, and TE = thioesterase. KR^{0} is an inactive KR domain in module 4.

PKS Thioesterase Domains

PKS thioesterase (TE) domains catalyze the release of the thioester-tethered intermediate from the ACP domain of the final module (30). TE domains in different

pathways can catalyze either hydrolysis to produce a linear carboxylic acid product (34) or intramolecular cyclization via a nucleophilic substrate hydroxyl to produce a macrolactone (29) (Figure 1-3). TE domains use a catalytic triad of a nucleophilic serine, histidine and aspartate like other serine hydrolases. The serine attacks the thioester bond of the ACP tethered substrate to form a TE-substrate covalent complex, releasing the ACP for the next round of substrate transfer (35). The covalent complex is then hydrolyzed by an active site water to produce a linear acid, or the substrate is positioned in the active site such that a substrate hydroxyl is in proximity to the serine ester to act as a nucleophile and cyclize the substrate (36).

Like homologous TE domains of fatty acid synthases, PKS TE domains have an alpha-beta hydrolase fold, however the active site lids of FAS and PKS TE domains differ. FAS TE domains are monomeric and have a hydrophobic binding cleft where the substrate binds (37). In PKS TEs the active site lids form a dimerization interface that keeps the lid permanently closed. The active site is located at the center of a tunnel through the TE domain. It has been hypothesized that substrates are delivered by the ACP to the N-terminal side of the tunnel and products exit via the C-terminal tunnel opening (36, 38). TE domains in macrolide antibiotic biosynthesis pathways catalyze the formation of 12, 14 and 16-membered rings. The TE domain from the pikromycin biosynthetic pathway catalyzes formation of both 12 and 14-membered rings *in vivo* and *in vitro*, suggesting that this domain is catalytically flexible (39-41). To date structural studies of PKS TE domains have focused only on domains that catalyze the formation of 12 or 14-membered rings and thus the full breadth of TE domains remain unexplored (36, 38, 42). Many biochemical studies of PKS TE domains have been limited to small molecule substrates that are minimally functionalized

making it difficult to draw strong conclusions about the substrate selectivity of these domains (43, 44). The few experiments employing natural substrates have provided significant insights into the factors controlling cyclization in PKS TE domains. The Pik TE was shown to exclusively cyclize the natural substrate but hydrolyze an analog with a hydroxyl in place of a ketone (45). Subsequent structural studies with covalent affinity labels lead to the hypothesis that the rigidity imposed by the substrate enone functionality assists in cyclization (36). Given their role in product offloading, it is essential to understand the substrate specificity of TE domains to successfully engineer new biosynthetic pathways or utilize them for macrolactone formation in chemo-enzymatic synthesis.

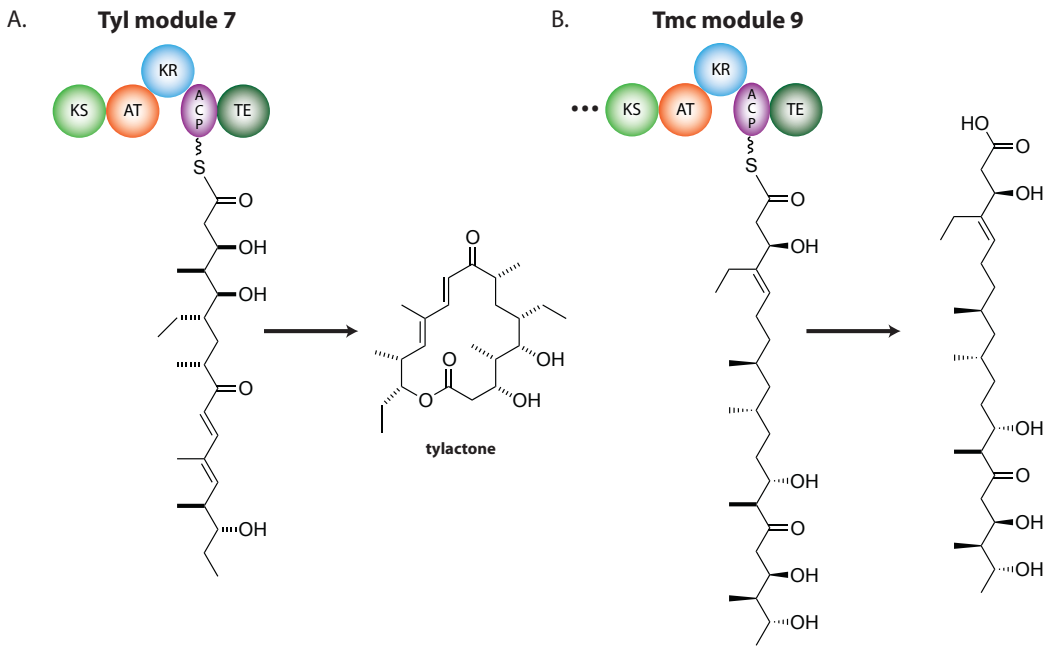


Figure 1-3 Reactions catalyzed by PKS TE domains. PKS domains catalyze cyclization (A) to produce macrolactones including tylactone or hydrolysis to produce carboxylic acids as is the case in the tautomycetin (Tmc) PKS.

Post PKS Tailoring Steps

Both linear and cyclized PKS products are frequently modified through additional enzyme-catalyzed reactions after their release from the pathway (46). Many macrolide antibiotics are modified by oxidoreductases and glycosyltransferases. These modifications significantly increase the diversity of functional groups included in the final product (Figure 1-4). Oxidoreductases add alchohol, ketone, aldehyde and epoxide functionalities and glycoslytransferases catalyze the transfer of heavily modified sugars to hydroxyl groups on the macrolactone ring (46). These tailoring steps frequently occur in a specific order necessitating a thorough understanding of the determinants of substrate specificity for PKS engineering (47, 48).

Sugars used in natural product biosynthesis originate from primary metabolism (49). They are activated through coupling with a nucleotide base and modified by amination, isomerization, methylation, reduction, dehydration and carbamoylation (50). In many bacterial systems the sugar nucleotide is in the form of a thymidine diphosphate (TDP) conjugate originally derived from glucose-1-phosphate (Figure 1-5). Many modifications including *N*- and *C*-methylation occur prior to the glycosylation step however *O*-methylation generally occurs once the sugar is transferred (51). Several enzymes involved in TDP-sugar metabolism have been engineered to alter specificity and activity, however similar work has not been carried out on post-glycosyltransfer enzymes (52-54).

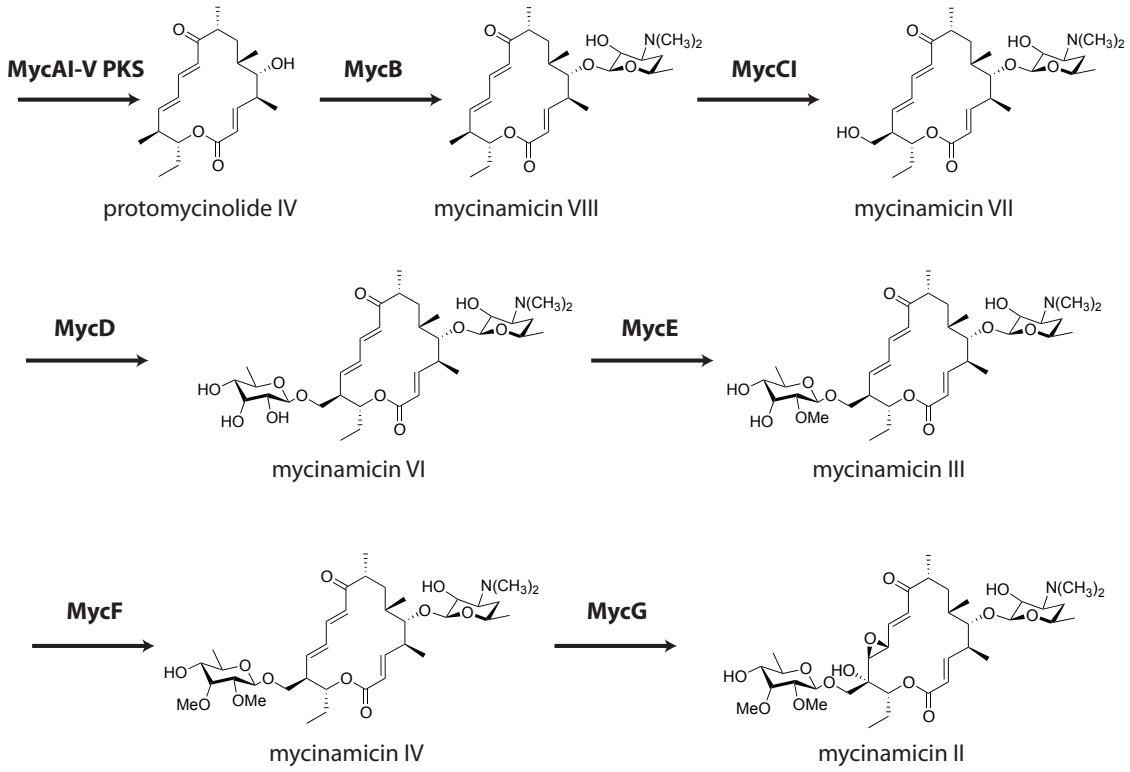


Figure 1-4 Post-PKS tailoring steps in the mycinamicin biosynthetic pathway (55). Sequential reactions in the mycinamicin pathway modify the macrolactone through hydroxylation, epoxidation, glycosylation and sugar methylation. Sugar *O*-methylation occurs after the glycosylation step.

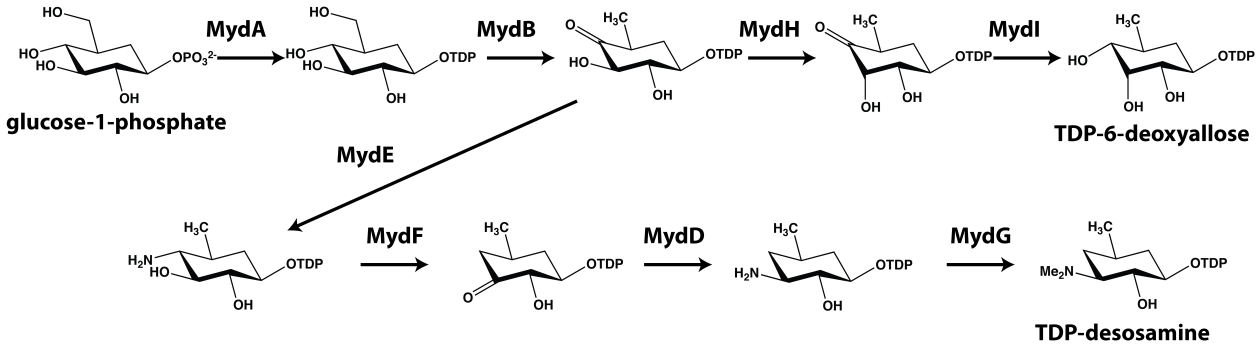


Figure 1-5 Enzymatic steps converting glucose-1-phosphate to functionalized sugars used in the biosynthesis of mycinamicin II (50).

Natural Product Sugar *O*-methyltransferases

Natural product sugar *O*-methyltransferases act after glycosyltransfer occurs. There are two families of sugar *O*-methyltransferases which were both originally characterized in the tylosin biosynthetic pathway. TylE is the founding member of the 6-deoxyallose 2-*O*-methyltransferase family, and TylF is the founding member of the javosyl-3-*O*-methyltransferase family (56, 57). Both the TylE and TylF families are *S*-adenosyl methionine (SAM) and magnesium dependent methyltransferases (57, 58). Despite these similarities the two methyltransferases belong to two distinct families and have very low (13%) sequence identity. Representatives of the two families from the mycinamicin biosynthetic pathway have been characterized *in vitro* and have different sensitivity to chelating agents, and different pH and temperature activity profiles (58). Members of the TylE family methylate at the 2′ or 3′ position of sugars while members of the TylF family methylate at the 3′ or 4′ position (58-60). In the multi-methylating systems that have been characterized *in vitro*, TylE family members act before TylF family members (56, 58, 60). There are annotated biosynthetic pathways that contain up to three sugar *O*-methyltransferases resulting in mono-, di- and tri-*O*-methylated sugars. Subsequent modifications occur to either the aglycone or the sugar in some pathways.

One member of each family has been characterized structurally. MycE, the 2′-*O*-methyltransferase from the mycinamicin biosynthetic pathway, is a compact tetramer (61). The active site occurs at the interface between three subunits of the tetramer and undergoes significant ordering upon binding the cosubstrate SAM and substrate mycinamicin VI. In the active site the 2′ and 3′ hydroxyl groups of the substrate sugar coordinate the magnesium ion, and a conserved histidine residue forms a hydrogen bond with the methyl acceptor and is

hypothesized to serve as a catalytic base (61). NovP is a TylF homolog that methylates the 4′-hydroxyl of the aminocoumarin antibiotic novobiocin (59). The structure of NovP with *S*-adenosyl-L-homocysteine (SAH) and magnesium indicates that TylF family members are dimeric and have structurally independent active sites (62). Without substrate present in the NovP crystal structure, it is unclear how MycE and NovP direct methylation to different sugar hydroxyl groups. Furthermore, the mechanisms by which TylE and TylF family members select the properly methylated substrate and the differences in the regiochemistry of methyltransfer between homologs (i.e., 2′ or 3′ for MycE and 3′ or 4′ for MycF) from different pathways remain unexplained. Insights into these areas will pave the way for pathway and protein engineering to produce new natural products and will increase the general understanding of these complex biosynthetic pathways.

PKS Bioengineering

The potential utility of natural products and challenges to fully synthetic production make the pathways that produce these molecules desirable targets for bioengineering (31). Furthermore, the biosynthetic pathways are frequently found clustered in the genome, aiding in their identification and simplifying manipulation of the pathway (63). These factors have inspired efforts to increase the productivity of natural pathways and engineer novel pathways to produce new natural product analogs. Several approaches have been used in engineering modular PKSs including swapping, introduction or reordering of full PKS modules, inserting, deleting or swapping individual domains, or substituting specific amino acids to alter catalytic activity or substrate specificity (64, 65). Although each approach has resulted in the production of new molecules, the amount of product observed is often significantly lower

than natural pathways (66). Recent advances in the understanding of intermodular substrate transfer in PKS pathways (67, 68) and the overall architecture of PKS modules may facilitate more efficient PKS engineering (69, 70). Engineering efforts have also focused on tailoring enzymes using gene knockout or knock-in approaches to target specific catalytic steps (66). Understanding the substrate specificity of natural product biosynthetic enzymes and identifying promiscuous or highly active variants will be key to the construction of highly productive engineered pathways.

Thesis Overview

The following chapters describe the structural characterization of the mycinamicin 3′-*O*-methyltransferase MycF. The series of structures illustrate the large conformational change required for cofactor exchange and include the first substrate complex of a TylF family member. Additionally, the determinants of substrate specificity and regiochemistry for this family of enzymes are discussed. Through biochemical experiments the substrate flexibility of the TylF/MycF family is explored and amino acid substitutions that increase activity and relax substrate specificity are described. This work represents the first example of enzyme engineering for post-glycosyl transfer enzymes in natural product biosynthetic pathways and demonstrates the utility of a structure-guided approach. The structural characterization of the tylosin thioesterase domain is also described. This first representative structure of a 16-membered ring forming thioesterase is compared with previously characterized 14-membered ring forming TEs. The potential applications for pathway engineering and biocatalytic strategies are discussed.

Chapter 2 Structural Basis of Substrate Specificity and Regiochemistry in the MycF/TylF Family of Sugar O-Methyltransferases.

Summary

Sugar moieties in clinically useful natural products are frequently modified by *O*-methylation. Two families of metal- and *S*-adenosyl methionine (SAM)-dependent enzymes commonly catalyze these modifications. In the biosynthesis of the macrolide antibiotic mycinamicin, members of each family catalyze methylation first at the 2′- and then the 3′-hydroxyl groups of a 6′-deoxyallose sugar. The structural basis of pathway ordering and substrate specificity is unknown. We solved a series of crystal structures of MycF, the 3′-*O*-methyltransferase, including the free enzyme and a binary complex with *S*-adenosyl homocysteine (SAH), and ternary complexes with SAH and the natural macrolide substrate, the product, an upstream pathway intermediate and an unnatural substrate. These structures present the first view of a MycF family member in complex with its substrate. In the absence of SAM/SAH, much of the active site lid is disordered. SAM binding induces substantial ordering that creates the binding site for the natural substrate, and a bound metal ion plays a central role in positioning the substrate for catalysis. The open active site suggests that MycF may accept other hydrophobic substrates attached to a javose sugar. Using the MycF substrate complex as a guide, we modeled the substrate complex of a 4′-specific enzyme, NovP. We identify active site residues that correlate with the 3′- or 4′- specificity of MycF family members and define the protein and substrate features that direct the regiochemistry

of methyltransfer. This classification scheme will be useful in the annotation of new secondary metabolite pathways that utilize this family of enzymes. Based upon the MycF substrate complex we identified single amino acid substitutions that relax the 2′-methoxy specificity and fortuitously increase the native activity of the enzyme with the natural substrate. Using this engineered variant we produced a new mycinamicin analog demonstrating the utility of structural information to facilitate bioengineering approaches for the chemoenzymatic synthesis of complex small molecules containing modified sugars. The timing and placement of these tailoring steps often impact final stage C-H functionalization reactions mediated by P450 monooxygenases.

Introduction

Highly modified sugars are commonly incorporated into macrolide natural products (71). Modifications, including *O*-, *N*- and *C*-methylations, acetylation, oxidation and epimerization, result in a diverse array of sugars not observed in primary metabolism (50). Modifications alter the ability of macrolide natural products to interact with targets and impact their solubility and membrane permeability. Natural product sugar *O*-methyltransferases have been shown to function *in vitro* on non-natural substrates, and with increased knowledge of the high resolution structures of substrate complexes, we anticipate the ability to more accurately assess the flexibility and specificity of this group of enzymes (72).

Sugars are incorporated into natural products as nucleotide mono or diphosphate conjugates originally derived from glucose or fructose. Thymidine diphosphate (TDP) conjugates are the most common in bacterial systems (50). These TDP-sugars are frequently

modified through epimerization, dehydration, reduction, amination or methylation (49). They are then transferred to natural product scaffolds by glycosyltransferases (51). Once incorporated, additional modifications including O and N methylation or acetylation are common. Many enzymes involved in TDP sugar modification have been structurally and functionally characterized, including those involved in O, N and C methylation of TDP-sugars (71). These enzymes may be particularly useful in pathway engineering as they recognize the same TDP moiety to select substrates. In contrast, sugar modifying enzymes that act after glycosyl transfer may interact with functional groups specific to their individual substrates limiting their utility designing new pathways. There is very little structural or functional information on how downstream sugar methyltransferases interact with their substrates leaving their potential for bioengineering unclear.

Mycinamicins, which are produced by *Micromonospora griseorubida,* are a clinically promising group of macrolide antibiotics because they exhibit better activity against antibiotic resistant bacterial strains than clinically used antibitiotics *(23, 55)*. They are composed of a 16-membered macrolactone core with two *O*-linked modified sugars. The mycinose sugar on the final product is derived from 6-deoxyallose. Following transfer to the macrolide core, the sugar is modified sequentially by two magnesium- and *S*-adenosylmethionine (SAM)-dependent methyltransferases. MycE methylates the 2′-hydroxyl of 6′-deoxyallose to form javose and MycF methylates javose at the 3′ position to generate the mycinose sugar (Figure 2-1) (58). MycE and MycF carry out essentially the same reaction but are highly specific for their respective substrates: MycE methylates only the 2′ sugar hydroxyl group, MycF methylates only the 3′ hydroxyl, and MycF is specific for 2′-methoxy sugars. While MycE and MycF have low sequence identity, each has many

homologs in annotated bacterial secondary metabolite pathways that catalyze sugar methylation in macrolide, anthracycline, and aminocoumarin antibiotics in addition to glycopeptidolipids (73-79). All characterized MycE homologs methylate sugar 2′ or 3′ hydroxyl groups while MycF homologs methylate at the 3′ or 4′ hydroxyl positions (59, 60). We previously reported crystal structures of MycE (2′-O-methyltransfease) in binary and ternary complexes with co-substrate and substrate (61). The substrate and metal-free structure of NovP, a MycF homolog, has been reported and the catalytic activity of the homolog TylF has been characterized (56, 57). However the substrate specificity and regioselectivity of the MycF methyltransferase family remains unclear (61, 62). Here we report a series of nine crystal structures with corresponding biochemical analysis that provide insights into substrate specificity, pathway ordering, and mechanism of MycF that is critical for construction of the fully elaborated macrolide antibiotic mycinamicin II. Furthermore, we demonstrate the utility of a structure guided approach to generating new macrolide antibiotics.

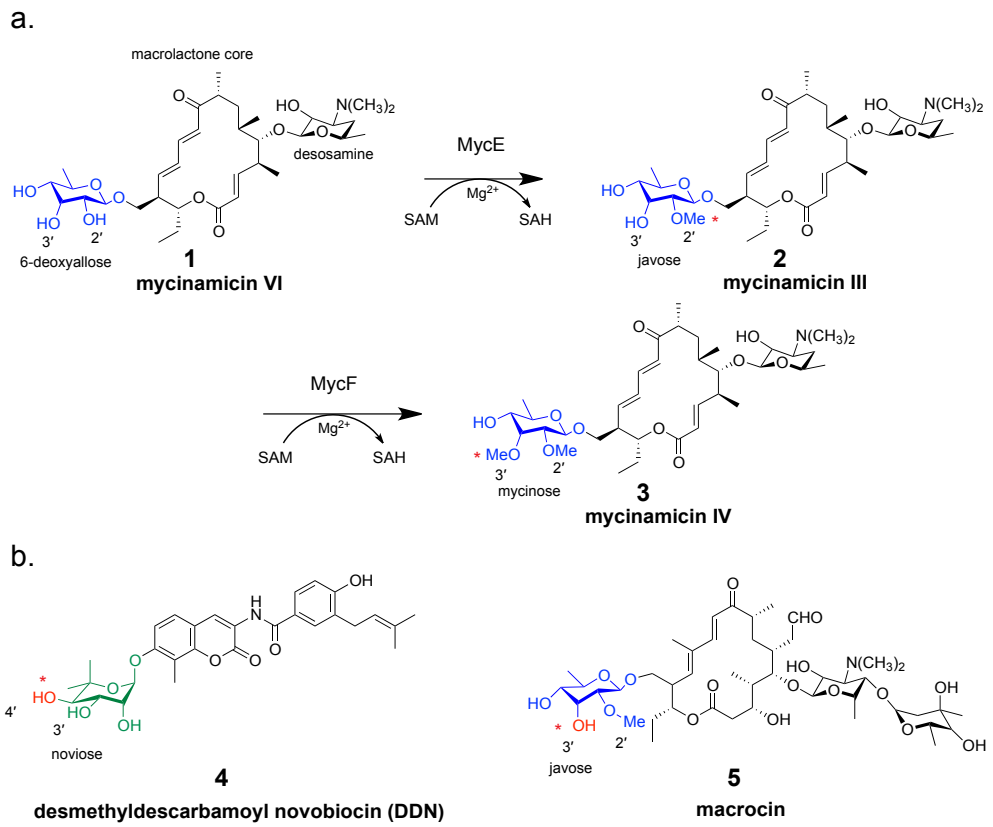


Figure 2-1 Sugar methylation in natural product biosynthesis. a) Ordered sugar *O*-methylation in the mycinamicin pathway. MycE catalyzes the 2′-*O*-methylation of the 6′-deoxyallose of mycinamicin VI (1) to generate the javose of mycinamicin III (2). MycF catalyzes the 3′-*O*-methylation of javose to form the mycinose of mycinamicin IV (3). The modified sugar is highlighted in blue and sites of methylation are denoted with red asterisks. Both enzymes are SAM and Mg^{2+} dependent. b) Substrates for the 4′-*O*-methytransferase NovP, desmethyldescarbamoyl novobiocin (4), and the 3′-*O*-methytransferase TylF, macrocin (5). The sites of methylation are denoted with red asterisks.

Methods

Protein Expression, Purification and Mutagenesis.

The bacterial pellet from a 1 L of culture *E. coli* bearing pET28b-mycF (58) was resuspended in 35 mL of Buffer A (50 mM Tris pH 8.0, 300 mM NaCl, 10% glycerol) with 5 mg lysozyme, 2 mg DNase and 2 mM $MgCl_2$, incubated for 30 minutes and lysed by

sonication. The soluble fraction was loaded onto a 5 mL HiTrap Ni NTA column (GE Healthcare) and bound protein was eluted with a gradient from 20-400 mM imidazole in Buffer A. Fractions containing MycF were pooled, concentrated to 5 mL and loaded onto a HiLoad 16/60 Superdex S200 gel filtration column (GE Healthcare) pre-equilibrated with a buffer containing 20 mM Tris pH 8.0, 150 mM NaCl and 10% glycerol. The apparent molecular weight is 66 kDa consistent with a MycF dimer. Fractions were pooled and concentrated to 20 mg/mL and flash frozen in liquid N_2. The single substitution variants were made with the Quik Change II Site-Directed Mutagenesis Kit (Stratagene) and the E35Q/E139A double mutant was made using the QuikChange Lightning Multi Site Directed Mutagenisis Kit (Strategene) following the manufacturers protocols using pET28b-mycF as the template. Forward and reverse primers can be found in Table 2-1**.** DNA sequences were verified at the University of Michigan DNA Sequencing Core. MycF variants were expressed and purified in the same manner as WT MycF.

Table 2-1 Mutagenesis primers

D191A	5′-GGATGGACGGCGCCTCCTACGGCGC-3′
	5′-GCGCCGTAGGAGGCGCCGTCCATCC-3′
D191N	5′-CGGATGGACGGCAACTCCTACGGCG-3′
	5′-CGCCGTAGGAGTTGCCGTCCATCCG-3′
F118Y	5′-CGACTCGTTCCAGGGCTACCCGAAGATCAC-3′
	5′-GTGATCTTCGGGTAGCCCTGGAACGAGTCG-3′
L143A	5′-GTACAACGAGGCGGTCGATGCGCCGACCAGC-3′
	5′-GCTGGTCGGCGCATCGACCGCCTCGTTGTAC-3′
L143S	5′-ACCAGTACAACGAGGCGGTCGATTCTCCGACCAGCCTG-3′
	5′-CAGGCTGGTCGGAGAATCGACCGCCTCGTTGTACTGGT-3′
L143N	5′-ACCAGTACAACGAGGCGGTCGATAATCCGACCAGCCTG-3′
	5′-CAGGCTGGTCGGATTATCGACCGCCTCGTTGTACTGGT-3′
L143Q	5′-GGCGGTCGATCAGCCGACCAGCC-3′
	5′-GGCTGGTCGGCTGATCGACCGCC-3′
M56A	5′-GACCGTGGCGCACACCGCGATCGGAATGAAACGG-3′
	5′-CCGTTTCATTCCGATCGCGGTGTGCGCCACGGTC-3′
E35Q	5′-GGCCTGATCACCCAGGCCGCGTTCG-3′
E139A	5′-CCACCAGTACAACGCGGCGGTCGATCTGC-3′

Crystallization.

Initial crystals of MycF were grown by hanging-drop vapor diffusion with a 1:1 ratio of protein solution (10 mg/mL MycF in 20 mM Tris 8.0, 150 mM NaCl, 10% glycerol, 1 mM SAH) and well solution (20-30% PEG 3350, 100 mM BisTrisPropane pH 6.5). These crystals were used to solve structures of the free enzyme, SAH binary complex, and SAH substrate ternary complex. However, the diffraction was poorly reproducible and streaky, and the crystals were non-isomorphous. Remarkably, crystals grew in three space groups (C2, $P2_12_12_1$, and $P22_12_1$), and the two orthorhombic crystals shared nearly identical unit cell constants (Table 2-2). Poor lattice contacts in the b direction involving the Glu35 and Glu139 side chains likely caused the streaky diffraction and led to variable orthorhombic space groups, which differed by translation of layers of molecules in the a-c plane (Figure 2-2). Several double mutants (E35Q/E139A, E35N/E139A and E35N/E139Q) were designed in an effort to improve crystal quality and reproducibility. The E35Q/E139A mutant resulted in crystals of space group $P2_12_12_1$ with less streaky diffraction and a d_{min} of 2 Å or better with similar unit cell constants to wild type orthorhombic crystals. The crystals were grown in the same manner as wild type crystals using a well solution containing 20-30% PEG 5000 MME, 100 mM ammonium acetate, and 100 mM BisTrisPropane pH 6.5. These crystals were used to solve structures of MycF in complex with mycinamicin IV (3), mycinamicin VI (1) and macrocin (5). For mycinamicin III (2), mycinamicin IV (3), mycinamicin VI (1) and macrocin (5) complexes the crystals were soaked 4-6 hours in well solution supplemented with 10 mM ligand. For data collection the crystals were

harvested on MicroMesh mounts (MiTeGen), cryoprotected in well solution plus 10% glycerol and flash frozen in liquid N_2.

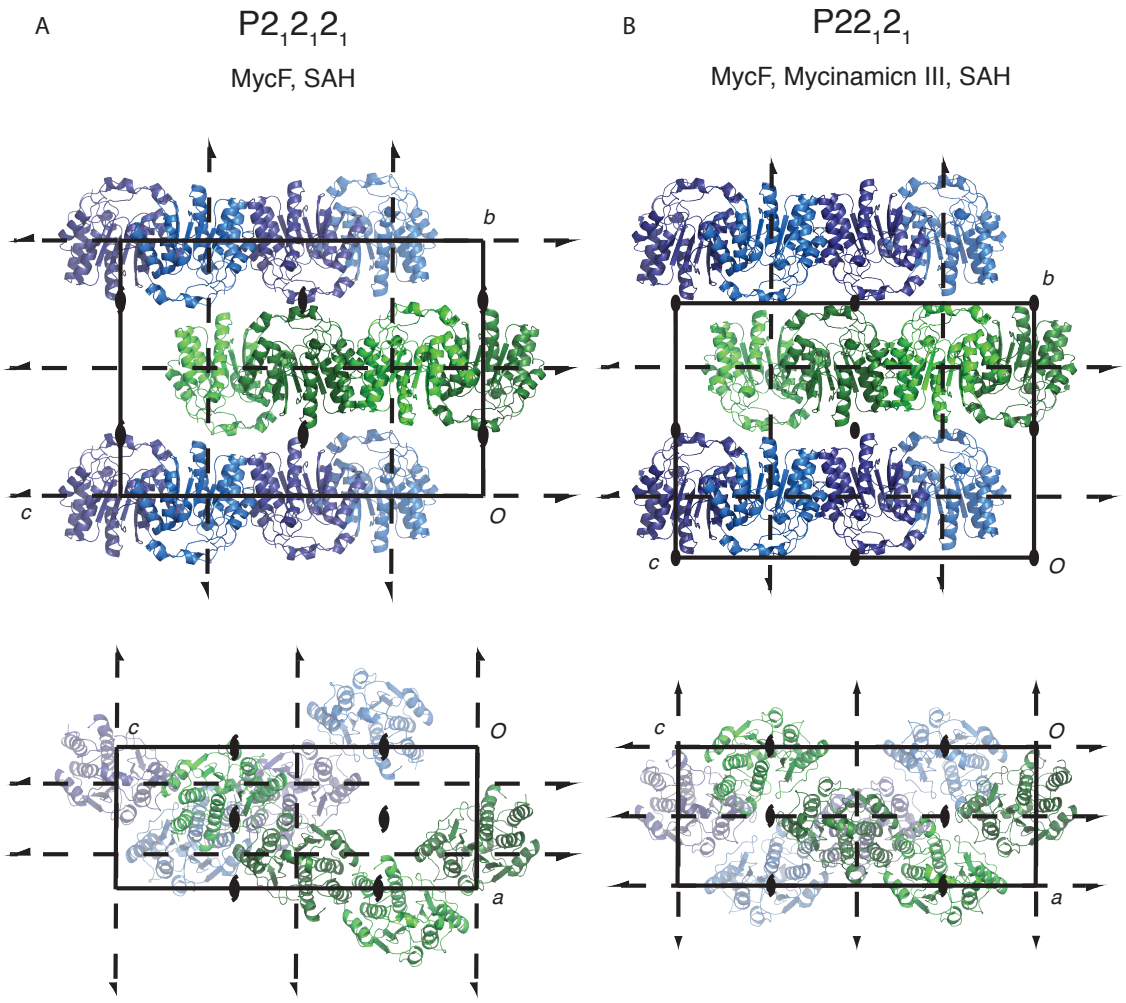


Figure 2-2 Crystals of MycF grown under the same conditions have distinct space groups but share a single set of cell constants. A) Crystal packing of the MycF SAH complex in $P2_12_12_1$ shown in two orthogonal views. B) The crystal packing of the MycF mycinamicin III SAH complex ($P2_12_12_1$) is shown in the same orientation. The difference in space group is caused by a translation of layers within the *a-c* plane, resulting in the loss of a 2-fold screw axis and formation of a 2-fold axis.

Data Collection and Structure Determination.

Data were collected at beamlines 23ID-B and 23ID-D (GM/CA) at the Advanced Photon Source. Diffraction data were collected using a 20 μM mini-beam with an oscillation of 0.5 degrees per frame using a 1 second exposure on a Mar300 CCD detector. 180 degrees of data were collected for each data set. The data processing and structure refinement statistics are shown in Table 2-2. The data were processed with HKL2000 (80) or XDS (Table 2-3, 2-4, 2-5, 2-6, 2-7, 2-8 and 2-9). The Se substructure of MycF was solved by the single wavelength anomalous diffraction and sites were found using HySS (81) and the initial model built with RESOLVE (82), both as implemented by AutoSol in the Phenix software suite (83). Twenty of 32 expected Se sites were located in the asymmetric unit. Coot (84) was used for model building and the models were refined with Refmac5 (85, 86) in the CCP4 suite (87, 88). A single copy of MycF was used as a search model for molecular replacement of subsequent complexes using Phaser (89). Mg^{2+} was modeled into all active sites and Refmac5 was used to find coordination bonds. The Grade server (Global Phasing Ltd., http://grade.globalphasing.org/ (90)) was used to generate initial models and refinement restraints for mycinamicin III, mycinamicin IV, and macrocin. Waters were automatically added in Coot and manually edited. The resulting models were validated with MolProbity (91) (Figures 2-3, 2-4, 2-5, 2-6, 2-7, 2-8, and 2-9).

Table 2-2 Crystallographic Summary

Crystal Contents	MycF WT[1] Mg - Disordered lid	MycF WT[1] Mg, SAH	MycF WT Mycinamcin III (**2**), Mg, SAH
Data			
Space group	C2	$P2_12_12_1$	$P22_12_1$
Cell (Å)	123.86 148.59 66.86	50.06 89.89 127.66	49.86 90.85 128.64
angles (°)	90 120.23 90		
d_{min} (Å)	2.50 (2.50-2.49)	2.40 (2.49-2.40)	1.65 (1.68-1.65)
I/σ	13.3 (2.1)	15.1 (2.9)	12.6 (2.1)
R_{sym}	0.080 (0.540)	0.114 (0.572)	0.073 (0.588)
Multiplicity	5.8 (5.8)	6.4 (5.9)	3.6 (3.6)
Completeness	100.0 (100.0)	98.4 (89.1)	99.4 (99.3)
No. of unique reflections	36246	22836	71066
Refinement			
R_{work}/ R_{free}	0.198/ 0.226	0.233/ 0.273	0.150/ 0.178
RMSD bonds (Å)	0.009	0.006	0.009
RMSD angles (°)	1.189	0.961	1.164
Ramachandran (%)			
Allowed	99.48	100	100
Outliers	0.52	0	0
Average B-factors (Å^2)			
Protein	55.0	56.7	16.7
Ligands/Ions	49.3	56.0	28.7
Water	48.6	49.6	25.9

[1] The structures of APO MycF and MycF in complex with SAH were solved by David L. Akey.

Crystal Contents	MycF E35Q E139A Mycinamicin IV (**3**), Mg, SAH	MycF E35Q E139A Macrocin (**5**), Mg, SAH	MycF E35Q E139A Mycinamcin VI (**1**), Mg, SAH
Data			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell (Å)	50.16 92.59 128.55	50.23 91.62 128.47	50.27 91.87 128.28
angles (°)			
d_{min} (Å)	1.39 (1.41-1.39)	1.75 (1.81-1.75)	1.45 (1.50-1.45)
I/σ	31.2 (1.7)	11.5 (3.8)	22.0 (4.2)
R_{sym}	0.070 (0.705)	0.17 (0.715)	0.119 (0.591)
Multiplicity	6.5 (3.3)	7.2 (6.9)	6.9 (6.1)
Completeness	87.6 (40.9)	99.9 (100.0)	93.0 (94.9)
No. of unique reflections	120684	60744	105396
Refinement			
R_{work}/ R_{free}	0.145/0.167	0.162/0.196	0.180/0.209
RMSD bonds (Å)	0.0005	0.009	0.010
RMSD angles (°)	1.202	1.387	1.427
Ramachandran (%)			
Allowed	100	100	100
Outliers	0	0	0
Average B-factors (Å^2)			
Protein	13.9	21.1	21.2
Ligands/Ions	20.3	30.9	32.0
Water	33.8	36.0	32.0

Crystal Contents	MycF E35Q M56A E139A Mg, SAH	MycF E35Q M56A E139A Mycinamicin III (**2**), Mg, SAH	MycF E35Q M56A E139A Mycinamcin VI (**1**), Mg, SAH
Data			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell (Å)	50.39 89.93 128.67	50.35 92.52 127.80	50.27 91.75 127.77
angles (°)			
d_{min} (Å)	1.48 (1.48-1.40)	1.53 (1.53-1.44)	1.59 (1.69-1.59)
I/σ	12.9 (0.9)	11.6 (1.4)	12.9 (2.0)
R_{sym}	0.072 (0.901)	0.077 (0.561)	0.078 (1.03)
Multiplicity	6.5 (3.3)	5.8 (3.1)	6.6 (6.6)
Completeness	85.5 (43.7)	99.0 (93.6)	98.5 (96.7)
No. of unique reflections	99384	107004	79006
Refinement			
R_{work}/ R_{free}	0.166/0.189	0.161/0.178	0.175/0.195
RMSD bonds (Å)	0.007	0.010	0.008
RMSD angles (°)	1.251	1.467	1.354
Ramachandran (%)			
Allowed	100	100	100
Outliers	0	0	0
Average B-factors (Å^2)			
Protein	18.7	13.7	19.0
Ligands/Ions	14.8	18.7	33.2
Water	31.9	27.2	32.1

Table 2-3 Scaling Statistics for MycF WT Mycinamcin III (2), Mg, SAH

Shell Limit (Å)		Average	Average		Linear
Lower	Upper	I	error	Chi^2	R-factor
50	4.48	318.3	18.7	0.432	0.034
4.48	3.55	412.4	23.5	0.46	0.034
3.55	3.11	249.8	14.7	0.571	0.041
3.11	2.82	138.2	8.9	0.782	0.057
2.82	2.62	100.5	6.8	0.962	0.072
2.62	2.46	80.6	5.8	0.79	0.068
2.46	2.34	72.1	5.5	0.733	0.069
2.34	2.24	64.6	5.3	0.761	0.078
2.24	2.15	54.4	4.9	0.865	0.097
2.15	2.08	47	4.7	0.957	0.117
2.08	2.01	37.3	4.3	1.017	0.144
2.01	1.96	30.8	4.1	0.98	0.167
1.96	1.9	26.2	4	0.995	0.196
1.9	1.86	20.6	3.9	0.992	0.24
1.86	1.82	17.4	3.8	0.995	0.282
1.82	1.78	14.6	3.8	0.96	0.331
1.78	1.74	12.9	3.9	0.99	0.385
1.74	1.71	10.9	3.9	0.929	0.445
1.71	1.68	9.5	4	0.924	0.515
1.68	1.65	8.6	4.1	0.952	0.588
All reflections		87.9	7	0.851	0.073

Table 2-4 Scaling Statistics for MycF E35Q E139A Mycinamicin IV (3), Mg, SAH

Shell Limit (Å)		Average	Average		Linear
Lower	Upper	I	error	Chi^2	R-factor
50	3.77	819	18.1	1.488	0.042
3.77	2.99	642.3	14.1	2.012	0.052
2.99	2.62	295.5	6.9	2.165	0.065
2.62	2.38	210.6	5.2	1.677	0.06
2.38	2.21	178.8	4.7	1.971	0.074
2.21	2.08	143.3	4.1	2.08	0.086
2.08	1.97	105.7	3.4	1.801	0.091
1.97	1.89	82.1	3.1	1.583	0.099
1.89	1.81	56.1	2.6	1.451	0.118
1.81	1.75	43.4	2.4	1.384	0.14
1.75	1.7	35	2.3	1.299	0.168
1.7	1.65	27.8	2.2	1.282	0.207
1.65	1.6	23.7	2.5	1.232	0.235
1.6	1.57	20.5	3.1	1.186	0.272
1.57	1.53	17.6	3.3	1.136	0.329
1.53	1.5	16	3.6	1.092	0.37
1.5	1.47	13.9	3.8	1.022	0.43
1.47	1.44	11.5	4.1	0.97	0.528
1.44	1.41	9.9	4.5	0.91	0.597
1.41	1.39	8.3	4.9	0.897	0.705
All reflections		159.4	5.1	1.562	0.07

Table 2-5 Scaling Statistics for MycF E35Q E139A Macrocin (5), Mg, SAH

Shell Limit (Å)		Average	Average		Linear
Lower	Upper	I	error	Chi^2	R-factor
50	3.77	514.8	36	0.636	0.095
3.77	2.99	383.2	26.2	0.851	0.12
2.99	2.61	185.9	13.7	0.939	0.144
2.61	2.38	131.1	10.5	1.008	0.17
2.38	2.2	117.1	10.3	1.048	0.2
2.2	2.07	101.7	10.3	1.035	0.239
2.07	1.97	71.4	9.1	0.985	0.314
1.97	1.89	58.1	9.2	0.942	0.405
1.89	1.81	41.4	8.7	0.821	0.557
1.81	1.75	35	9.1	0.797	0.715
All reflections		166.8	14.5	0.907	0.17

Table 2-6 Scaling Statistics for MycF E35Q E139A Mycinamcin VI (1), Mg, SAH

Shell Limit (Å)		Average	Average		Linear
Lower	Upper	I	error	Chi^2	R-factor
50	3.77	476.4	20.5	0.771	0.052
3.77	2.99	304.3	13.9	1.022	0.064
2.99	2.61	124.8	6.2	1.191	0.091
2.61	2.38	79.1	4.7	1.214	0.116
2.38	2.2	63.4	5.5	1.431	0.163
2.2	2.07	45.2	4.3	1.482	0.228
2.07	1.97	30	3.8	1.481	0.313
1.97	1.89	23.2	4.2	1.49	0.418
1.89	1.81	13.3	3.5	1.756	0.702
1.81	1.75	10.1	3.4	1.752	0.881
All reflections		117.7	7	1.356	0.103

Table 2-7 Scaling Statistics for MycF E35Q M56A E139A, Mg, SAH

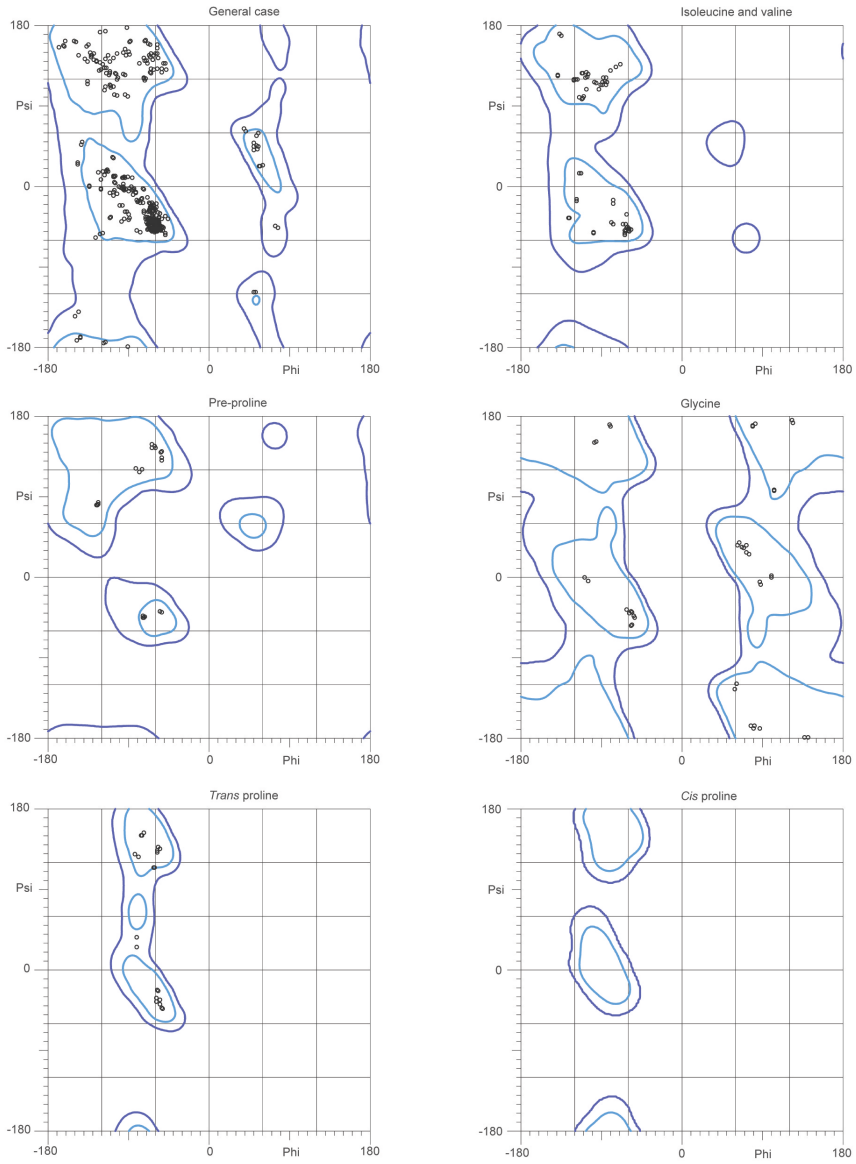
Resolution	Number of reflections			Completeness	R-factor	I/sigma	CC(1/2)
Limit	Observed	Unique	Possible				
4.18	28605	4666	4681	99.70%	4.40%	34.63	99.8
2.96	50288	8063	8072	99.90%	5.20%	31.93	99.8
2.42	66224	10295	10304	99.90%	6.50%	25.35	99.6
2.1	76714	12064	12075	99.90%	8.40%	19.16	99.5
1.87	89370	13660	13668	99.90%	12.90%	12.64	99.1
1.71	99705	15096	15097	100.00%	22.50%	7.27	98.1
1.58	78657	15295	16285	93.90%	35.80%	3.87	93.9
1.48	43219	12115	17488	69.30%	55.30%	1.84	81.6
1.4	21595	8130	18586	43.70%	90.10%	0.85	50.7
total	554377	99384	116256	85.50%	7.20%	12.9	99.8

Table 2-8 Scaling Statistics for MycF E35Q M56A E139A, Mycinamicin III (2), Mg, SAH

Resolution Limit	Number of reflections			Completeness	R-factor	I/sigma	CC(1/2)
	Observed	Unique	Possible				
4.31	27596	4362	4366	99.90%	4.10%	33.12	99.9
3.05	47414	7525	7531	99.90%	4.80%	30.1	99.8
2.5	61632	9560	9564	100.00%	6.30%	22.51	99.8
2.16	74626	11285	11285	100.00%	8.00%	17.94	99.7
1.93	82506	12678	12679	100.00%	10.60%	13.35	99.5
1.77	92205	14024	14025	100.00%	16.30%	8.66	99.0
1.64	100836	15188	15189	100.00%	26.60%	5.54	98.0
1.53	87434	16239	16246	100.00%	40.10%	3.15	93.8
1.44	49452	16143	17247	93.60%	56.10%	1.44	80.8
total	623701	107004	108132	99.00%	7.70%	11.57	99.9

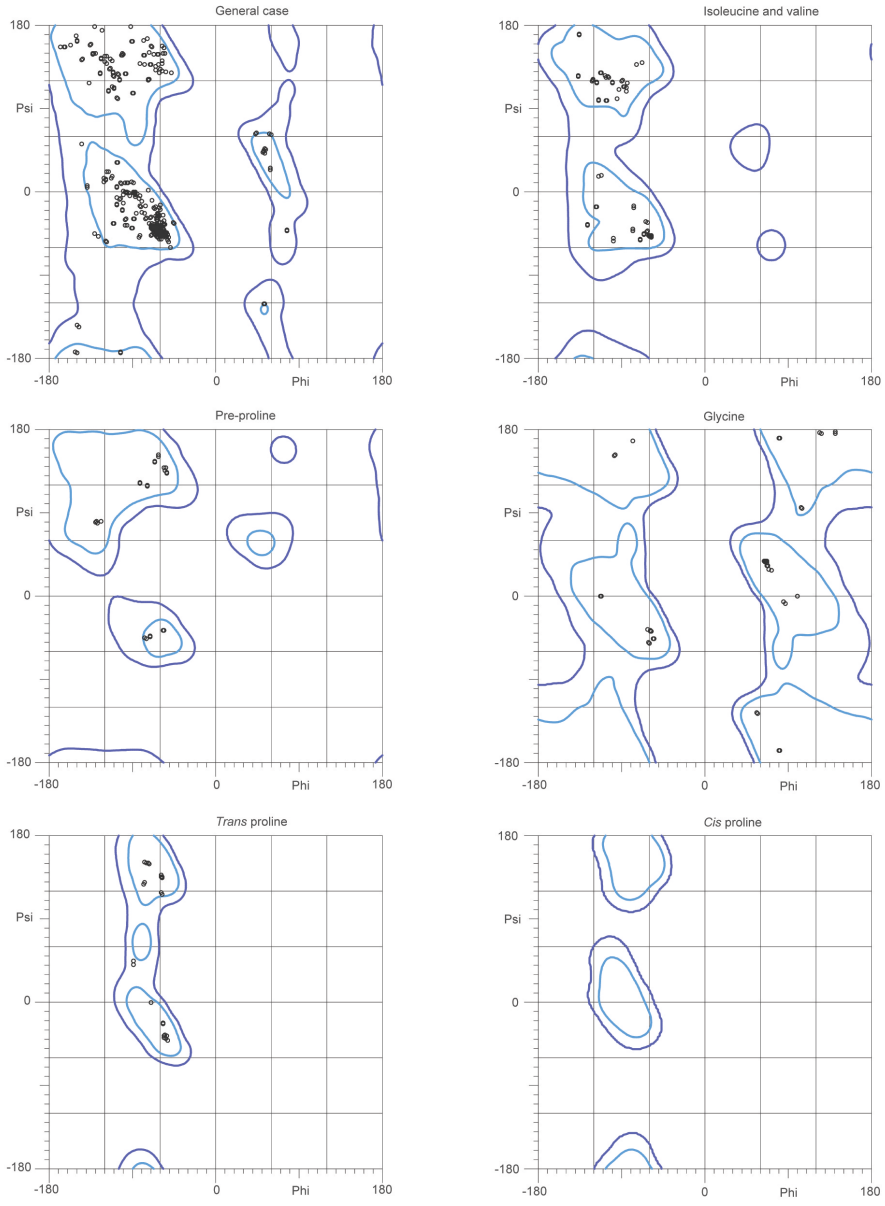
Table 2-9 Scaling Statistics for MycF E35Q M56A E139A, Mycinamcin VI (1), Mg, SAH

Resolution Limit	Number of reflections			Completeness	R-factor	I/sigma	CC(1/2)
	Observed	Unique	Possible				
4.75	19993	3241	3261	99.40%	3.70%	33.12	99.9
3.37	34769	5556	5580	99.60%	4.00%	30.1	99.8
2.75	46905	7092	7126	99.50%	5.20%	22.51	99.8
2.38	54520	8280	8334	99.40%	7.30%	17.94	99.7
2.13	62864	9353	9440	99.10%	10.50%	13.35	99.5
1.95	67623	10227	10358	98.70%	17.00%	8.66	99.0
1.8	74859	11064	11235	98.50%	32.20%	5.54	98.0
1.69	77378	11814	12047	98.10%	56.60%	3.15	93.8
1.59	81514	12379	12802	96.70%	97.70%	1.44	80.8
total	520425	79006	80183	98.50%	7.90%	12.91	99.9



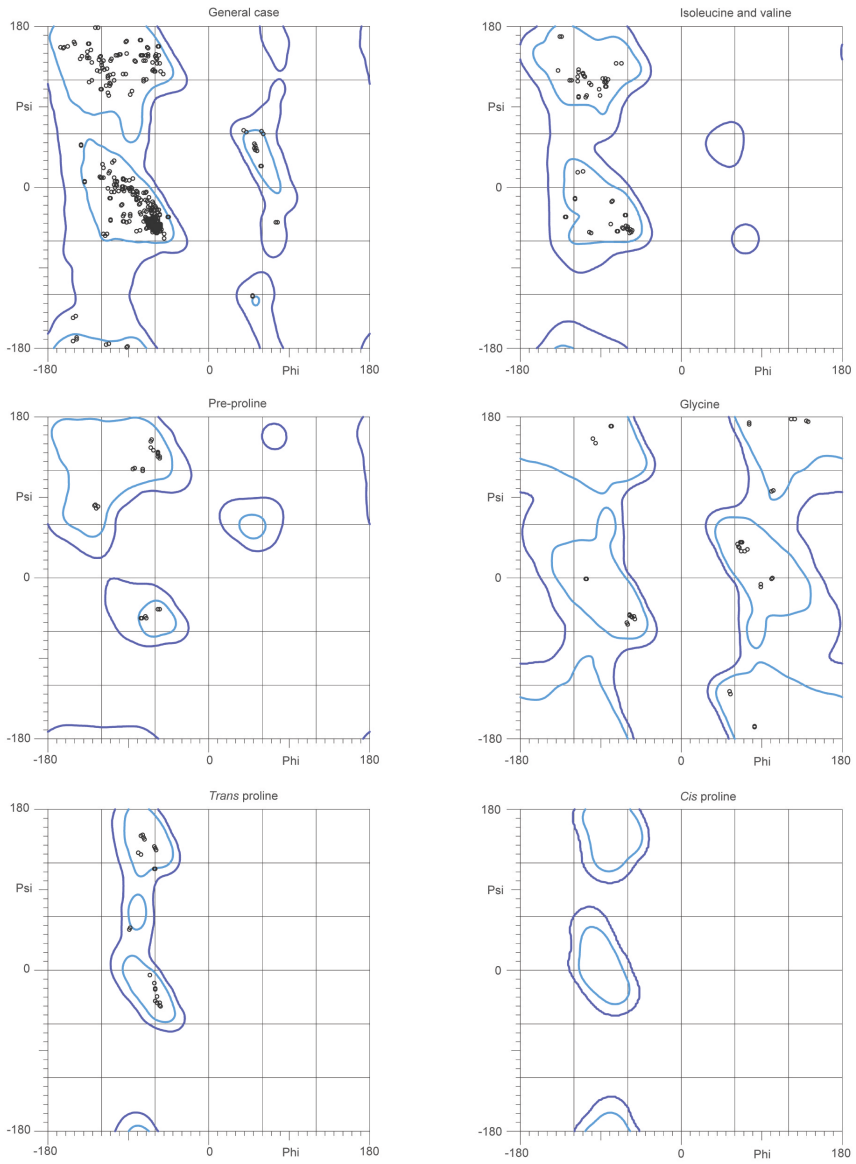
96.6% (517/535) of all residues were in favored (98%) regions.
100.0% (535/535) of all residues were in allowed (>99.8%) regions.

Figure 2-3 Ramachandran Plot for MycF WT, mycinamicin III (2), and SAH complex



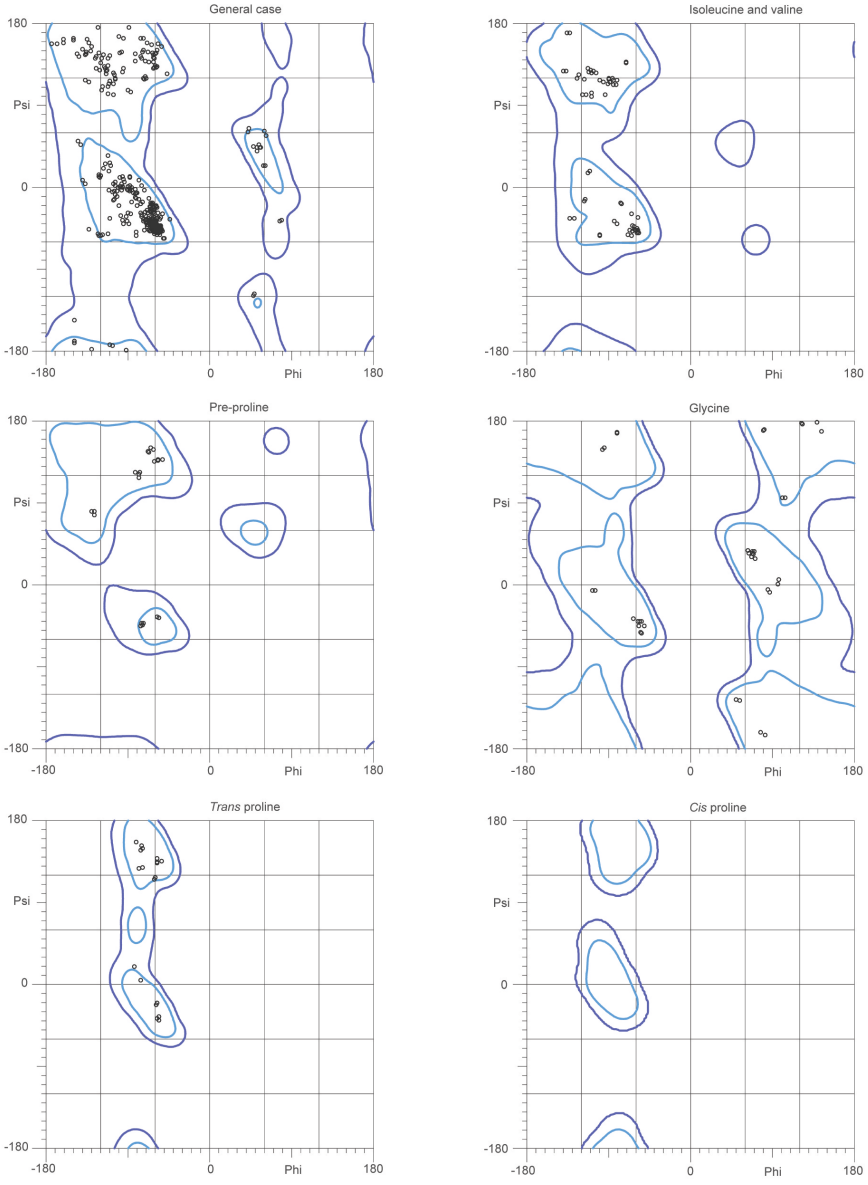
96.6% (517/535) of all residues were in favored (98%) regions.
100.0% (535/535) of all residues were in allowed (>99.8%) regions.

Figure 2-4 Ramachandran Plot for MycF E35Q E139A, mycinamicin IV (3), and SAH



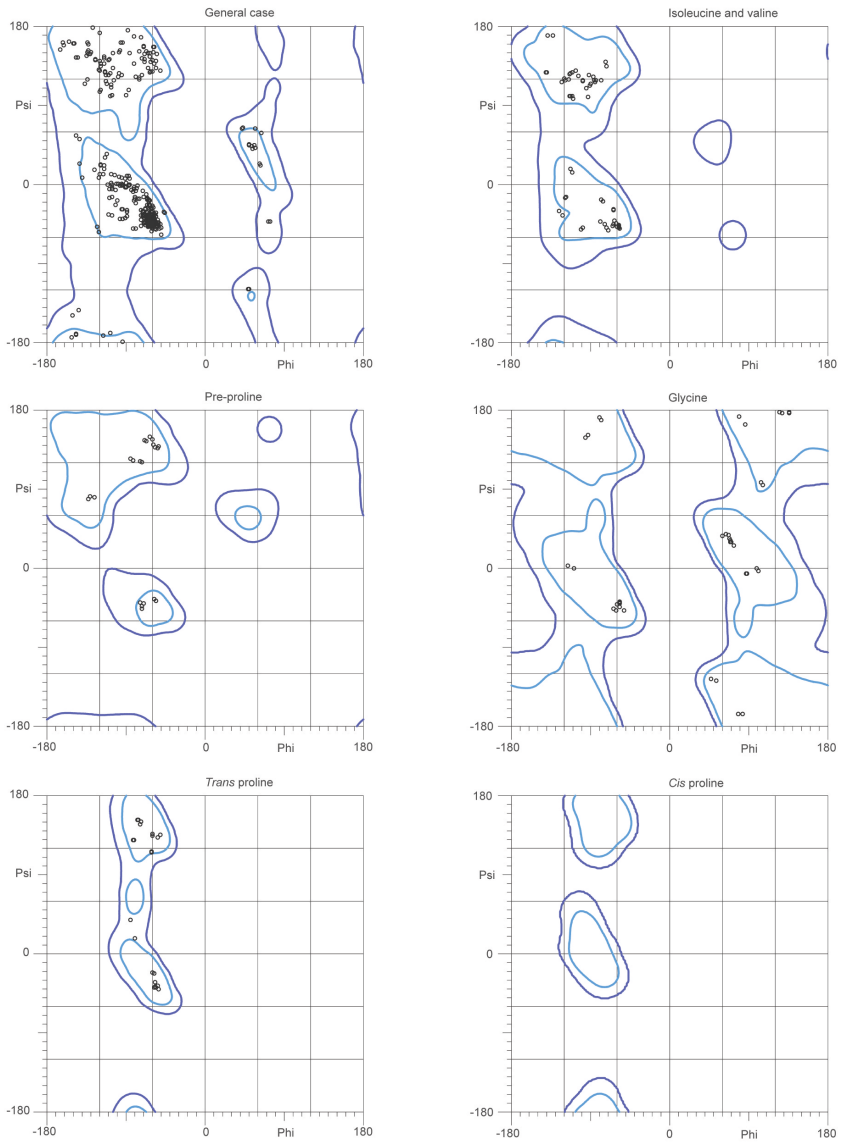
96.9% (531/548) of all residues were in favored (98%) regions.
100.0% (548/548) of all residues were in allowed (>99.8%) regions.

Figure 2-5 Ramachandran Plot for MycF E35Q E139A, Macrocin (5), and SAH



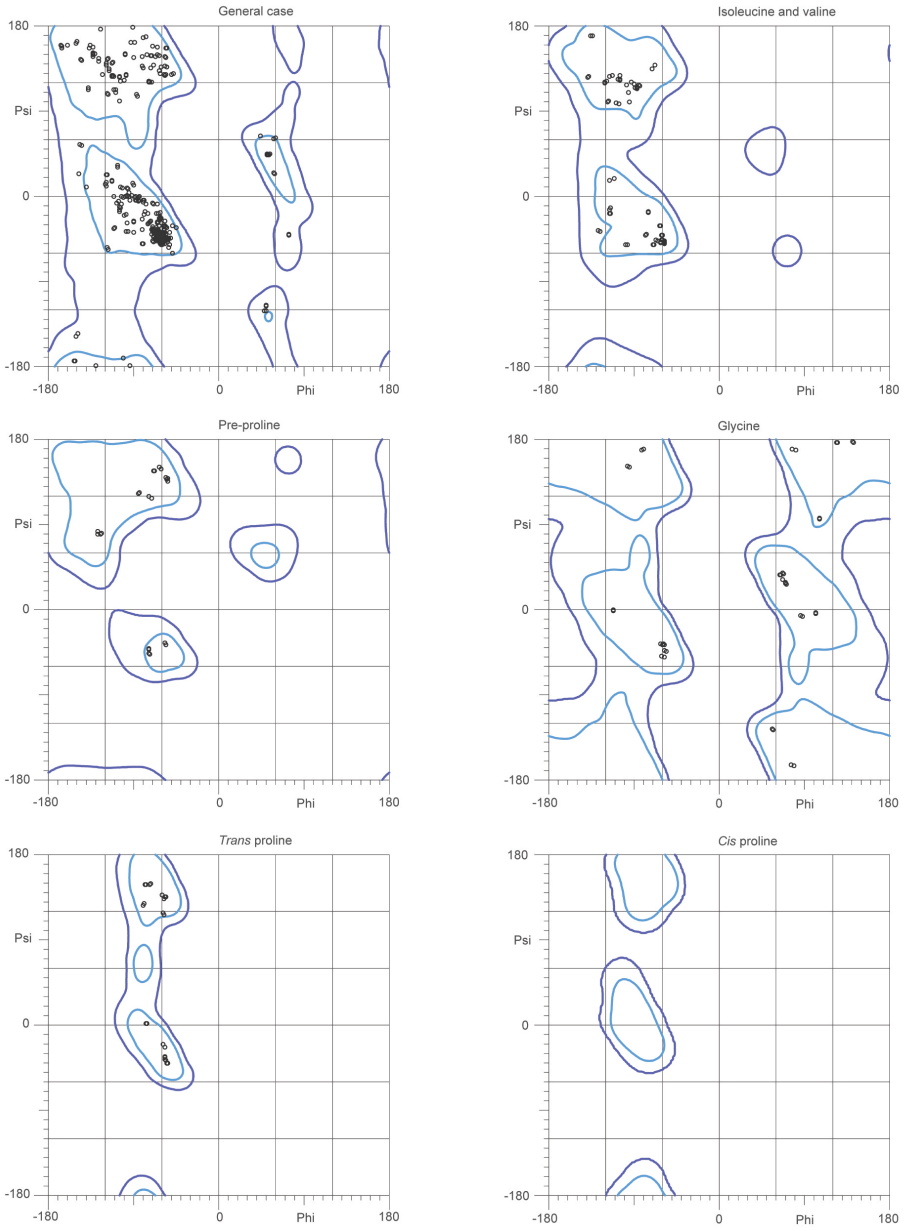
96.3% (500/519) of all residues were in favored (98%) regions.
100.0% (519/519) of all residues were in allowed (>99.8%) regions.

Figure 2-6 Ramachandran Plot for MycF E35Q E139A, mycinamicin VI (1), and SAH



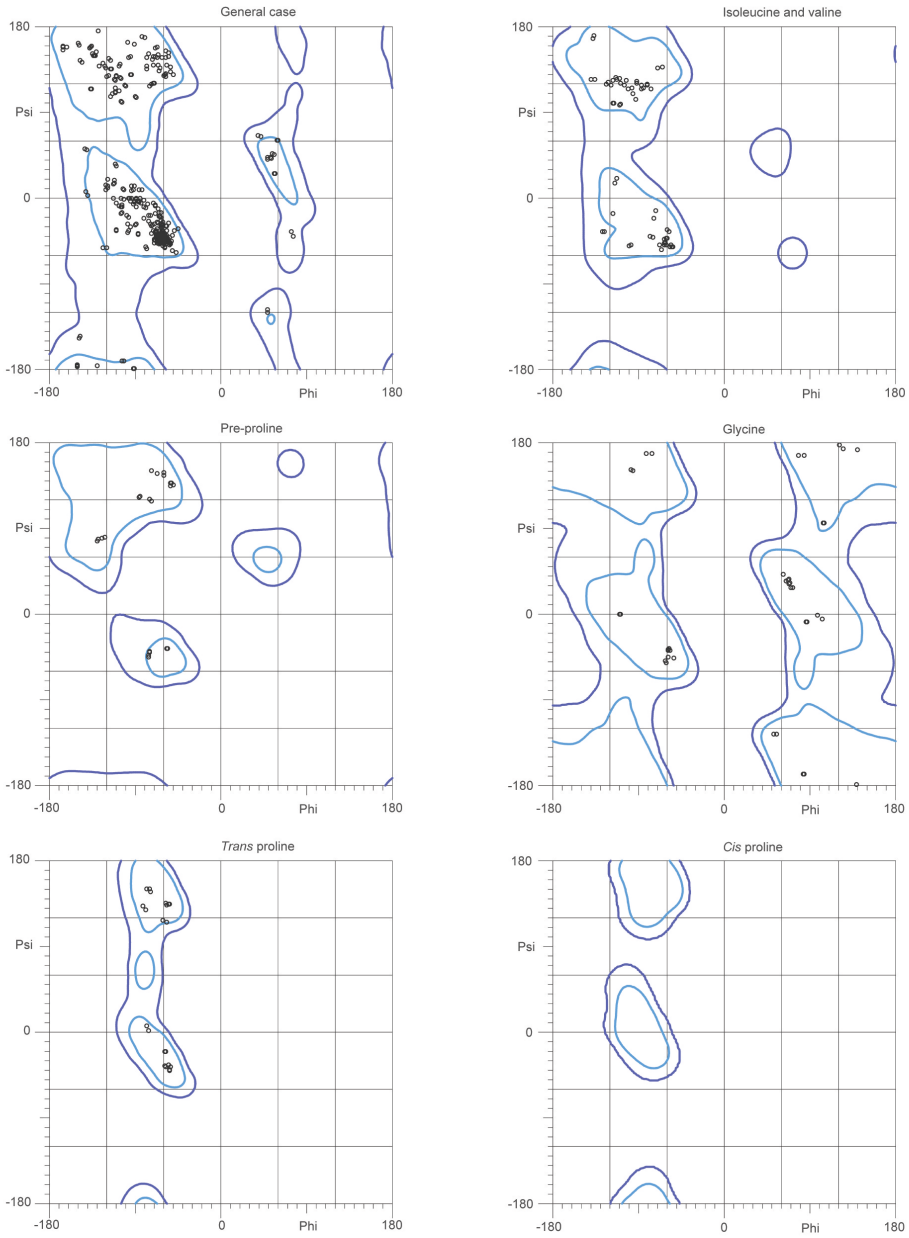
96.0% (485/505) of all residues were in favored (98%) regions.
100.0% (505/505) of all residues were in allowed (>99.8%) regions.

Figure 2-7 Ramachandran Plot for MycF E35Q M56A E139A, and SAH



96.8% (488/504) of all residues were in favored (98%) regions.
100.0% (504/504) of all residues were in allowed (>99.8%) regions.

Figure 2-8 Ramachandran Plot for MycF E35Q M56A E139A, mycinamicin III (2), and SAH



96.2% (483/502) of all residues were in favored (98%) regions.
100.0% (502/502) of all residues were in allowed (>99.8%) regions.

Figure 2-9 Ramachandran Plot for MycF E35Q M56A E139A, mycinamicin VI (1), and SAH

Sequence Alignments, Modeling, Free Energy Calculations and Figures.

MycF homologs were identified with BLAST (92) and verified manually. Sequences were aligned with ClustalW (93) and displayed in Jalview (94). The free energy calculations for mycinamicin III and mycinamicin VI conformations were performed using Spartan'10 version 1.10 (Spartan'10, Wavefunction, Inc. Irvine, CA (95)). Structure alignments and figures were made in PyMol (96). The Grade server (Global Phasing Ltd., http://grade.globalphasing.org/ (90)) was used to generate the initial model of desmethyldescarbamoyl novobiocin. Coot (84) was used for superposition of the MycF and NovP (PDB ID 2WK1) structures, remodeling of the active site residues and placing Mg^{2+} and substrate.

Enzyme Activity Assays.

Enzyme activity assays were carried out in 100 µL total volume containing 50 mM Tris buffer (pH 8.0), 5 µM MycF, 250 µM mycinamicin III, 10 mM $MgCl_2$. The reaction was initiated by addition of 750 µM SAM, incubated at 30 °C for 30 min and quenched by addition of 100 µL 100% methanol. Mycinamicin VI reactions were performed in the same manner with 20 µM MycF and incubated for 20 hours. Precipitated protein was removed by 30 minutes centrifugation (14000 x g) at 4 °C, and the components of the reaction mixture were seperated by reverse phase HPLC on an Agilent Zorbax C-18 column with a linear gradient of 20-100% acetonitrile, 0.1% formic acid over 20 min with a flow rate of 1 mL/min and the absorbance was monitored at 280 nm. The percent activity was quantified using the integrated peak area for the substrate and product to calculate the percent conversion. The percent conversion for each variant was normalized to wild type MycF to give percent

activity. Reactions were performed in duplicate. Components of reaction mixtures were identified in the HPLC elution trace by comparison with standards of mycinamicin VI, mycinamicin III, mycinamicin IV. The product of the MycF reaction with mycinamicin III and SAM has been characterized by LC-MS and reported previously (58).

Characterization of the New Mycinamicin Analog.

The new mycinamicin analog was prepared by incubating MycF M56A with 3 mg of mycinamicin VI under the previously described reaction conditions for 20 hours. The reaction was followed by analytical HPLC. The total reaction was dried under nitrogen and resuspended in 1 mL 100% methanol and insoluble material was removed by centrifugation. The supernatant was purified using preparative HPLC and fractions containing the new mycinamicin were pooled and dried under nitrogen. The purified product was dissolved in 300 μL deuterated acetone for NMR analysis. All NMR spectra were acquired on a Varian INOVA 600 MHz spectrometer at the Life Sciences Institute NMR Facility at the University of Michigan. Final chemical shift assignments were made using ^{1}H, ^{13}C, COSY, HSQCAD and HMBCAD spectra (Table 2-10 and Figures 2-10, 2-11, 2-12, 2-13, and 2-14).

Table 2-10 Chemical shifts for new mycinamicin analog

	δ_C	δ_H
C1		
C2	121.5	5.93
C3	151.4	6.54
C4	41.7	2.76
C5	88	3.36
C6	34	1.22
C7	33.18	1.72
C8	45.3	2.47
C9	203.1	
C10	123.7	6.46
C11	141.2	7.01
C12	133.5	6.24
C13	141.2	6.02
C14	49.8	2.53
C15	74.1	4.91
C16	23.6	1.59
C17	9.7	0.9
C18	19.4	1.23
C19	17.6	0.99
C20	17.7	1.14
C21	68.4	3.68
C1′	105.7	4.28
C2′	70	3.16
C3′	66.5	2.53
C4′	29.56	1.74
C5′	69.5	3.55
C6′	21.2	1.18
N(CH3)2	40.4	2.29
C1′′	101.9	4.5
C2′′	72.7	3.41
C3′′	82.6	3.65
C4′′	73.9	3.18
C5′′	70.9	3.15
C6′′	17.9	1.19
C7''	61.7	3.58

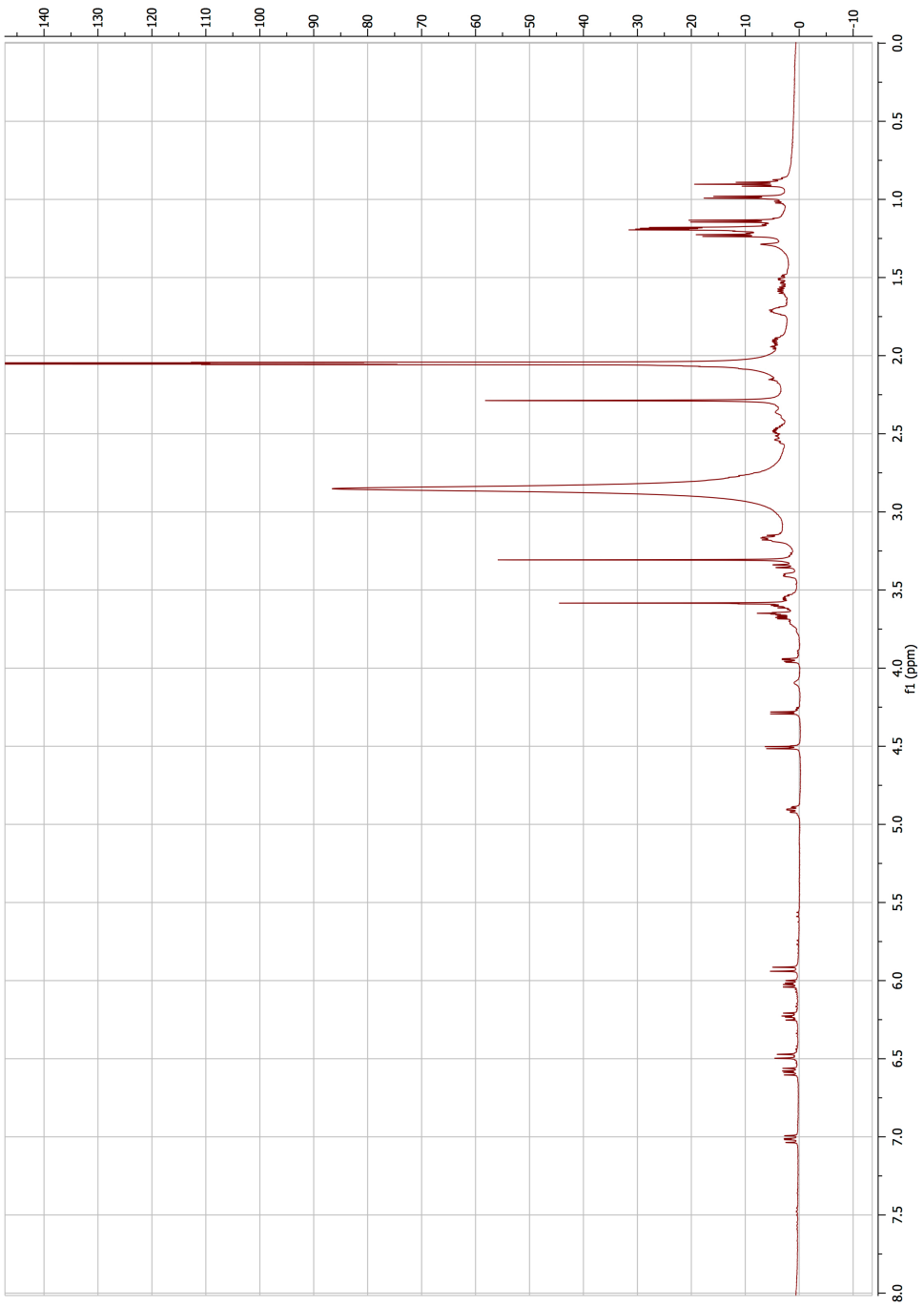


Figure 2-10 ^{1}H spectrum of new mycinamicin

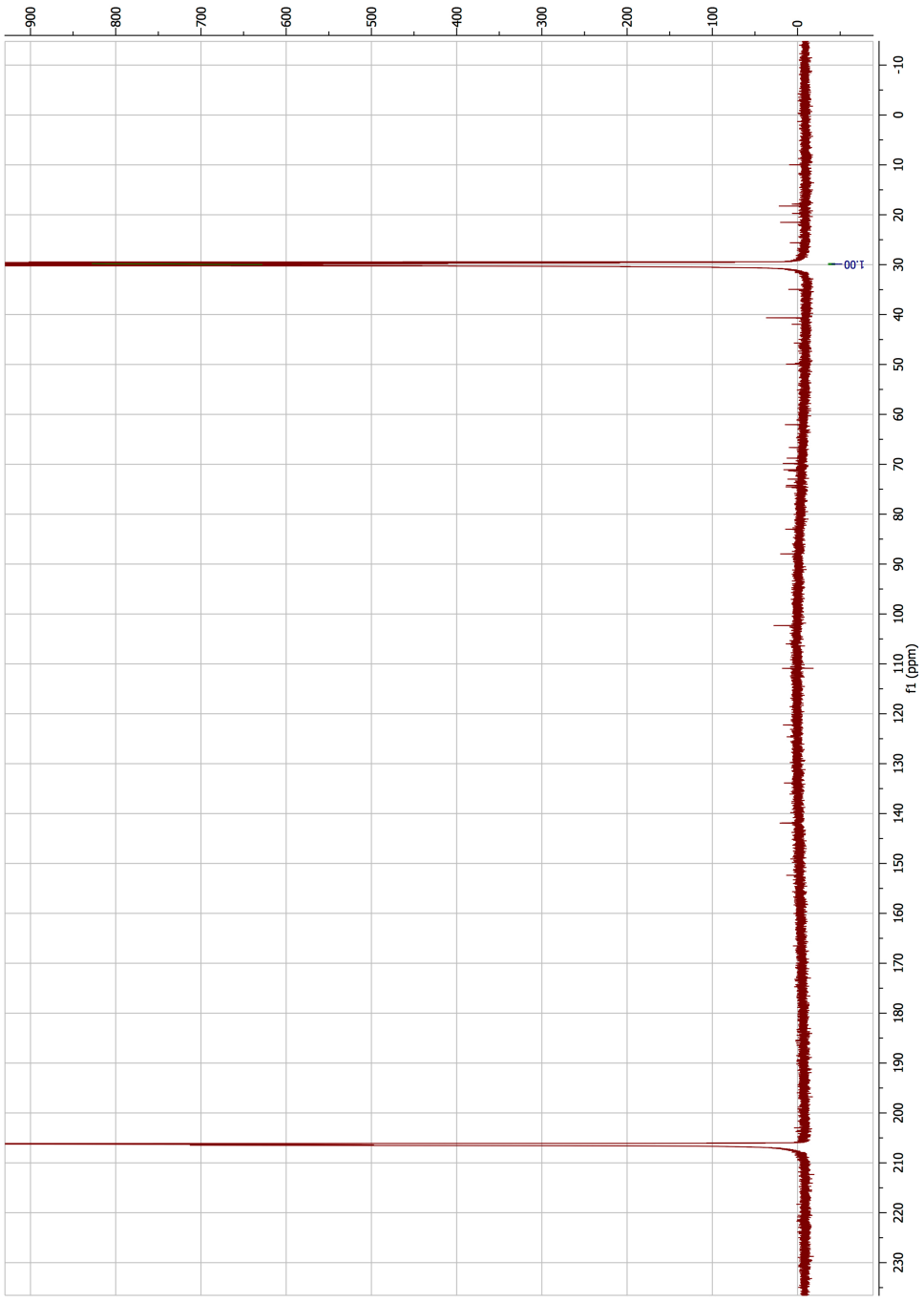


Figure 2-11 ^{13}C spectrum of new mycinamicin

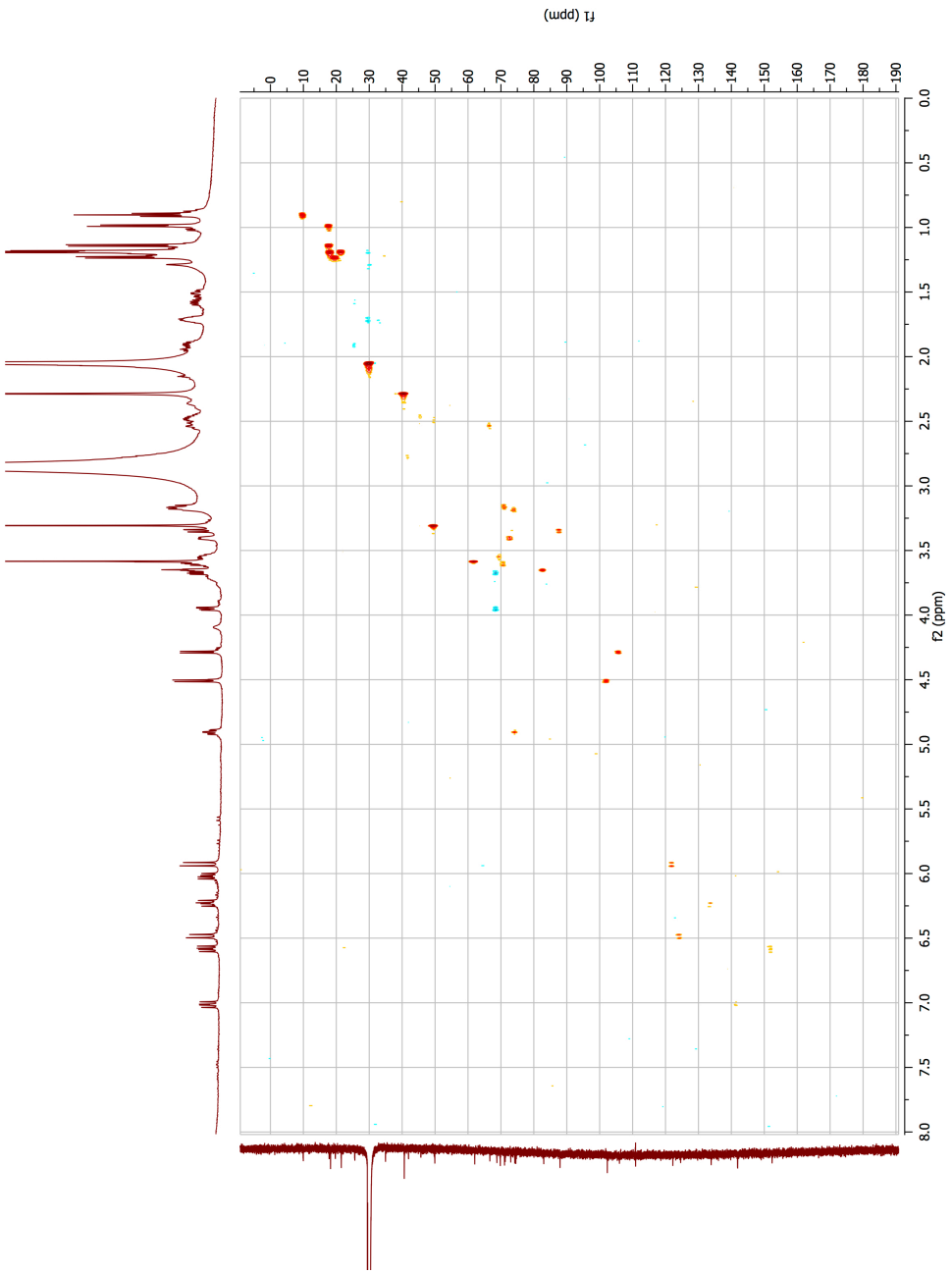


Figure 2-12 HSQCAD spectrum of new mycinamicin

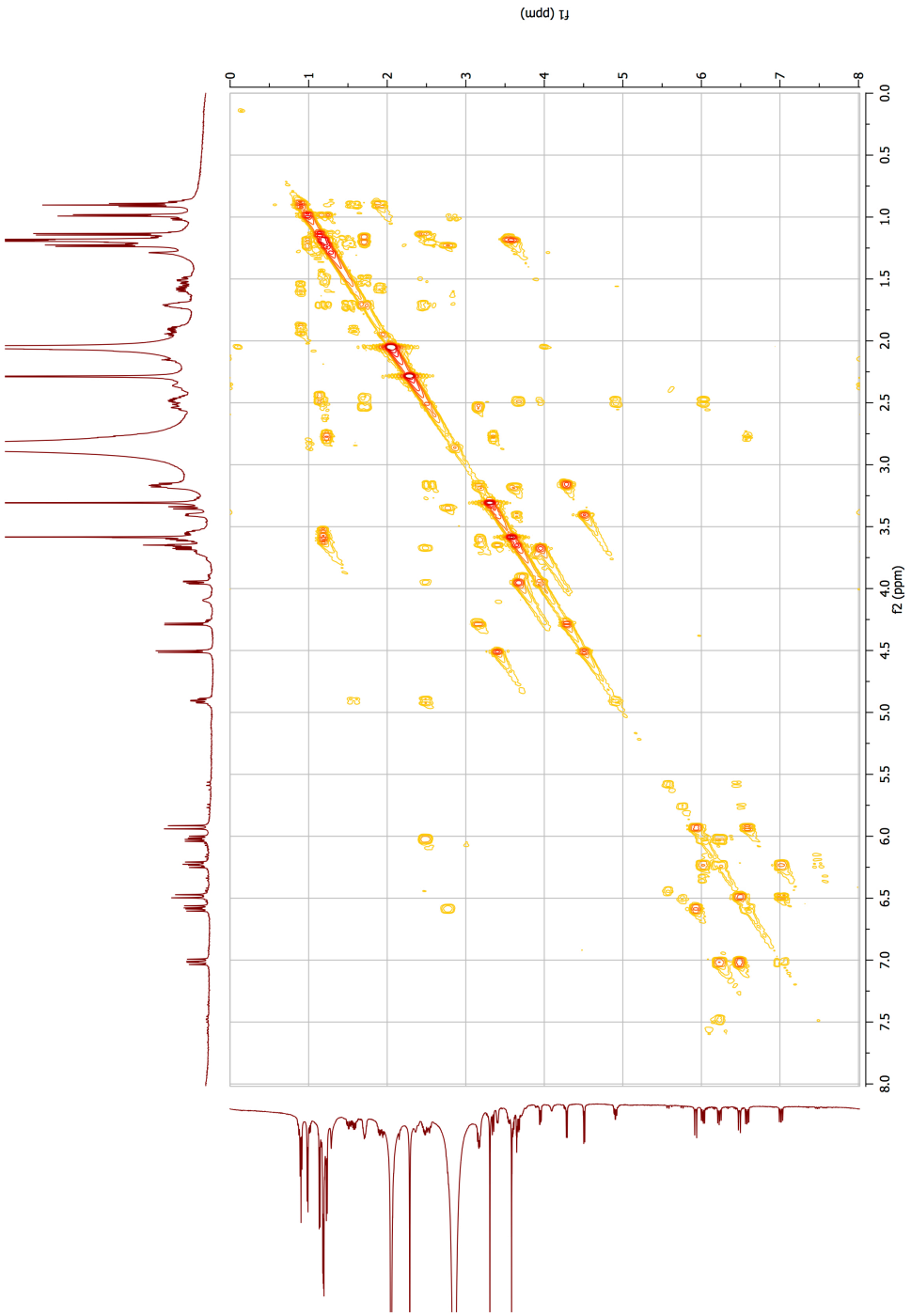


Figure 2-13 COSY spectrum of new mycinamicin

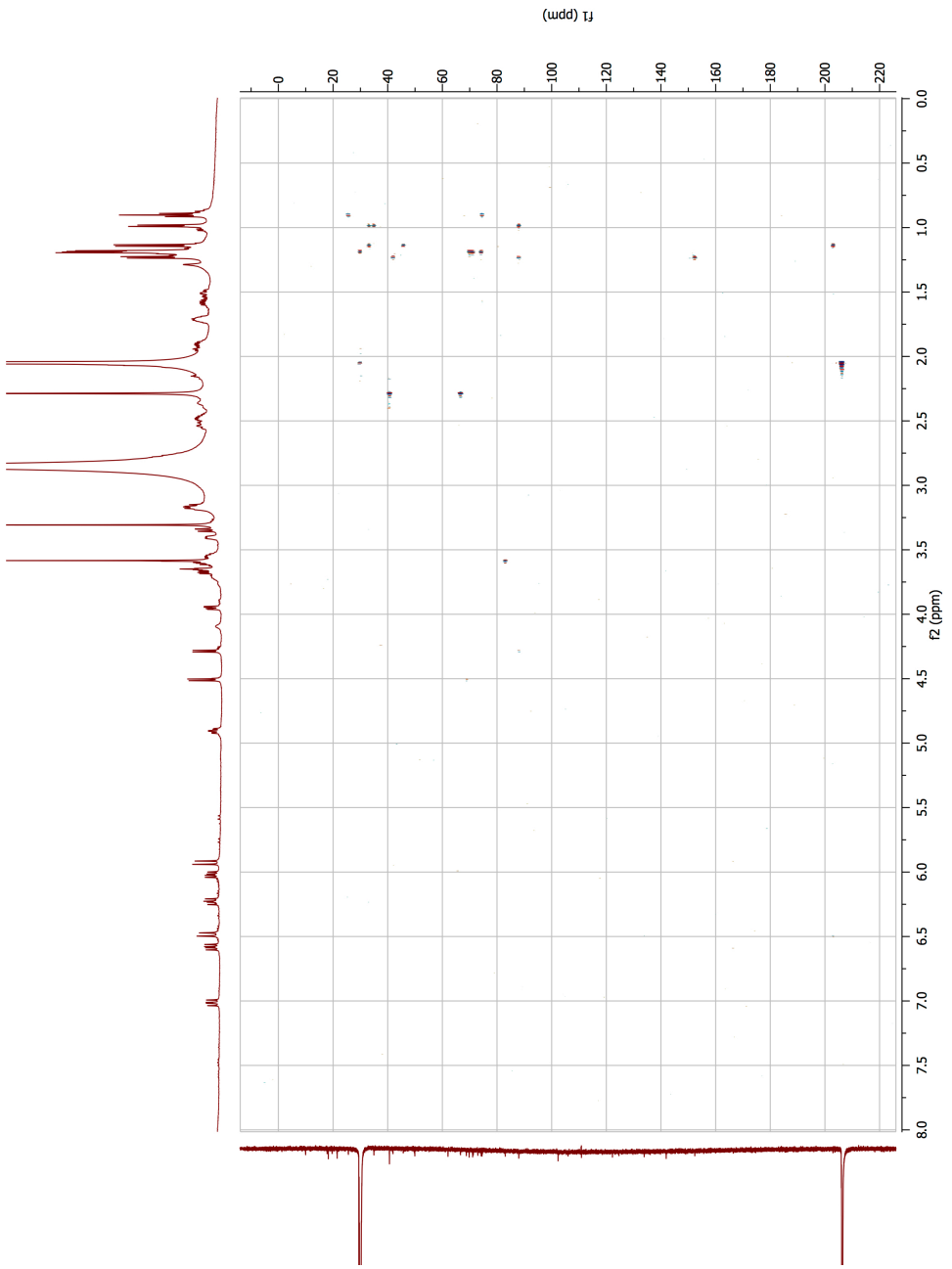


Figure 2-14 HMBCAD spectrum of new mycinamicin

Results

In order to understand MycF substrate specificity, we determined crystal structures (Table 2-2) for six different states of MycF, including the free enzyme, a binary complex with *S*-adenosylhomocysteine (SAH) and four ternary complexes with SAH and the substrate (mycinamicin III, **2**), the product (mycinamicin IV, **3**), an upstream pathway intermediate (mycinamicin VI; natural MycE substrate, **1**) and a non-natural substrate from the related tylosin pathway (macrocin, **5**). Crystals of wild type MycF diffracted with streaky patterns that were poorly reproducible but nevertheless yielded three structures of the free enzyme, binary SAH complex, and ternary substrate complex. The three crystal structures were in non-isomorphous space groups. Two crystal forms were related with identical crystallization conditions and crystal morphology and nearly identical unit cell constants. The streaky diffraction and multiple crystal forms were caused by a lattice contact that slipped to avoid repulsive contacts of glutamate side chains in neighboring molecules. We improved reproducibility and crystal quality by engineering substitutions at these positions. The double-substituted variant E35Q/E139A crystallized reproducibly, had non-streaky diffraction, and led to high-resolution structures of the ternary complex with product, upstream intermediate, non-natural substrate, and three structures of an active site variant. The double substitution had no impact on the catalytic activity of MycF (Table 2-11).

Table 2-11 Relative activity of MycF variants

MycF variant	% Activity with MycF substrate (mycinamicin III, **2**)
Wild type	100 ± 7
D191A	1.7 ± 0.1
D191N	1.9 ± 0.2
M56A	340 ± 21
F118Y	11 ± 3
L143A	199 ± 11
L143S	272 ± 6
L143N	219 ± 4
L143Q	28 ± 2
M56A/L143A	135 ± 5
M56A/L143S	68 ± 5
M56A/L143N	66 ± 4
E35Q/E139A	101 ± 1

Overall Structure.

MycF has a Rossmann-like fold common among small molecule methyltransferases (Figure 2-15) and, as expected, a nearly identical structure to NovP (r.m.s.d. of 0.414 Å for 203 Cα atoms, 54% overall sequence identity) (62, 71). Both in solution and in crystals, MycF is a dimer with an N-terminal helix (α1, residues 6-19) at the subunit interface. A metal ion is bound in the active site in all crystal structures. Two regions of the structure form an active site lid: an N-terminal "lid loop" (residues 25-51) between helices α1 and α2, and an α-helical "lid domain" (residues 116-144) between strand β1 and helix α4 (Figure 2-15b).

Substrate Binding Orders the Active site Lids.

The active site of MycF becomes progressively more ordered as substrates bind. In the free enzyme, both parts of the active site lid are disordered (Figure 2-15a,d). The binary and ternary complex structures include SAH, the desmethylated product of the SAM methyl donor, with its sulfur atom near the bound metal. SAM/SAH binding to MycF closes the helical lid domain over the adenosine and partially orders the lid loop. These contacts also occur in the SAH complex of NovP (62). Three amino acids (34-36) in the lid loop are disordered in the SAH complex (Figure 2-15b), whereas seven residues (corresponding to MycF 30-37) in the N-terminal lid loop are disordered in NovP and no metal ion is bound (62).

The ordered lid loop and helical lid domain form the active site entry channel and create a binding site for the substrate 16-membered ring macrocycle (Figure 2-15b,e). The entry channel is a funnel shaped cleft with magnesium and SAM at the bottom. The substrate binds with the javose sugar at the narrow bottom end of the funnel, which is 8 Å across. The macrolactone core rests in the tapered upper portion of the funnel, which is 10 x 20 Å at its widest and is formed by hydrophobic residues from both the ordered helical lid domain and lid loop including Leu32, Tyr49, Met132, Leu134, Tyr137 and Val141. Upon substrate binding the side chains of Leu32 and Tyr49 shift towards the substrate, and the final three residues of the N-terminal lid loop (34-36) become ordered (Figure 2-15c,f).

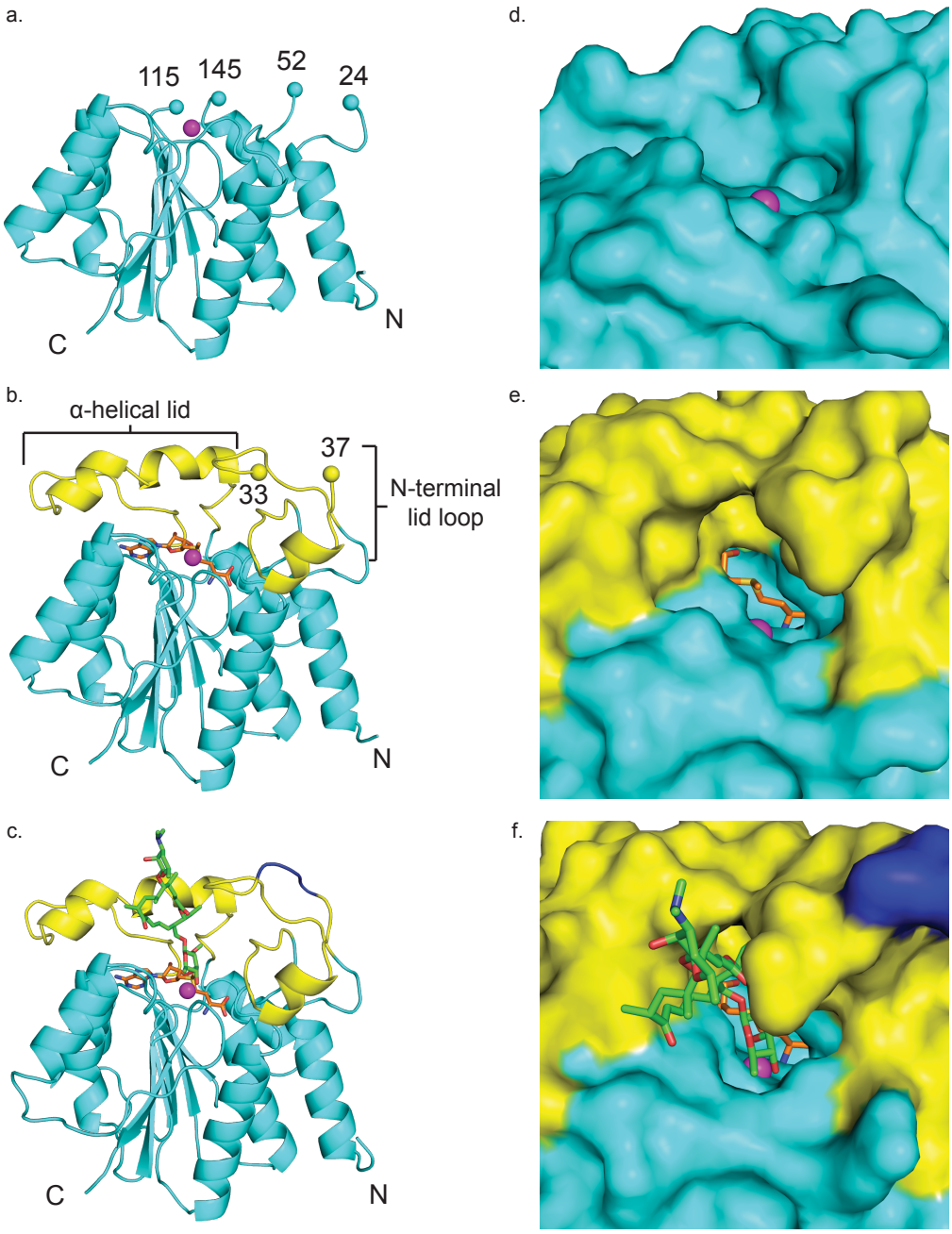

a.
115
145
52
24
C
N
b.
α-helical lid
37
33
N-terminal
lid loop
C
N
c.
C
N
d.
e.
f.

Figure 2-15 MycF substrate binding pocket. SAM binding orders the active site lids and creates a binding pocket for mycinamicin III (2). a) MycF-Mg^{2+}. In the absence of SAM the α-helical lid domain and N-terminal lid loop are unstructured, and the active site Mg^{2+} (magenta sphere) is solvent exposed. b) MycF-SAH. When SAM/SAH (orange sticks) is bound, the α-helical lid domain (yellow) closes over the cofactor. The N-terminal lid loop is ordered except for residues 34-36. c) MycF-SAH-substrate. Interactions with the substrate, mycinamicin III (2, green sticks), order the final three residues in the N-terminal lid loop (dark blue). Boundaries of disordered regions are highlighted with spheres and labeled at the final ordered residue. d-f) Surface representations of the structures in a-c, respectively, highlight the formation of the substrate binding cleft.

Active Site Organization.

The structures of MycF provide the first view of metal binding to a MycF/TylF family member. A metal ion is bound in the active site in all crystallized states of the enzyme even though no metal was included during the final purification step. The metal is presumed to be Mg^{2+} based on the octahedral coordination by oxygen ligands and the Mg^{2+} dependence of MycF/TylF family members (56-59). The Mg^{2+} binds in a negatively charged pocket where three Asp side chains (Asp189, Asp216 and Asp217) are monodentate metal ligands in all structures (Figure 2-16). In the SAH binary complex, Asp191 is also a Mg^{2+} ligand. The Mg^{2+} displays hexavalent coordination in all structures, as expected, with substrate hydroxyl groups or water occupying the other sites.

The Mg^{2+} ion orients the substrate for methylation at the 3′-hydroxy and stabilizes the presumed 3′-hydroxylate intermediate. The 3′- and 4′-hydroxyl groups of the mycinamicin III (**2**) javose sugar coordinate the Mg^{2+}, placing the 3′-hydroxyl in position for methylation, 4.7 Å from the sulfur atom of SAH or 3.1 Å from the modeled position of the SAM methyl group. The Gln246 side chain also stabilizes the substrate position by accepting a hydrogen bond from the 4′-hydroxyl. In the substrate complex, the javose 3′-hydroxyl displaces

Asp191 from the Mg^{2+} ligand field, and forms a hydrogen bond with the Asp carboxylate. Asp191 is the only basic residue near the site of methylation, hydrogen bonds with the 3′-hydroxyl group, and is conserved in the TylF/MycF family. Thus, it is the prime candidate for the catalytic base (Figure 2-16). Consistent with this hypothesis, asparagine and alanine substitutions at Asp191 impaired the methyltransfer reaction nearly 100 fold (Table 2-10). This is further supported by quantum mechanical calculations on the magnesium dependent catechol *O*-methyltransferase, which indicate that the Mg^{2+} positions the substrates for catalysis but is not required to lower the pKa of the substrate (97). In the related but structurally distinct TylE/MycE family, the analogous position is occupied by a histidine that is proposed to serve as the catalytic base (61).

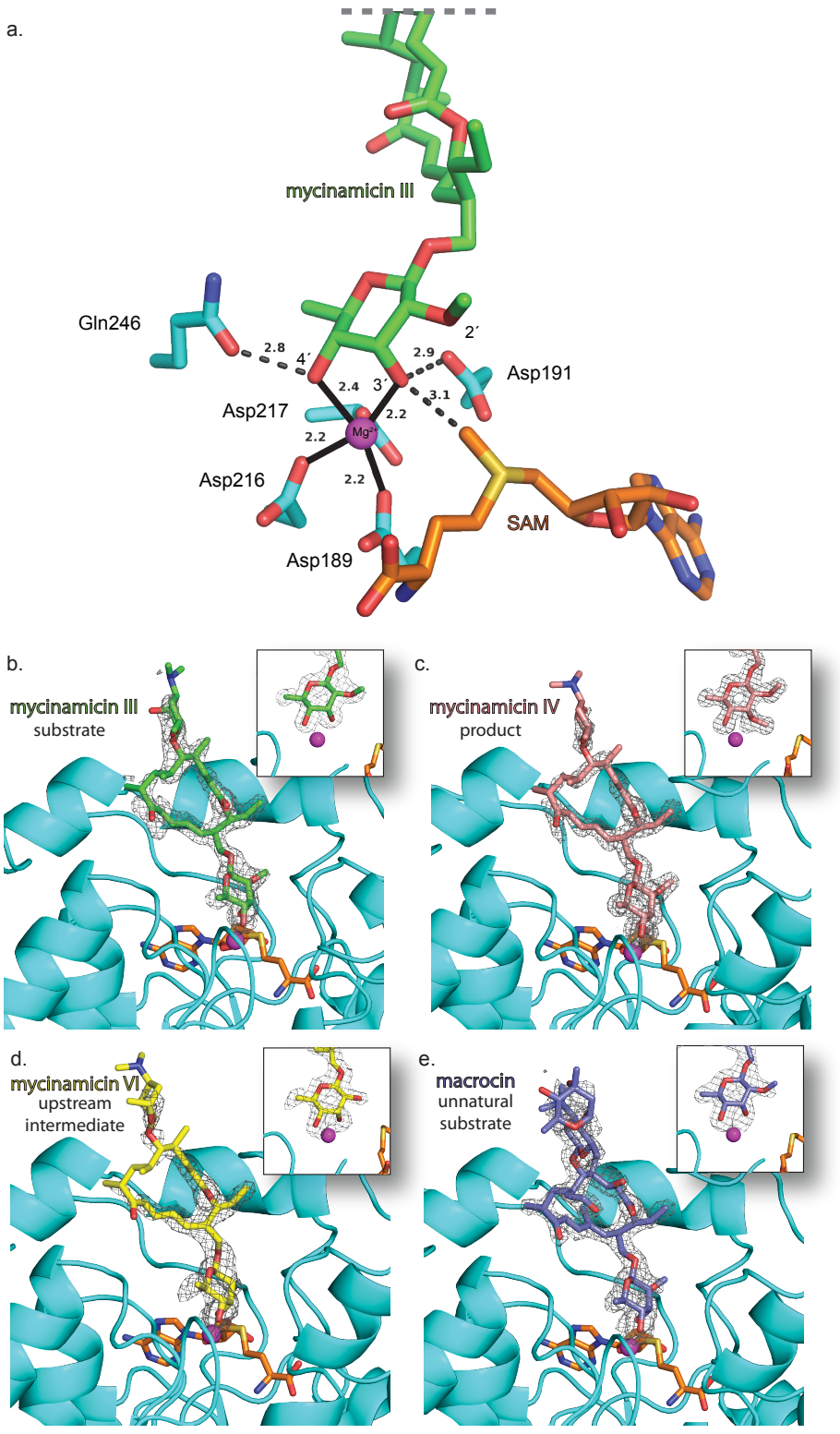

a.
mycinamicin III
Gln246
2.8
4´
2.9
2´
2.4
3´
3.1
Asp191
Asp217
2.2
2.2
Mg2+
Asp216
2.2
SAM
Asp189
b.
mycinamicin III
substrate
c.
mycinamicin IV
product
d.
mycinamicin VI
upstream
intermediate
e.
macrocin
unnatural
substrate

Figure 2-16 MycF active site. MycF is specific for the javose sugar and does not make specific contacts with the macrolactone ring or desosamine sugar. Mg^{2+} positions the substrate near Asp191 and the methyl donor. a) Substrate binding. The detailed view shows the three Asp residues that coordinate Mg^{2+}. Coordination bonds between Mg^{2+} and the 3′- and 4′-hydroxy groups of mycinamicin III (2) position the substrate for catalysis. The SAM methyl group (modeled) is 3.1 Å from the 3′-hydroxy. Asp191 is positioned to act as the catalytic base at 2.9 Å from the 3′-hydroxy. b-e) Ligand binding in the active site. The F_O-F_C omit density, contoured at 3σ, for b) the natural substrate (mycinamicin III, 2, green), c) the product (mycinamicin IV, 3, pink), d) the upstream pathway intermediate (mycinamicin VI, 1, yellow), and e) an unnatural substrate from the tylosin pathway (macrocin, 5, purple). The javose sugar (inset) and macrocycle can be built into density definitively for substrate (b), product (c) and the unnatural substrate (e). The upstream intermediate mycinamicin VI (1), which is not a substrate for MycF, is not as well ordered but con be confidently modeled in the same conformation as the substrate and product (d).

Comparison To Related Methyltransferases.

The MycF active site shares similarity to other metal dependent *O*-methyltransferases which include catechol *O*-methyltransferase (98) (1VID), alfalfa caffeoyl coenzyme A 3-*O*-methyltransferase (99) (1SUI, 1SUS), TDP-rhamnose 3-*O*-methyltransferase CalS11 (100) (3TOS, 4GF5) and mycinamicin 2′-*O*-methyltransferase, MycE (61) (3SSN). All structurally characterized metal dependent *O*-MTs share a conserved SAM binding site, metal binding site and catalytic base position (Lys in catechol *O*-MT and caffeoyl CoA *O*-MT, His in MycE, and Asp in NovP, CalS11 and MycF.) Superposition of the CalS11 and MycF (rms 3.5 Å for 258 residues) shows that both enzymes have lid regions inserted between helices α1 and α2, and between strand β1 and helix α4. In CalS11 helix α1 does not pack against the MT core, but interacts with neighboring subunits to form a decamer with C5 symmetry. Although the positions of these insertions are conserved, their topology and the difference in oligomeric states results in different active site access points relative to the

cosubstrate and metal binding site. The active site of CalS11 is significantly more open than MycF, potentially due to the lack of relevant substrates (*42)*. Additionally, its metal binding site may be an artifact of the crystallization conditions, as a bound K^+ ion interacts with only two of three conserved metal ligands.

Together the structures of the mycinamicin 2′-*O*-methyltransferase, MycE (PDB ID 3SSN) (61) and mycinamicin 3′-*O*-methyltransferase, MycF, provide the first structural view of ordered sugar methylation in macrolide biosynthesis. The positions of the metal, SAM and catalytic base are conserved on the core methyltransferase fold. The MycF and MycE active sites differ in the positions of their active site lids, resulting in significant differences in the substrate positions and conformations of the ternary complexes (Figure 2-17a). In MycE the substrate (**1**) 2′- and 3′-hydroxyls coordinate the magnesium and the sugar adopts a chair conformation with four axial substituents, placing the 2′-hydroxy proximal to the methyl group of SAM. In contrast, MycF positions the substrate such that the sugar has a chair conformation with only one axial substituent, the 3′- and 4′-hydroxyl groups coordinate the magnesium, and the 3′-hydroxyl is nearest to SAM (Figure 2-17b). Although the MycE substrate conformation appears less favorable (four axial substituents *vs*. one), the calculated free energy difference between these two substrate poses is small, 24 kJ/mol (5.7 kcal/mol). This can readily be rationalized by the magnesium coordination bonds and hydrophobic interactions of the macrolactone core, which are more extensive with MycE than with MycF (61).

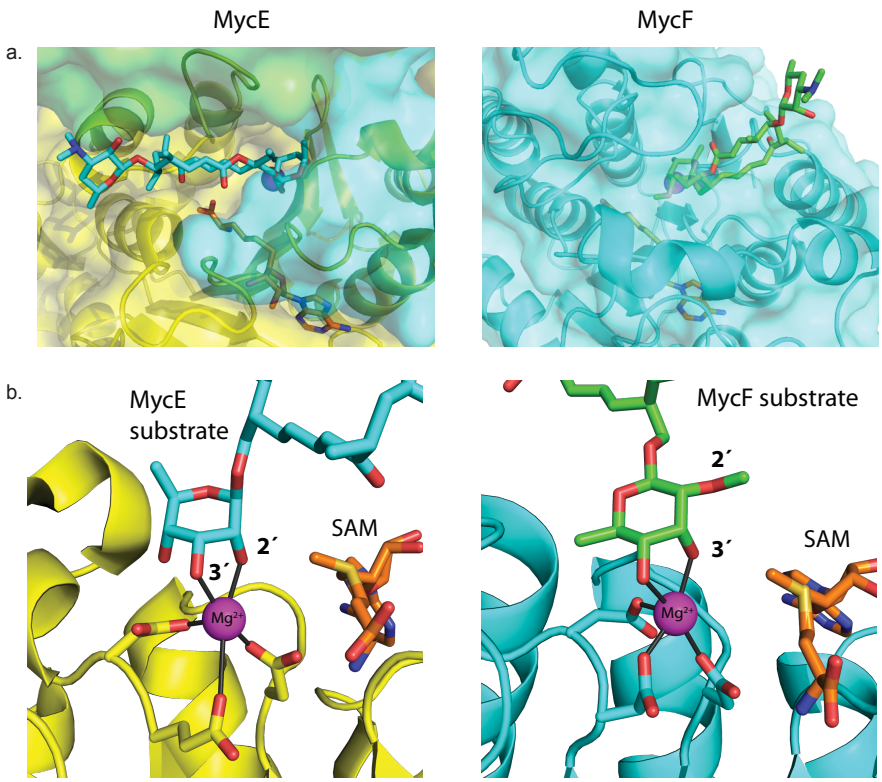


Figure 2-17 Comparison of MycE and MycF active sites. a) Surface representation. The active sites of MycE (PDB ID 3SSN, left; yellow, green, cyan) and MycF (right; cyan) are shown with SAM in orange, Mg^{2+} in magenta, mycinamicin VI (1) in cyan and mycinamicin III (2) in green. The active site lids of MycE and MycF direct entry on opposite sides of the catalytic Mg^{2+} and SAM, which are shown in identical orientation for the two enzymes. b) Detailed view of the MycE (yellow) and MycF (cyan) active sites. The different orientation of substrates results in differences in the magnesium ligands, sugar conformation and regiochemistry of methyltransfer. Coloring is the same as in (a).

Substrate Specificity.

The MycF ternary complex with SAH and the natural substrate mycinamicin III (**2**) provides insight into the substrate specificity of MycF/TylF family members and suggests that MycF may be active with alternative substrates. In the substrate ternary complex, the javose sugar and macrolactone ring are clearly defined in the electron density whereas the desosamine sugar, attached at C11, is less well ordered, makes no protein contacts and thus

appears not to contribute to substrate specificity (Figure 2-16b). The macrolactone is located in a hydrophobic region at the opening of the active site funnel where it is partially solvent-exposed and forms no specific hydrogen bonding contacts with MycF (Figure 2-18). This suggests that MycF might accept javose substrates bearing different core macrolactones, differentially substituted macrolactones, or non-macrolactone hydrophobic substituents.

To assess the flexibility of MycF to accept alternative substrates we examined the activity of MycF with macrocin (**5**), the javose-containing intermediate from the tylosin biosynthetic pathway. Both mycinamicin III (**2**) and macrocin (**5**) are 16-membered ring macrolides bearing javose and desosamine sugars. Macrocin has an additional mycarose sugar linked to the desosamine moiety (Figure 2-1b). Under our assay conditions, MycF methylated the 3′ hydroxy of both macrocin (**5**) and the native substrate, mycinamicin III (**2**), with similar efficiency (Figure 2-19). As MycF and TylF have 70% amino acid sequence identity and similar substrates, it is not surprising that they catalyze methytransfer with each other's substrates. These results confirm that MycF does not interact with the desosamine portion of the substrate and tolerates changes to the macrolactone core. In the SAH-macrocin (**5**) ternary complex, the non-natural substrate is bound in the same position as the natural substrate and the differences in the macrolactone structure do not alter the position of javose in the active site (Figure 2-16e).

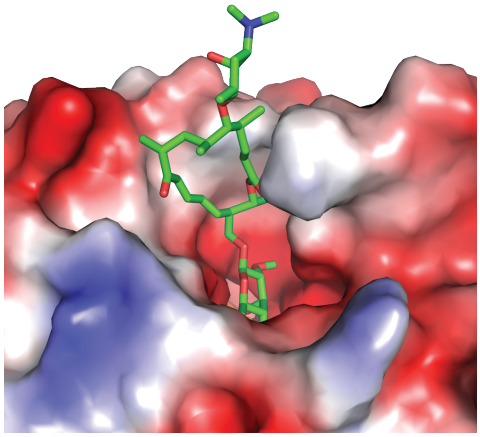


Figure 2-18 MycF active site electrostatic surface potential. MycF forms no specific contacts with the macrolide core. The electrostatic surface potential of the MycF active site is shown with the substrate (mycinamicin III, 2) in green sticks. The substrate macrolactone interacts with a hydrophobic surface suggesting that MycF can accept a variety of hydrophobic substrates with the javose sugar.

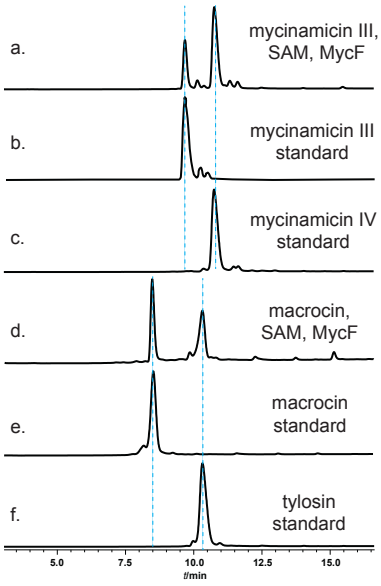


Figure 2-19 HPLC analysis of MycF reaction mixtures. MycF is active on the unnatural substrate macrocin. a) HPLC chromatogram of mycinamicin III (2) incubated with MycF and SAM shows conversion of mycinamicin III (2) to mycinamicin IV (3), confirmed with standards in (b) and (c). d) MycF catalyzed methylation of macrocin (5). Standards of macrocin (5) and tylosin are shown in (e) and (f). Experimental conditions are described in Methods.

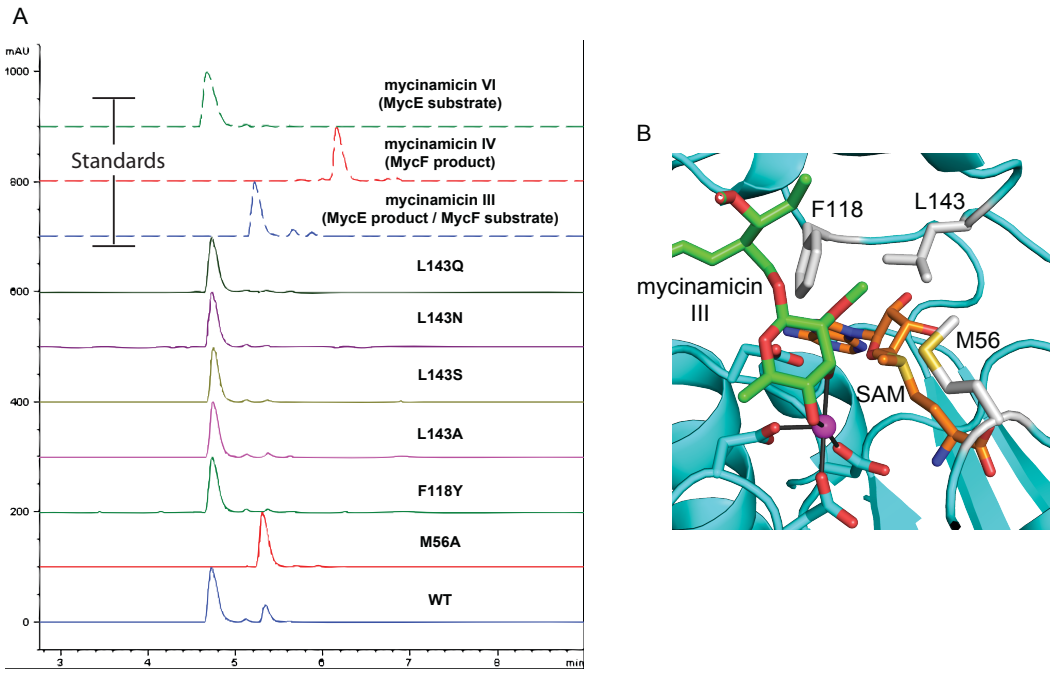


Figure 2-20 MycF activity with MycE substrate (mycinamicin VI). A) HPLC chromatograms of MycF variants with the MycE substrate (mycinamicin VI). Standards of mycinamicin VI, III, and IV are shown. WT MycF and MycF M56A show production of a new product that has a retention time similar to mycinamicin III. B) View of the MycF active site highlighting the 2′-methoxy specificity pocket. Residues selected for substitution experiments are shown in grey sticks.

Pathway Ordering.

Sugar methylation occurs in a specific order in the biosynthesis of mycinamicin, tylosin and spinosyn (56-58, 60). Members of the MycF family methylate the javose 3'-hydroxy only after the 2'-hydroxy of 6'-deoxyallose has been methylated (creating javose) under standard assay conditions. A key motivation for our study was to determine how MycF distinguishes between the substrate mycinamicin III (**2**) and mycinamicin VI (**1**), which is poor substrate, as these molecules differ only in the 2' substituent (methoxy *vs*. hydroxy). It

was not immediately apparent from the structure of the MycF-SAH-substrate ternary complex why the MycE substrate mycinamicin VI (**1**) is not methylated with a similar efficiency to the natural substrate. Indeed, with long incubation times and high enzyme concentration MycF converted mycinamicin VI to a previously unreported product (Figure 2-20a). The javose 2'-methoxy binds in a hydrophobic pocket created by Met56, Phe118, Leu143 and Val141 (Figure 2-20b). We first considered that this pocket might be too hydrophobic for a 2'-hydroxyl group, and substituted alanine or a polar side chain at each of three hydrophobic amino acids in the pocket (M56A, F118Y, L143A, L143S, L143N, L143Q). Two substitutions decreased the hydrophobic surface area (M56A, L134A) and four increased the polarity of the hydrophobic pocket (F118Y, L143S, L143N, L143Q). The M56A substitution showed increased conversion of mycinamicin VI to a single new product relative to WT MycF under extended incubation conditions (Figure 2-20a). None of the variants had detectable activity with mycinamicin VI (**1**) under standard assay conditions although, interestingly, four of the substitutions increased activity with the natural substrate mycinamicin III (**2**) (Table 2-11). We made double-substituted MycF variants to see if the substitutions had additive or synergistic effects, however the double substitutions had activity comparable to wild type with the natural substrate and no activity with the MycE substrate (Table 2-11). The previously uncharacterized compound had a slightly different retention time than mycinamicin III under our chromatography conditions and mass spectrometry showed a 16 Da shift relative to the starting material consistent with a single methylation (Figure 2-21). Using multidimensional NMR experiments we confirmed that the product is a new mycinamicin analog methylated at the 3'-hydroxyl (Table 2-10, Figures 2-22 and 2-23).