Overcoming the challenges associated with MbtH-like protein dependence of nonribosomal peptide synthetases.

By Karla J. Esquilín-Lebrón

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Microbiology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: 12/10/2018

The dissertation is approved by the following members of the Final Oral Committee:

Michael G. Thomas, Professor, Bacteriology
Silvia Cavagnero, Professor, Chemistry
Garret Suen, Professor, Bacteriology
Daniel Amador Noguez, Assistant Professor, Bacteriology
Brian Pfleger, Professor, Chemical and Biological Engineering

An important group of bioactive natural products produced by bacteria and fungi are the nonribosomal peptides. The biosynthesis of these structurally diverse metabolites is derived by the modular enzymology of nonribosomal peptide synthetases (NRPSs). The modularity of NRPSs reside in a set of repeating catalytic domains that work together to recognize and incorporate aryl or amino acid precursors into the final product. The repetitive nature of NRPSs makes them perfect candidates for the rational design of these enzymes by the rearrangement of modules or domains to modify the nonribosomal peptide chemical scaffolds. Successes of NRPS combinatorial biosynthesis approaches have been scarce. Recent findings by the Thomas laboratory and others revealed that many NRPSs in bacteria require members of the MbtH-like proteins (MLP) superfamily for their solubility or function. MLP/NRPS interactions and their complexity are an important factor that the field has not addressed during combinatorial biosynthesis efforts. Proper pairing of MLP/NRPS systems might be the key to successful combinatorial biosynthesis efforts. In this thesis I propose two approaches to overcome the MLP challenge in combinatorial biosynthesis. First, the identification or evolution of a “universal” MLP that can act as a molecular tool and interact with multiple NRPSs. In Chapter 2, I characterized an orphan MLP, MXAN_3118, from *M. xanthus* DK1622 and identify it can naturally interact with at least seven cognate NRPS systems and the enterobactin system from *E. coli*. The ability of MXAN_3118 to naturally interact with multiple NRPSs suggests this enzyme is a candidate universal MLP that can potentially interact with MLP-dependent NRPSs from different systems.

My second approach is to understand how we can change an MLP-dependent NRPS to be independent of the MLP. In Chapter 3, I studied the recently discovered enterobactin (ENT) system from yeast and identified that after acquisition of the biosynthetic gene cluster and further evolution the yeast system is MLP-independent. I hypothesize that the yeast ENT system lost the requirement for MLP from a combination of factors; including changes in the protein structure and the yeast cell physiology. The work in this thesis to overcome the MLP-dependence of NRPS

systems, or MLP challenge, provided groundwork that can contribute to the future success of natural product engineering.

ACKNOWLEDGEMENTS

Words cannot describe how grateful I am to have been part of the Thomas lab, the Microbiology doctoral training program and the University of Wisconsin-Madison; but this journey would not have been possible without the great support and love of my family, friends and mentors.

Thank you, Michael! For being a great mentor, your support, guidance, encouragement (even when I did not believe in myself) and for pushing me every day to be a better scientist. I will miss our conversations in the office and classes on the white board while I sat down on the floor taking notes.

Thanks to all of the Thomas lab members past and present; Becca, Vlad, Hyunjun, Sophia, Kurt & Kai. I am very grateful I got to have a great lab family and that I got to share this journey with the best team I could have ask for, Becca & Vlad. Your support, help with experiments, endless laughs, mental breaks for a diva coffee and group hugs were key to those loooong days at lab.

My family in Puerto Rico, it was not easy being away from home but your love, encouragement and daily support motivated me to never give up and continue every day to accomplish the goal! To my brother Giancarlo, thank you for showing me that with perseverance and hard work we can accomplish any goal we set ourselves for. I am so proud of you and I cannot wait to celebrate the great things to come.

To my godmother Maris Fonseca thanks for being a role model and for encouraging me to apply to UW-Madison, On Wisconsin!

My forever labmate Hildamarie, thank you for your loyal and supportive friendship during all of these years. From our undergrad adventures to your visits to Madison and long talks to support each other in this crazy journey called graduate school! Thanks for always being there when I needed you.

My family in Madison (Zulmarie, Pepe, Jehziel, Elian, Jara, Jarel, Nacho, Lynn, Ali, Michael and Rocco) and my family in Chicago (Dianicha & Hector) GRACIAS! For opening the doors of your houses and family to me, your support, food, and the love was always there!

A special thanks to my “combo” of boricuas and friends that Madison gave me; Lory, Diana, Emmanuel, Patricia, Sofia, Cristina, Marti, Kemardo, Dianiris, Natalia, Mayra, Matt, Jeysika, Paulina, Camilo, Pedro, Will, Ben, Maria. Thanks for the dinners, gatherings, writing clubs and any other excuse to hang-out and support each other was key to keep me focused, supported and loved away from home.

My UW-Madison family; SciMed, Abbey and Sara, the community provided me with a supportive environment and a family on campus that help me feel that I could accomplish this goal.

I want to extend my gratitude to our collaborators Dr. Larry Shimkets and Dr. Tye Boynton from University of Georgia-Athens. Also, to the Forest, Gourse and Hittinger labs that provided help and supplies for some of my experiments.

This thesis is dedicated to the two most important women in my life; my mom Jeanette and my grandma Monsi.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1: Introduction

Chapter 2: An orphan MbtH-like protein interacts with multiple nonribosomal peptide synthetases in *Myxococcus xanthus* DK1622

Chapter3: Characterization of an enterobactin-producing nonribosomal peptide synthetase from yeast

Chapter 4: Conclusions and Future Directions

LIST OF TABLES

Chapter 1: Introduction

Chapter 2: An orphan MbtH-like protein interacts with multiple nonribosomal peptide synthetases in *Myxococcus xanthus* DK1622

Chapter3: Characterization of an enterobactin-producing nonribosomal peptide synthetase from yeast

Chapter 1

Introduction.

An important group of natural products characterized by their diverse chemical structures and broad range of biological activities are the nonribosomal peptides. Due to their complex chemical scaffolds the chemical synthesis of nonribosomal peptides are a significant challenge and current efforts are focused on genetic engineering (or combinatorial biosynthesis) approaches to rationally design new nonribosomal peptide variants (1). Combinatorial biosynthesis of these natural product biosynthesis pathways often has resulted in unsatisfactory results. To optimize efforts to derive new compounds we first need detailed mechanistic insight into the assembly enzymology of these bioactive molecules. The biosynthesis of structurally diverse nonribosomal peptides, and their potential as targets for combinatorial biosynthesis, is driven by the shared modular enzymology of nonribosomal peptide synthetases (NRPSs) that assemble these natural products (2). The challenge in understanding how to rationally reprogram these enzymes was increased by the recent finding that there are two types of NRPS systems: those that are dependent on an accessory protein belonging to the MbtH-like protein (MLP) superfamily, and those that are MLP independent. The complexity of the NRPS/MLP interactions introduces new challenges for metabolic engineering of MLP-dependent NRPSs. This thesis investigates two mechanisms for overcoming these challenges: 1) the identification of a “universal” MLP that can functionally interact with any targeted NRPS, and 2) understanding how to convert an MLP-dependent NRPS into one that is MLP independent.

NRPS enzymology.

The repetitive enzymatic nature of NRPS makes them excellent candidates for rational combinatorial biosynthesis to generate functional, hybrid megasynthases of designer “unnatural” natural products. The genes coding for all the enzymes needed to assemble a nonribosomal peptide are usually found within biosynthetic gene clusters (BGCs). Natural products are biosynthesized in an assembly line fashion by the repetitive enzymology of NRPSs (3). NRPS

assembly lines are divided into modules, with an NRPS module consisting of the catalytic domains that are responsible for the incorporation of a specific precursor into the final product. These modules can include an adenylation (A) domain, thiolation (T) domain, and a condensation (C) domain (2) (Fig. 1). The A domains are known as the “gate-keepers” of the assembly line since their role is to recognize and activate in an ATP-dependent manner a specific precursor to generate an acyl-AMP intermediate. In contrast to the ribosome, NRPS A domain substrate specificity is not limited to the 20 proteinogenic amino acids as building blocks. A domains from the range of known NRPSs activate and incorporate more than 500 different substrates that range from nonproteinogenic amino acids, fatty acids and alpha-hydroxy acids (4). The diverse selection of building blocks gives these class of natural products ample structural diversity and biological activity for the synthesis of nonribosomal peptides. The next step in the assembly line-like enzymology is catalyzed by the T domain where the thiol of the 4'-phosphopantotetheinyl group attacks the acyl-AMP intermediate, releasing AMP, and forming an acyl-*S*-T intermediate (Fig.1). Subsequently, the C domain from a downstream module will catalyze a peptide bond between acyl-S-T intermediates tethered to upstream and downstream modules. Following peptide bond formation, the intermediate can undergo additional condensation reactions or the nonribosomal peptide can be released by the thioesterase (TE) domain at the C-terminus of the NRPS. The three core catalytic domains are not the only domains in a NRPS module; additional tailoring domains can be found in NRPS assembly lines. These domains can be responsible for further substrate modifications, for example an epimerase domain can change the stereochemistry of an amino acid substrate contributing to structure complexity (3).

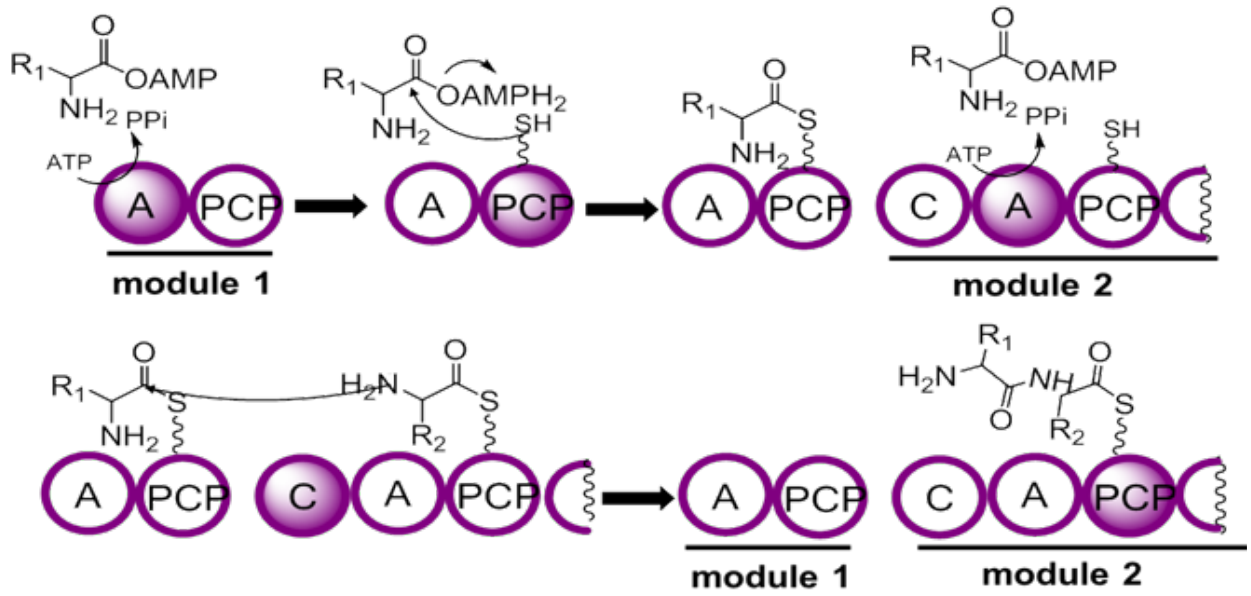


Figure 1. Schematic representation of the core catalytic reactions of an NRPS. The core catalytic domains are adenylation (A), peptidyl carrier protein (PCP) and condensation domain. The 4'-Ppant cofactor is represented by the SH bond to the PCP domain.

The catalytic domains described above were thought to be the core of NRPS enzymology, but recently our laboratory (5) and others (6–8) found that there are two classes of NRPSs: those that have A domains that function independent of any additional factor, and those that have A domains that require the presence of an accessory protein belonging to the MbtH-like protein (MLP) superfamily (5–7). Since nonribosomal peptides compose an important group of drugs currently used in the clinics and there is a desire to generate designer versions of these natural products, it is imperative we understand how to deal with this dichotomy of NRPSs.

Efforts to understand the role(s) of MLPs in NRPS enzymology.

MLP-dependent A domains are found in NRPSs involved in the production of an array of medically-relevant natural products like antibiotics, siderophores and lipopeptides (9). MLPs are small proteins (~70 amino acids) associated with NRPS-encoding gene clusters in bacteria (5). The first MLP to be identified was YbdZ from the siderophore enterobactin (ENT) biosynthetic

gene cluster (BGC), yet the protein superfamily was named after MbtH, the protein coded by the eight gene in the mycobactin biosynthetic gene cluster (10, 11). MLPs were initially overlooked in NRPS enzymology studies because not all of the NRPS-encoding gene clusters code for one and initial studies suggested that there was no role for MLPs in NRPS biosynthesis. One of these early studies determined that *orf1* (*ybdZ)* the MLP-encoding gene from enterobactin biosynthesis in *E. coli* was not produced in the organism (10). Reichert *et al.* (1992) used ^{35}S-Methionine to label the only methionine (Met) in YbdZ overlooking the fact that *E. coli* often cleaves the first Met from its proteins. Interestingly in this study EntF was observed to be more stable when co-expressed with *orf1*. Also, successful *in vitro* reconstitution studies of ENT were performed in the absence of exogenously added YbdZ to the reaction supporting the hypothesis that is not required. However, in this assay all of the proteins were overproduced in and purified from an *E. coli* overexpression strain that expressed a chromosomal copy of *ybdZ* (12). Studies by Stegmann *et al*. (2006) also missed the role of the MLP because they deleted the MLP encoded within the balhimycin BGC, did not observed a change in metabolite production, and assumed the MLP is not essential for balhimycin biosynthesis. However, future studies showed that a second MLP-encoding gene was present in the genome of the organism and functionally replaced the natural MLP partner for the balhimycin NRPS (13).

Overlooked work on the siderophore pyoverdine from *Pseudomonas* first demonstrated the essential nature of MLPs by showing that the deletion of the MLP-encoding gene PA2412 abolished pyoverdine production and the bacterium's survival under iron limited conditions (14). Subsequent work on MLPs encoded in actinomycete genomes showed that MLPs are essential for the production of clorobiocin, CDA and coelichelin (15–17). These studies required the researchers to inactivate all MLP-encoding genes to abolish metabolite production and revealed that if a producing organism had more than one MLP-encoding gene the deletion of only one may not be sufficient to abolish metabolite production (15, 16). MLPs from other NRPS encoding gene

clusters or noncognate MLPs can have a level of interactions with noncognate NRPS systems, or crosstalk, and influence their metabolite production (8, 16, 18). These findings led the field to question the role of MLPs and determine if previously identified NRPSs were also MLP-dependent.

Biochemical studies of MLPs.

Here, I summarize biochemical and structural findings from the past 10 years that have been focused on understanding the role of MLPs and their interactions with NRPSs. Biochemical characterization of NRPSs in the presence and absence of MLPs establish that MLPs can co-elute with NRPSs, can influence the solubility of a heterologously produced NRPS, can influence affinity (K_m) for the amino acid substrate and influence the aminoacylation reaction for a functional NRPS (5, 6, 19, 20). Since MLPs can have multiple impacts on NRPS enzymology the use of only one assay is insufficient to determine if an NRPS is MLP-dependent (20). Another level of complexity of these interactions is that not all of the A domains in a particular NRPS require an MLP and A domains in the same NRPS can have different levels of interactions with an MLP partner (21, 22). These findings suggested that MLPs act as chaperone-like proteins to ensure the NRPS has the correct conformation for efficient catalysis (5–7).

The understanding that MLPs are essential for the solubility and function of their cognate NRPS partner provided a means for characterizing NRPSs that have proven recalcitrant to biochemical characterization (7, 8, 21–24). These findings provided evidence of the essentiality of MLPs for combinatorial biosynthesis, but also prompted the question of how heterologous hosts with cognate MLPs will influence the NRPSs from other systems. Recent findings from our lab determined that pairing non-cognate MLPs with the MLP-dependent NRPS from ENT biosynthesis can have a variety of impacts on NRPS function (18). These findings established

that noncognate MLP/NRPS interactions are complex and a functional pairing cannot be determined simply by sequence similarity to the cognate MLP or only by assessing NRPS solubility or substrate activity. It also was shown that an inappropriate MLP/NRPS pair can have a negative impact on the loading of the acylated substrate into the PCP domain. A structure-function analysis of MLP/NRPS interactions using an alanine scan of the ENT MLP, YbdZ suggested that topological interactions between the MLP and NRPS are needed for proper pairing and functionality. From the eleven residues identified to be required for functional MLP/NRPS interactions it was identified that the N-terminal region of YbdZ is important for protein-protein interactions (co-purification and affinity) with the NRPS and that the residues in the MLP-A domain interface as well as those in the C-terminal region of YbdZ are critical for the MLP stability. Interestingly, the NRPS EntF can influence the solubility of some of the YbdZ variants; this is the first time a NRPS protein is identified to influence an MLP. In summary, the variability of MLP/NRPS interactions makes them a complex task to disentangle.

Structural studies of MLPs and MLP bound NRPSs.

Structural studies of MLPs, which are small proteins of approximately 70 amino acids, and their association with NRPSs have contributed to our understanding of these MLP-NRPS interactions. The first structures of MLPs revealed a unique fold composed of three antiparallel beta-sheets and one to two alpha helices with three conserved tryptophan residues located on the same face of the protein with no catalytic core (17, 25). The first structure that looked at an MLP bound to an NRPS was solved for Slgn1, an NRPS with the MLP fused to the N-terminus of the A domain. The Slgn1 structure showed that the MLP binds away from the catalytic core of the A domain (26). In the MLP-A domain interface, two of the conserved tryptophan residues from the MLP form a pocket into which an A domain alanine residue fits. Mutation of the alanine residue to a glutamate abolished the activity of the A domain. This observation was confirmed in the structure of the

NRPS protein EntF bound to either the cognate MLP, YbdZ or a noncognate MLP, PA2412. This led to the hypothesis that MLP-dependent A domains have a non-polar residue in the alanine position that enables interaction with the MLP (27). Although these early studies did not reveal a role for MLPs in NRPS enzymology, they did highlight the MLP regions that interact with the A domain, which were further explored through bioinformatics studies(28).

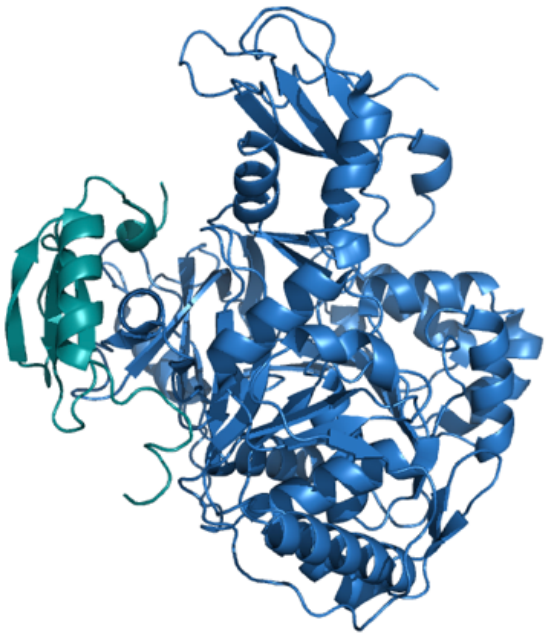


Figure 2. Cartoon representation of the structure of YbdZ (teal) bound structure of the adenylation domain (blue) from the EntF module. (PDB 5JA1)

The structure of EntF in complex with YbdZ and also without YbdZ revealed that upon binding of the MLP there were no significant conformational changes (27). Subsequent studies of an MLP bound to the DhbF NRPS provided evidence that MLP binding results in structural changes of the NRPS that are not detected by crystallographic methods but are observable under native gel conditions (29). The crystallographic conditions used in both structures, EntF/YbdZ and DhbF/MLP, used the Ser-adenosine vinylsulfonamide substrate analog that trapped the A domain in a thioester-forming conformation with the T domain, corresponding to a snapshot of A domain activity (27, 30). The lack of a detectable conformational change may be due to the crystallization

conditions and additional studies using cryogenic electron microscopy may provide a more accurate picture of how MLPs influence NRPS catalysis.

The MLP challenge.

The variable interactions of MLPs with NRPS systems adds another complication to combinatorial biosynthesis efforts. One of the goals in combinatorial biosynthesis is to rationally engineer NRPS assembly lines by rearrange modules, domains or subdomains form different systems and organisms to obtain a novel structure (1). The MLP challenge has not been taken into account previously in combinatorial biosynthesis efforts. There are three major aspects of combinatorial biosynthesis in which MLPs may influence success.

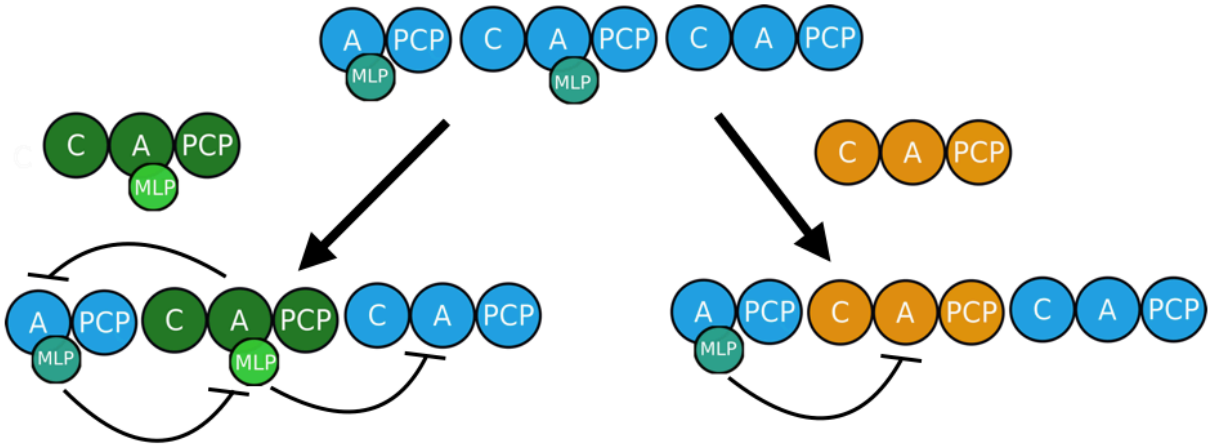


Figure 3. Diagram representing two examples of combinatorial biosynthesis incorporating an MLP-dependent (green) and MLP-independent (orange) NRPS modules. Possible negative interactions between MLPs and NRPSs are highlighted. Proper pairing of MLP/NRPS systems might be a key to successful combinatorial biosynthesis efforts.

First, during the engineering of a hybrid NRPS an MLP-dependent module can be incorporated and the resulting non-cognate MLP/NRPS pairings can inhibit NRPS function. As represented in Figure 3, the non-cognate MLP (green) can inhibit interactions between the cognate MLP/NRPS pairs (blue) and the cognate MLP can also inhibit the non-cognate NRPS (green). Second, the

non-cognate MLP may also inhibit a cognate MLP-independent NRPSs. Third, if an MLP-independent NRPS module or A domain is incorporated into an MLP-dependent NRPS system the cognate MLP (blue) may inhibit the MLP-independent noncognate NRPS. We can expect that if we do not have a correct pairing of MLP-NRPS interactions we will not have a functional hybrid NRPS system. In the following Chapters, I propose two approaches to overcome the MLP challenge: 1) the identification or evolution of a “universal” MLP that can interact with multiple NRPSs; similar to the role of the phosphopantotheinyl transferase Sfp widely used in NRPS and PKS enzymology (Chapter 2) and 2) the understanding of how we can change an MLP-dependent NRPS to be independent of the MLP (Chapter 3).

Chapter 2

An orphan MbtH-like protein interacts with multiple nonribosomal peptide synthetases in *Myxococcus xanthus* DK1622.

Karla J. Esquilín-Lebrón[1], Tye O. Boynton[2], Lawrence J. Shimkets[2] and Michael G. Thomas[1]*.
[1]Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin. [2]Department of Microbiology, University of Georgia, Athens, Georgia.

*****Modified from: Esquilín-Lebrón KJ, Boynton TO, Shimkets LJ, Thomas MG. 2018. An orphan MbtH-like protein interacts with multiple nonribosomal peptide synthetases in *Myxococcus xanthus* DK1622. *J Bacteriol* 200:e00346-18. https://doi.org/10.1128/JB.00346-18.

The co-authors contributed to the research in the following ways:

Prof. Michael G. Thomas performed initial *in silico* analysis of orphan MLPs.

Karla J. Esquilín-Lebrón performed cloning, heterologous expression, protein purification and characterized all NRPS modules from MXAN_3634-6 and other MXAN proteins, performed RT-PCR analysis, developed the B2H system and phenotypic assays for *M. xanthus* strains, and was involved in the *in silico* analysis of orphan MLPs.

Dr. Tye O. Boynton, under the supervision of Dr. Lawrence J. Shimkets, created the in-frame deletions of MXAN_3118 and MXAN_3634-6 from *Myxococcus xanthus* DK1622.

Abstract

One mechanism by which bacteria and fungi produce bioactive natural products is through the use of nonribosomal peptide synthetases (NRPSs). Many NRPSs in bacteria require members of the MbtH-like protein (MLP) superfamily for their solubility or function. Although MLPs are known to interact with adenylation domains of NRPSs, the role MLPs play in NRPS enzymology has yet to be elucidated. MLPs are nearly always encoded within the biosynthetic gene clusters (BGCs) that also code for the NRPSs that interacts with the MLP. Here, we identify 50 orphan MLPs from diverse bacteria. An orphan MLP is one that is encoded by a gene that is not directly adjacent to genes predicted to be involved in nonribosomal peptide biosynthesis. We targeted the orphan MLP MXAN_3118 from *Myxococcus xanthus* DK1622 for characterization. The *M. xanthus* DK1622 genome contains fifteen NRPS-encoding BGCs, but only one MLP-encoding gene (MXAN_3118). We tested the hypothesis that MXAN_3118 interacts with one or more NRPS using a combination of *in vivo* and *in vitro* assays. We determined that MXAN_3118 interacts with at least seven NRPS from distinct BGCs. We show that one of these BGCs codes for NRPS enzymology that likely produces a valine-rich natural product that inhibits clumping of *M. xanthus* DK1622 in liquid culture. MXAN_3118 is the first MLP to be identified that naturally interacts with multiple NRPS systems in a single organism. The finding of an MLP that naturally interacts with multiple NRPS systems suggests it may be harnessed as a “universal” MLP for generating functional hybrid NRPSs.

Importance

MLPs are essential accessory proteins for the function of many NRPSs. We identified 50 MLPs from diverse bacteria that are coded by genes that are not located near any NRPS-encoding BGCs. We define these as orphan MLPs due to their NRPS partner(s) being unknown. Investigations into the orphan MLP from *M. xanthus* DK1622 determined that it interacts with NRPSs from at least seven distinct BGCs. Support for these MLP-NRPS interactions came from the use of a bacterial two-hybrid assay and co-purification of the MLP with various NRPSs. The flexibility of this MLP to naturally interact with multiple NRPSs lead us to hypothesize that this MLP may be used as a “universal” MLP during the construction of functional hybrid NRPSs.

Introduction

An important group of bioactive natural products produced by bacteria and fungi are the nonribosomal peptides. The biosynthesis of these structurally diverse metabolites is derived by the modular enzymology of nonribosomal peptide synthetases (NRPSs). The modularity of NRPSs resides in a set of repeating catalytic domains that work together to recognize and incorporate aryl or amino acid precursors into the final nonribosomal peptide product. Typically a module includes an adenylation (A) domain, a thiolation (T) domain, and a condensation (C) domain cooperating in an assembly-line fashion to form a nonribosomal peptide (2). A domains are essential components of NRPSs as “gate-keepers” of the assembly line by selecting the precursor introduced by each module (2). Recently, it was recognized that there are two types of A domains in bacteria. Members of the first type function independent of any additional protein factor. Members of the second class require an accessory protein from the MbtH-like protein (MLP) superfamily to be functional (5–7, 19).

The MLP superfamily is named after MbtH, a protein encoded by the mycobactin biosynthetic gene cluster (BGC) in *Mycobacterium tuberculosis* (11). MLPs are small proteins of approximately 70 amino acids that are nearly always encoded within NRPS-encoding BGCs. MLPs were first identified to be essential for the production of the NRPS-assembled siderophore pyoverdine (14) and subsequently the production of clorobiocin, calcium-dependent antibiotic, coelichelin, and enterobactin (5, 14–17). Interestingly, some of these studies also showed that MLPs encoded by one BGC are able crosstalk with the NRPSs coded by other BGCs and influence metabolite production of that non-cognate system (8, 16); however, this crosstalk is only relevant when the gene coding for the cognate MLP is deleted.

The impact MLPs have on the production of nonribosomal peptide biosynthesis has led to biochemical and structural studies aimed at understanding the function of MLPs. Biochemical characterization of the NRPSs in the presence and absence of MLPs established that MLPs influence the solubility of overproduced NRPSs (5, 6, 19) and can also impact various aspects of NRPS enzymology including the K_m for the amino acid substrate or ATP (5–7, 27). We recently showed that incorrect MLP/NRPS pairings can result in decreased levels of amino acid thioesterification to the PCP and alterations of the overall NRPS turnover (18). These observations suggest that the long-term goal of generating hybrid NRPS to produce “unnatural” natural products will be complicated by the impact of incorrect MLP and NRPS pairings.

In addition to biochemical studies, there have been a number of structural studies. The initial structures of isolated MLPs determined they possess a unique protein fold and do not contain a clearly identifiable catalytic site (17, 25). Structures of MLP/NRPS complexes showed the MLPs bind to the A domains of NRPSs but at a location that suggests it does not conflict with other catalytic domains. Unfortunately, structural analyses of the NRPS EntF with and without an MLP failed to identify substantive conformational changes that would explain how an MLP imparts its activity on its partner NRPS (26, 27, 29). One original hypothesis about the function of MLPs was that they function as chaperone-like proteins to help NRPSs fold correctly (5, 6). Consistent with this hypothesis, under non-denaturing protein electrophoresis and size-exclusion chromatography, an MLP/NRPS complex migrates or elutes differently compared to the NRPS alone, suggesting that the structural changes caused by the MLP were not detected by the previous crystallographic methods (29). In summary, the specific role of MLPs in NRPS enzymology remains unclear.

MLPs are essential components of many NRPSs and our understanding of their role in catalysis will not only contribute to a better understanding of NRPS enzymology, but will also improve

combinatorial biosynthesis efforts. A significant number of medically-relevant nonribosomal peptides have MLP-dependent A domains, suggesting the construction of hybrid NRPSs to generate next-generation drugs will require an intimate knowledge of how MLPs impart their function on NRPSs (9). In support of this effort, our recent work investigated the effect of pairing non-cognate MLPs with the enterobactin NRPS to understand how MLPs influence NRPS enzymology. We found that non-cognate MLP/NRPS interactions are complex and improper pairing is detrimental to various steps in NRPS catalysis (18).

In this current study, we identified, from multiple bacteria, 50 orphan MLPs that are encoded by genes which are not located near any NRPS-encoding BGCs or genes predicted to be involved in nonribosomal peptide biosynthesis. We investigate the role that one of these orphan MLPs, MXAN_3118, plays in the bacterium *Myxococcus xanthus* DK1622. Using a combination of assays that investigate NRPS solubility, MLP/NRPS co-purification, and *in vivo* MLP/NRPS interactions, we determined that the only MLP encoded by *M. xanthus* DK1622 interacts with NRPS systems from at least seven different BGCs. This is the first example of an MLP that naturally functions with multiple NRPS systems in a single organism. As part of this investigation, we present evidence that one of the MLP-dependent NRPSs that is encoded by an orphan NRPS produces a valine-rich natural product that influences clumping of *M. xanthus* DK1622 in minimal media. The flexibility of MXAN_3118 to interact with multiple NRPS systems suggests that it may be exploited as a universal MLP for the combinatorial biosynthesis of functional hybrid NRPSs.

Results

MXAN_3118 is an orphan MLP in M. xanthus DK1622. The Basic Logic Alignment Search Tool (BLAST) (31), with the default settings, was used to scan the predicted proteome of *M. xanthus* DK1622 for proteins belonging to the MLP superfamily using YbdZ from *Escherichia coli* (accession number YP_588441) as the search query. A single protein, MXAN_3118, was identified. Unexpectedly, the gene coding for this MLP was not near any other genes encoding NRPS enzymology or enzymology known to be associated with the production of a nonribosomal peptide, making this an orphan MLP. Analyses of fully or partially sequenced bacterial genomes using MXAN_3118 as a search query in the Integrated Microbial Genomes & Microbiome Samples from the Joint Genome Institute (32), using the default settings and analyzing the proteins encoded by the 20 kbp on either side of the gene encoding a MLP, identified 43 additional orphan MLPs in other bacteria (Table 1). The presence of these orphan MLPs raised the question of whether they have partner NRPSs or whether they play some other role in bacterial physiology. The former proposal was considered because of our recent report that MXAN_3118 can functionally replace YbdZ, the MLP that naturally functions with the NRPS involved in enterobactin biosynthesis in *E. coli* (18). The latter proposal was considered since the genome of *Wenzhouxiangella marina* KCTC 42284 codes for an MLP but does not code for any NRPSs (33).

Phylum	Order	Species	Locus tag	Accession number
Proteobacteria	*Myxococcales*	*Myxococcus xanthus* DK1622	MXAN_3118	WP_011553168.1
		Myxococcus hansupus	Ga0131205_11350	WP_002640683.1
		Corallococcus coralloides	COCOR_04933	WP_014397734.1
		Myxococcus xanthus DSM 16526	Ga0131205_11350	WP_014397734.1
		Myxococcus xanthus DZ2	MXAN_05015	none[a]
		Myxococcus xanthus DZF1	MXF1DRAFT_05386	none[a]
		Myxococcus virescens DSM 2260	Ga0070493_12057	none[a]
		Myxococcus fulvus HW-1	LILAB_23315	YP_004667636
		Myxococcus macrosporus DSM14697	MYMAC_003063	WP_095958640.1
		Archangium gephyra DSM2261	Ga0070489_111132	WP_043407274.1
		Archangium violaceus Cb vi76	Q664_03470	WP_043389803.1
		Stigmatella aurantiaca DSM 17044	Ga0131204_109251	WP_075007919.1
		Stigmatella aurantiaca DW4/3-1	STAUR_1870	WP_002613859.1
		Stigmatella erecta DSM 16858	Ga0131206_105306	WP_093519887.1
		Chondromyces crocatus Cm C5	Ga0098230_11567	WP_050428970.1
		Chondromyces apiculatus DSM436[c]	CAP_5348	WP_04424566.1
		Cystobacter fuscus DSM 2262	D187_000707	WP_002622872.1
		Cystobacter ferrugineus Cbfe23	Ga0248592_1046	WP_071899239.1
		Cystobacter fuscus DSM 52655[c]	Ga0248597_118675	WP_095990516.1
		Melittangium boletus DSM14713	Ga0248600_116732	ATB28426.1
		Melittangium boletus DSM14713[c]	MEBOL_RS09450	WP_095977109.1
		Nannocystis exedens ATCC 25963	Ga0008035_00092	WP_096333190.1
		Sorangium cellulosum So ce 56	sce3255	WP_012235886.1
		Sorangium cellulosum So ce 56[c]	SCE_RS36450	WP_012239717.1
		Sorangium cellulosium So0157-2	SCE1572_19685	KYG04883.1
		Haliangium ochraceum DSM14365[c]	Hoch_4320	WP_012829413.1
		Vitiosangium sp. GDMCC 1.1324[c]	DAT35_RS11390	WP_108066965.1
	Burkholderiales	*Sphaerotilus natans* ATCC1338	Ga0111604_11866	WP_037483010.1
	Rhizobiales	*Methylosinus trichosporium* OB3b	MettrDRAFT_3629	WP_003613392.1
	Rhodocyclales	*Azovibrio restrictus* DSM 23866	G474DRAFT_01118	WP_026686118.1
	Ktedonobacteria	*Thermosporothrix hazakensis* ATCC BAA-1881	EI42DRAFT_02385	WP_111322093.1
	Chromatiales	*Wenzhouxiangella marina* KCTC 42284[b]	Ga0098257_112714	AKS43021.1
	Desulfobacterales	*Desulfofaba hansenii* DSM 12642	Ga205213_1005119	WP_100391142.1
	Sphingomonadales	*Sphingomonas* sp. LM7	Ga0136193_11785	WP_077508878.1
Chloroflexi	*Herpetosiphonales*	*Herpetosiphon aurantiacus* DSM 785	Haur_00026	ABX02678.1
		Herpetosiphon geysericola DSM 7119	Ga0098740_135275	WP_054536649.1
	Caldinileales	*Caldilinea* sp. AAV1	CLDDRAFT_02804	none[a]
Deinococcus-Thermus	*Deinococcales*	*Deinococcus misasensis* DSM 22328	Q3710DRAFT_01978	WP_034339197.1
Bacteroidetes	*Flavobacteria*	*Cyclobacterium lianum* CGMCC 1.6102	Ga0070154_10460	WP_073093884.1
		Cyclobacterium halophilum IBRC-M 10761	Ga0075264_101710	WP_092169494.1
	Cytophagales	*Algoriphagus machipongonensis* PR1	ALPR1_09740	WP_008200161.1
		Methylovulum psychrotolerans HV10_M2	Ga0226488_11228	WP_088620285.1
		Emticicia oligotrophica DSM 17448	Emtol_3519	AFK04647
		Rudanella lutea DSM19387	RudluDRAFT_0365	WP_019986628.1
	Sphingobacteria	*Mucilaginibacter* sp. PPCGB2223	BEL04_21110	WP_066007208.1
Cyanobacteria	*Cyanophyceae*	*Fischerella* sp. PCC9339	PCC9339DRAFT_06326	WP_017307780.1
		Oscillatoria sp. PCC10802	Osc10802DRAFT_4715	WP_017718167.1
		Crinalium epipsammum PCC 9333	Cri9333_2149	WP_015203141.1
	Nostocales	*Nostoc piscinale* CENA21	Ga0100955_11598	WP_062288309.1
Actinobacteria	*Actinomycetales*	*Streptomyces leeuwenhoekii* DSM 42122	Ga0133265_134129	WP_0293883247.1

Table 1. Organisms with orphan MbtH-like proteins.

[a] These orphan MLPs are found on contigs from genome sequencing projects that have yet to be completely assembled; thus, they have yet to be assigned an accession number.
[b] No NRPS-encoding gene was identified in this genome.
[c] These orphan MLPs were identified by BLAST search using MXAN_3118 as query. All other orphan MLPs were identified by Integrated Microbial Genomes & Microbiome Samples from the Joint Genome Institute.(32)

Interestingly, approximately half of the orphan MLPs identified were found in the *Myxococcales* order. Since the JGI IMG databank did not include all available genomic information, we identified all of the MLPs coded by *Myxococcales* using MXAN_3118 as a search query in BLAST using the default settings. Fifty-seven MLP homologs were identified and manually curated to identify a total of 27 orphan MLPs. A phylogenetic analysis of these 57 MLPs show that orphan MLPs are found throughout the tree and do not cluster into a single clade (Fig. 1). Interestingly, MLPs from *Myxococcus* species do form sister clades and these are the only MLPs encoded in the genomes of these organisms. This is somewhat surprising since these organisms are well-known to be prolific producers of nonribosomal peptides (34, 35). The similarity between these MLPs may suggest they have a common role in these organisms. We focused our attention on the orphan MLP from *M. xanthus* DK1622 since this bacterium is well-studied and can be genetically manipulated.

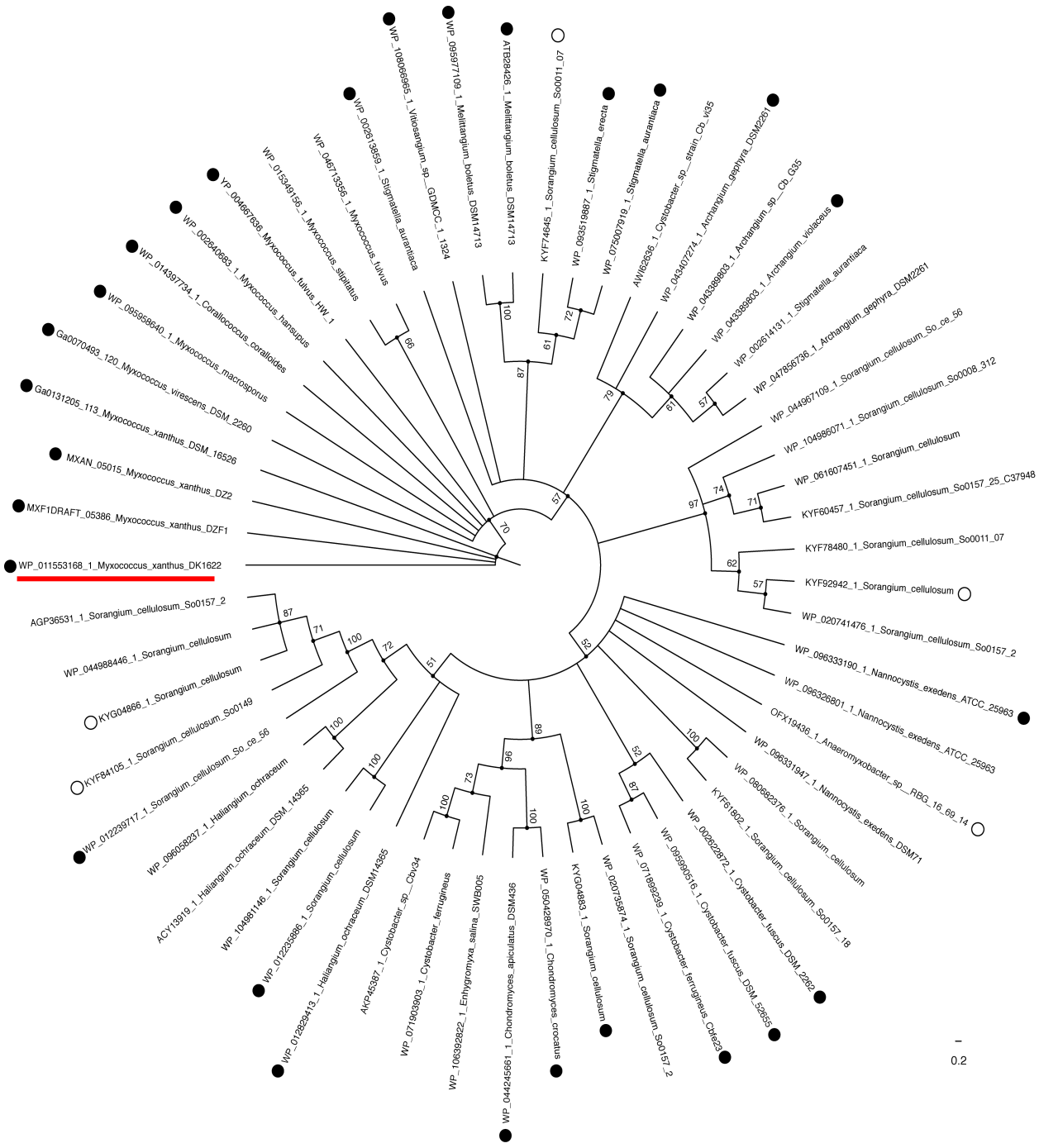


Figure 1. Phylogenetic analysis of *Myxococcales* MLP homologs. An alignment of amino acid sequences of 57 MLPs from the *Myxococcales* order was used to generate a Bayesian phylogenetic tree using Mr.Bayes 3.2.6 version (1). Tree was calculated using the Felsenstein (F81) substitution model plus Gamma. Bootstrap support is indicated by number at the node after 10,000,000 iterations. The filled circles represent orphan MLPs and the open circles indicate MLPs that are found on contigs from incomplete genome sequencing projects. MXAN_3118 is underlined in red. For each MLP, the name of the species and the accession number of the protein is noted. Scale bar indicates substitutions per site.

Identification of potential MXAN_3118 NRPS partners. We compared regions of the genomes surrounding the MXAN_3118 homologs in *Myxococcus* species. All *M. xanthus* strains, *M. macrosporus* DSM 14697, *M. virescens* DSM 2260, *M. fulvus* HW-1, and *M. hansupus* DSM 436 have similar genes around the gene coding the homolog of MXAN_3118 (Fig. 2). In contrast, *M. stipitatus* DSM 14675 and *M. fulvus* DSM 16525 have NRPS-encoding genes immediately upstream of the homolog of MXAN_3118. In *M. stipitatus* DSM14675, the NRPS-encoding gene is associated with a putative 63-kbp BGC followed by homologs to MXAN_3119 and MXAN_3120 (Fig. 2). The genome of *M. fulvus* DSM 16525 was not complete at the time of our analysis, but a similar BGC is expected based on the homology of the NRPS-encoding gene immediately upstream of the MXAN_3118 homolog and a second contig containing homologs of MXAN_3119 and MXAN_3120 and the majority of the genes observed in the 63-kbp BGC from *M. stipitatus* DSM14675. While *M. stipitatus* DSM14675 and *M. fulvus* DSM16525 appear to share this BGC, it is not found in the other *Myxococcus* species. One hypothesis is that these other species lost the corresponding BGC but retained the gene coding for the MLP due to its requirement for one or more of the other NRPSs that are encoded by these bacteria.

We (18) and others (27, 30) have reported that there is no signature amino acid sequence within an A domain that identifies it as an MLP-dependent or -independent megasynthase. Therefore, to identify candidate NRPSs in *M. xanthus* DK1622 for interactions with the MXAN_3118 we identified BGCs that code for an NRPS with significant amino acid identity/similarity (49%/64%) to MYSTI_03598 and code for enzymes that had similar characteristics to the enzymes coded by BGC in *M. stipitatus* DSM 14675. These characteristics were: 1) initiation with an acyl-CoA ligase domain, 2) a module with an A domain disrupted by an oxidase domain, and 3) a hybrid NRPS/polyketide synthase (PKS) system. From this analysis, we identified two potential BGCs

coding for NRPS components that were candidates for MXAN_3118 interactions (Fig. 3). These clusters were MXAN_3778-3779, with MXAN-3779 coding for the hybrid NRPS/PKS that

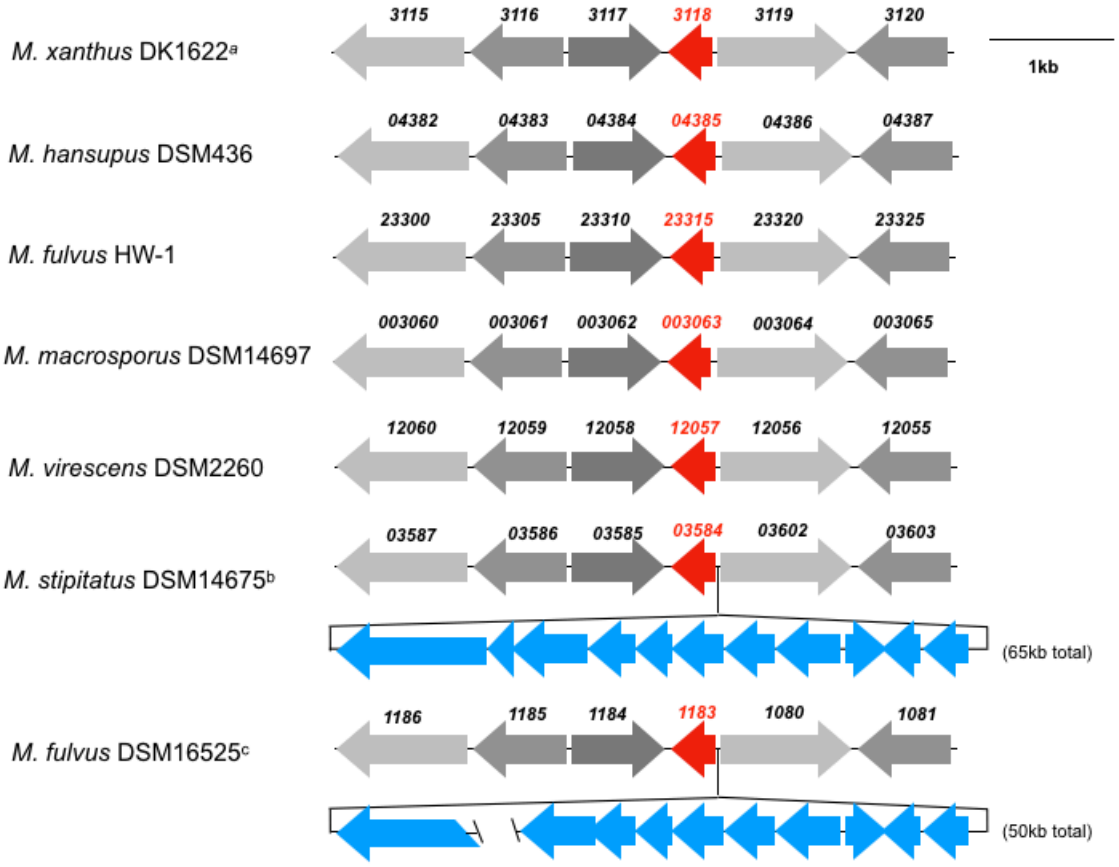


Figure 2. Schematic of *Myxococcus* species and strain genomic regions surrounding the MXAN_3118 homolog (red arrow). Similar gray shading identifies homologs of MXAN_3115-3117 and MXAN_3119-3120. The 1-kb bar indicates the scale for all genomic regions except those show in shades of blue, the size of which are noted in parentheses. Locus tag abbreviations: *M. xanthus* DK1226 (MXAN_3115-3120); *M. virescens* DSM 2260 (ga0070493_12060-12055); *M. fulvus* HW-1 (LILAB_23300-23325); *M. hansupus* DSM436 (A176_04382-04387); *M. macrosporus* DSM14697 (MYMAC_003060-65); *M. stipitatus* DSM14675 (MYSTI_03587-03603); *M. fulvus* DSM16525 (Ga0131203_1181, 1183-1187). Superscripts: a) *M. xanthus* DK1622 is representative of all other *M. xanthus* strains (DSM16525, DZ2, DF1); b) the blue genes are an NRPS-associated gene cluster covering 65 kb in *M. stipitatus* DSM14675. The NRPS-associated gene cluster has homolog ORFs in *M. fulvus* DSM 16525; c) genome sequence was incomplete and two contigs were identified, the first contig covers 24 kb, terminating in the middle of the NRPS-encoding gene 1181 and the second contig is 26 kb starts in the middle of gene 1071 to 1079, missing the homolog for gene MYSTI_03592 in the *M. fulvus* DSM 16525 cluster.

produces myxoprincomide (35), and an orphan BGC consisting of MXAN_3634-3636. We targeted select A domains or the associated modules from MXAN_3779 and MXAN_3634-3636 to evaluate them for interactions with MXAN_3118.

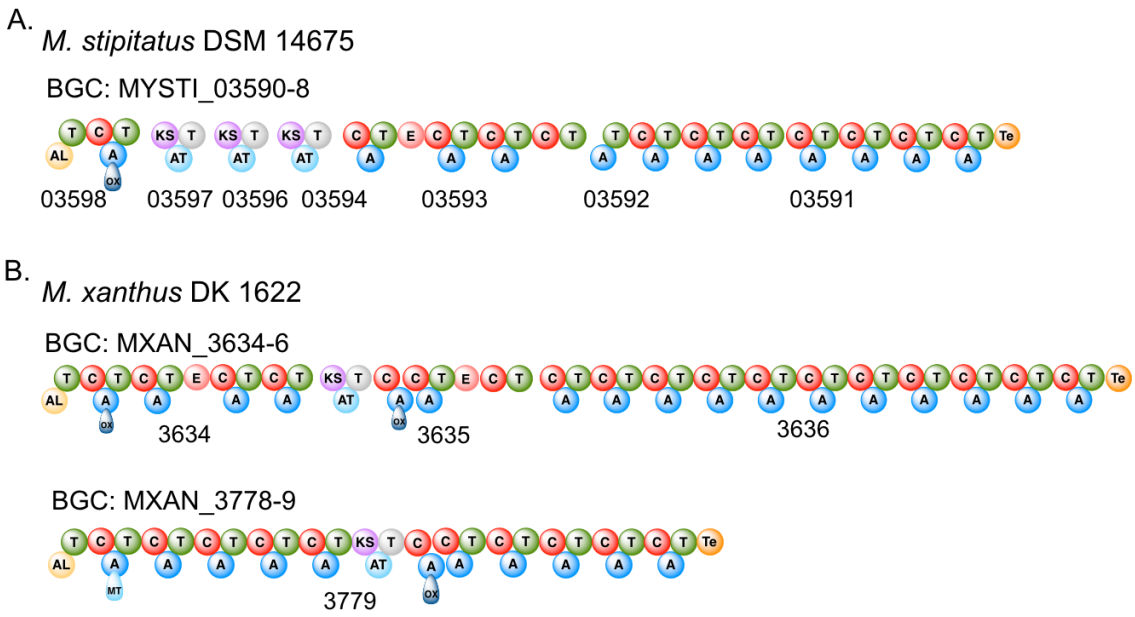


Figure 3. Potential NRPS partners for the MXAN_3118 orphan MLP. A) Schematic of the MYSTI_03590-8 NRPS/PKS hybrid megasynthase from *M. stipitatus* DSM 14675. B) Schematic of the MXAN_3779 and MXAN_3634-6 NRPS/PKS hybrid megasynthases from *M. xanthus* DK1622. Domain abbreviations: AL, acyl-CoA ligase; T, thiolation; C, condensation; ox, oxidation; KS, ketosynthase; AT, acyltransferase; E, epimerase; Te, thioesterase.

MXAN_3118 interacts with the myxoprincomide (Mpc) mixed NRPS/PKS megasynthase. Two A domains from the Mpc mixed NRPS/PKS were assessed to determine whether they interact with MXAN_3118 or not. We have previously established that no single assay is capable of detecting all MLP/NRPS interactions (18, 21). Due to this we used more than one assay to assess MXAN_3118 interactions with the A domains ($A5_{Mpc}$ and $A12_{Mpc}$) of two modules ($M5_{Mpc}$ and $M12_{Mpc}$) (Fig. 3A). One assay was an *in vivo E. coli*-based B2H assay to detect *in vivo* A domain/MXAN_3118 interactions. The other assays detected whether the solubility of an NRPS was influenced by the presence of MXAN_3118 or whether untagged MXAN_3118 protein co-

eluted with His-tagged NRPS component from a Ni-NTA column. We define an NRPS as MXAN_3118 dependent if one or more of the associated A domains or modules is positive in the B2H assay and is also positive for increased solubility in the presence of MXAN_3118 or co-elutes with MXAN_3118 from a Ni-NTA column when only the NRPS is His-tagged.

A B2H system, a well-established mechanism for detecting protein-protein interactions (36), has not been previously used to assess MLP/NRPS interactions. To investigate whether this system is able to detect these interactions, the enterobactin system was used as a positive control. Briefly, the A domain of EntF was fused to the α-subunit of RNA polymerase and YbdZ, the MLP of the enterobactin system (5), was fused to λcI. Constructs expressing these fusion proteins were introduced into an *E. coli* Δ*ybdZentF* strain containing a *lacZ* reporter that has increased expression when there are protein-protein interactions between the two fusion proteins (36). When the YbdZ-λcI fusion protein is present in the cells along with the EntF A domain-α-subunit fusion, 10,000-fold higher LacZ activity was detected compared to a strain producing the EntF A domain-α-subunit fusion and the λcI protein alone (Fig. 4B), confirming this assay is able to detect MLP/NRPS interactions.

The assay was repeated with MXAN_3118 fused to the λcI protein and $A5_{Mpc}$ or $A12_{Mpc}$ fused to the α-subunit of RNA polymerase. We defined the B2H assay as testing positive for MLP/NRPS interactions when the LacZ activity detected was statistically higher than the activity detected the negative control strain expressing the A domain-α-subunit fusion and the λcI protein alone. Both $A5_{Mpc}$ or $A12_{Mpc}$ were positive for interactions with MXAN_3118 (Fig. 4B) when compared to the negative control. These results suggest that both A domains from the Mpc NRPS are MXAN_3118 dependent. The variability in the activity detected by the fusions is not unexpected. As we have

shown (21), an MLP can have significantly different levels of interaction with each A domain from the same NRPS system.

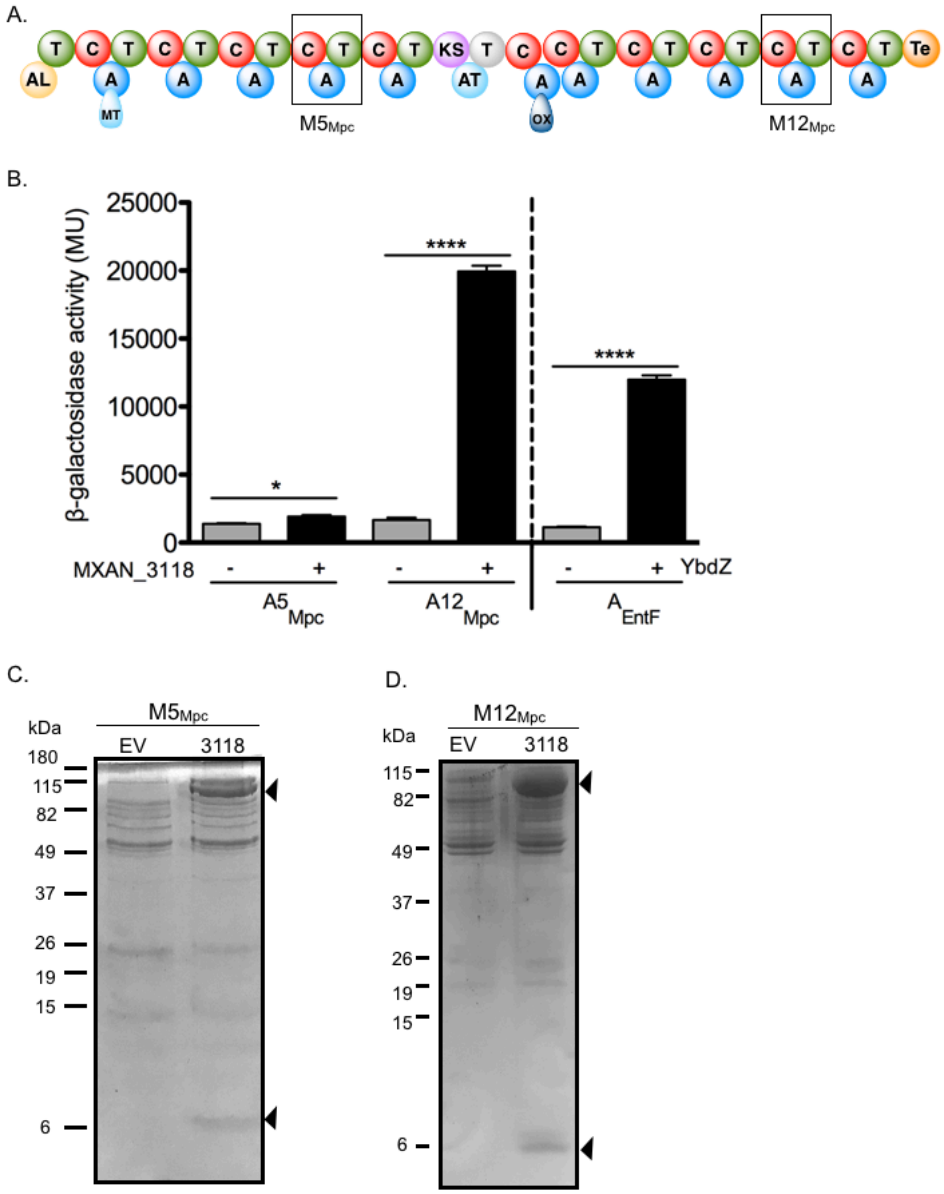


Figure 4. Myxoprincomide (Mpc) NRPS protein interactions with MXAN_3118. A) Schematic of the MXAN_3779 NRPS/PKS hybrid assembly line with boxes highlighting the fourth (M5 $_{Mpc}$) and tenth (M12 $_{Mpc}$) modules investigated for MXAN_3118 interactions. Domain abbreviations: AL, acyl-CoA ligase; T, thiolation; C, condensation; ox, oxidation; MT, methyltransferase; KS, ketosynthase; AT, acyltransferase; Te, thioesterase. B) Results from the B2H assay: Quantitative β-galactosidase assay of the α-subunit-A domain fusions with MXAN_3118-λcI fusion (+) or λcI alone (-). EntF (A domain) and YbdZ (MLP) were used as controls of MLP-NRPS positive interactions. Error bars show standard deviations between three independent assays. P values

were calculated using the Student's t test. (* means $p<0.01$, ** means $p<0.001$, **** means $p<0.0001$). C) Tris-tricine (16.8%) polyacrylamide gels and Coomassie blue staining of His-tagged $M5_{Mpc}$ eluting from an Ni-NTA column after overproduction without (EV, empty expression vector) or with (3118, expression vector expressing MXAN_3118) untagged MXAN_3118. Ten micrograms of protein were loaded in each lane. D) Tris-tricine (16.8%) polyacrylamide gels and Coomassie blue staining of His-tagged $M12_{Mpc}$ eluting from an Ni-NTA column after overproduction without (EV, empty expression vector) or with (3118, expression vector expressing MXAN_3118) untagged MXAN_3118. Seventeen micrograms of protein were loaded in each lane.

As a second test for MXAN_3118 dependence, the two modules containing the $A5_{Mpc}$ and $A12_{Mpc}$ domains were heterologously overproduced in *E. coli* with N-terminal His tags, co-overproduced with and without untagged MXAN_3118, and then partially purified by Ni-NTA chromatography to determine whether the two proteins co-elute. Both NRPS modules co-eluted with MXAN_3118 (Fig. 4C). In control reactions, strains lacking MXAN_3118 failed to produce a module that eluted from the column or had significantly less soluble module produced. Additionally, untagged MXAN_3118 had no affinity for the Ni-NTA resin. Based on these results, we conclude that at least these modules of the Mpc NRPS/PKS hybrid enzyme are MLP dependent. We did not determine whether DK1622 lacking MXAN_3118 produced Mpc because this strain produces very low levels of Mpc. A different strain of *M. xanthus* had to be used to produce enough Mpc for characterization and required a 300 L fermentation to produce enough metabolite for analysis (35).

MXAN_3118 interacts with all but one of the A domains from the megasynthase encoded by the orphan BGC MXAN_3634-3636. The same assays were used to assess whether components from the NRPS/PKS megasynthase coded by the BGC MXAN_3634-3636 interact with MXAN_3118. Initial studies focused on representative A domains from each subunit of the megasynthase (Fig. 5A). The B2H system detected significant interactions between the A domains from modules 3, 8, and 15, but not from the A domain of module 11 (Fig. 5B). As most of these A domains are positive for interactions with MXAN_3118, we investigated all the A

domains for interactions with MXAN_3118 by assessing whether the NRPS components co-purified with MXAN_3118 after heterologous co-overproduction in *E. coli*. MXAN_3118 co-purified

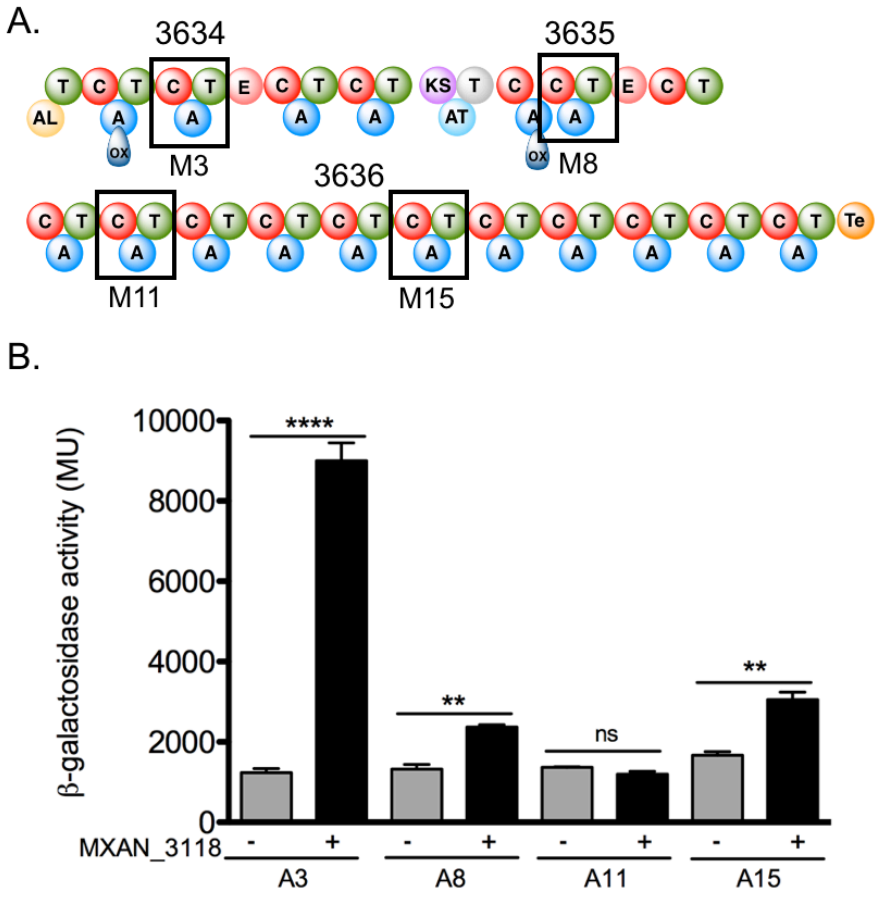


Figure 5. MXAN_3634-36 interactions with MXAN_3118. A) Schematic of the MXAN_3634-36 NRPS/PKS hybrid assembly line. Boxes highlight the third (M3), eighth (M8), eleven (M11) and fifteen (M15) module investigated for MLP interactions. Domain abbreviations: AL, acyl-CoA ligase; T, thiolation; C, condensation; ox, oxidation; KS, ketosynthase; AT, acyltransferase; E, epimerase; Te, thioesterase. B) Results from B2H assays: Quantitative β-galactosidase assay of the α-subunit-A domain fusions with MXAN_3118-λcI fusion (+) or λcI alone (-). Error bars show standard deviations between three independent cultures. P values were calculated using the Student's t test (* means p<0.01, ** means p<0.001, **** means p<0.0001, ns means not significant).

with the A domain or complete Module from 16 of the 17 NRPS modules (Fig. 6), including the A domain from module 11 that was negative in the B2H assay. These results support our finding that no single assay will detect all possible MLP/NRPS interactions (18). The only NRPS module

that tested negative in this assay was module 7, which contains an A domain that lacks some of the signature sequences for an A domain, is disrupted by an oxidase (Ox) domain, and also lacks a PCP domain (Fig. 5A). As discussed in the next section, our *in vitro* analysis of this A domain supports the hypothesis that this A domain is inactive.

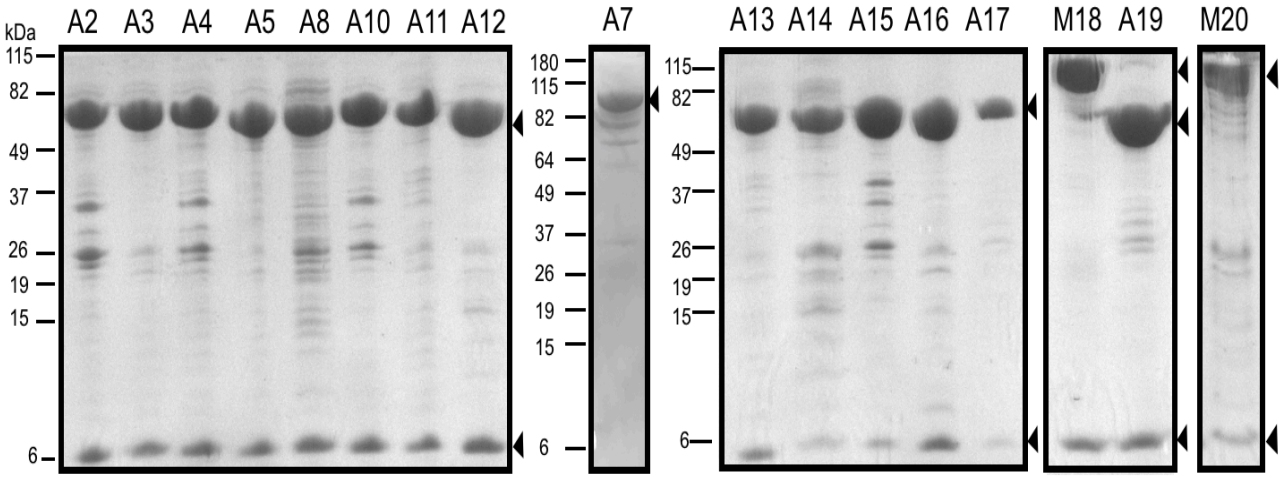


Figure 6. Orphan MXAN_3634-3636 NRPS/PKS megasynthase is MXAN_3118 dependent. Tris-tricine (16.8%) polyacrylamide gels and Coomassie blue staining to detect untagged MXAN_3118 co-elution with histidine-tagged A domains (e.g. A2) or modules (e.g. M18), as noted. Top arrow heads point to the A domain or module and bottom arrow heads point to MXAN_3118. (10 μg of protein loaded in each lane).

The orphan BGC MXAN_3634-3636 codes for an MXAN_3118-dependent megasynthase that is likely to produce a valine-rich natural product that influences aggregation of M. xanthus 1622 in minimal media. With each A domain or module partially purified, we assayed each enzyme for amino acid specificity using standard ATP/PP_i exchange assays (Fig. 7). Each A domain or module was assayed for activation of proteinogenic amino acids. We focused our efforts on these potential substrates since the BGC consisting of MXAN_3634-3636 does not code for any non-

proteinogenic amino acid biosynthesis enzymes and the amino acid specificity codes of each complete A domain suggested the activation of a proteinogenic amino acid (Table 2).

Table 2. Orphan NRPS/PKS MXAN_3634-6 A domain substrate specificity.

A domain/Module	Stachelhaus code[a]	Predicted AA	AA specificity[b]
A2	DILQLGMIWK	Gly	Val
A3	DVWHFSLVDK	Ser	Ser
A4	DAFFHGVTFK	Ile	Val
A5	DAFWLGGTFK	Val	Leu
A7	–	-	-
A8	DALTIAGVCK	Phe	Tyr
A10	DTENVGTAVK	Lys	Val
A11	DALWLGGTFK	Val	Val
A12	DAPQVGAVDK	Arg	Arg
A13	DALWLGGTFK	Val	Val
A14	DLTKVGHVGK	Asp	Asp
A15	DALWLGGTFK	Val	Val
A16	DALWLGGTFK	Val	Val
A17	DALWLGGTFK	Val	Val
M18	DALWLGGTFK	Val	Val
A19	DALWLGGTFK	Val	Val
M20	DLFNNALTYK	Ala	Gly/Ser

[a] Substrate specificity determined by NRPSpredictor2. (6)
[b] Amino acid (AA) activated by the purified NRPS.

Of the 17 NRPS A domains or modules, 16 of them activated an amino acid substrate. Module 7, containing an A domain we hypothesized to be inactive, failed to co-purify with MXAN_3118 and also failed to activate any amino acid substrate. These data suggest that this module does not incorporate an amino acid into the mixed nonribosomal peptide-polyketide product. Nine of the 16 remaining A domains activated the amino acid L-valine, suggesting that the final product is a valine-rich natural product. The initiating module of the megasynthase is a didomain module with a homolog of a medium-chain acyl-CoA ligase and a PCP domain. Repeated attempts to overproduce soluble forms of module 1 with and without MXAN_3118 in *E. coli* failed; thus, at this time we do not know if this module is active.

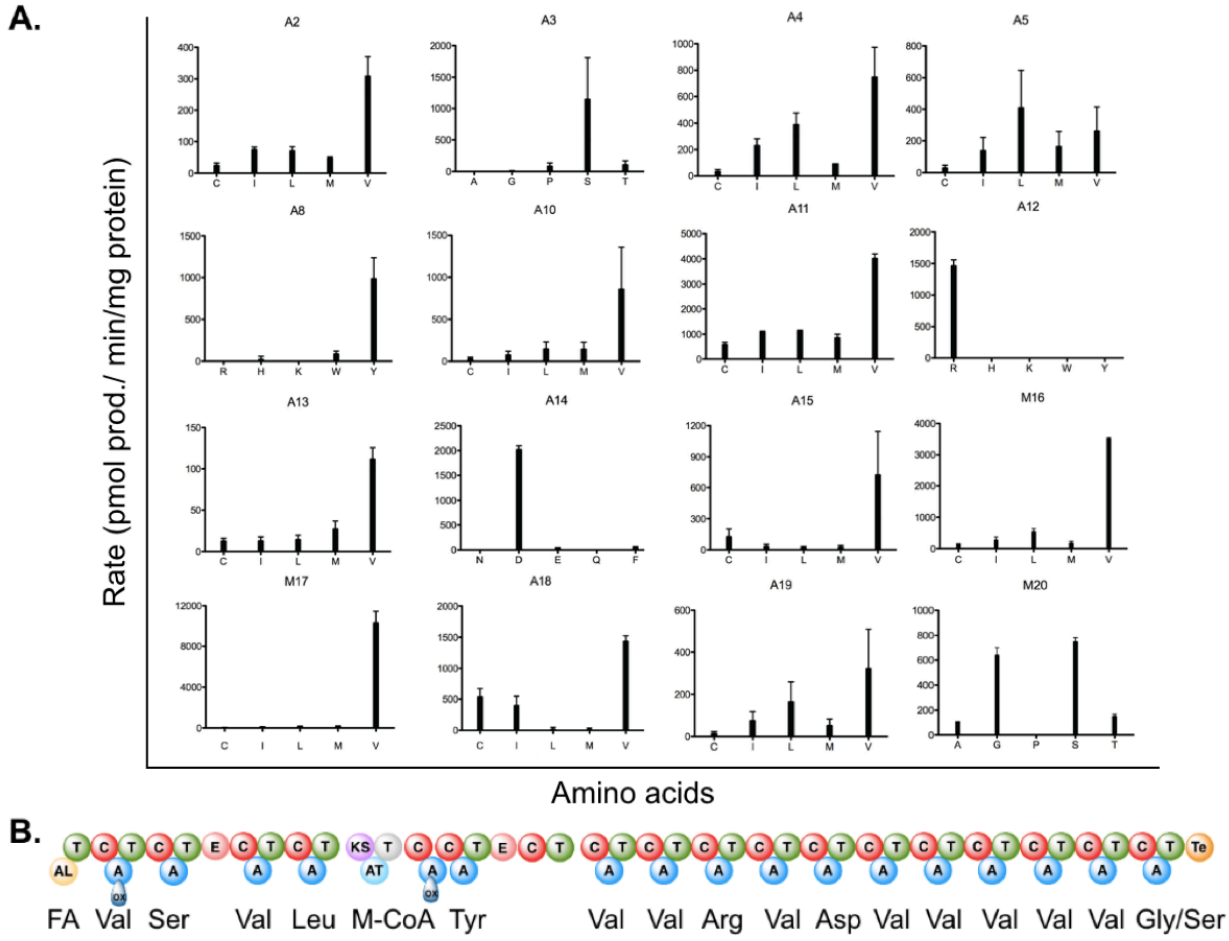


Figure 7. Amino acid specificity of the A domains from the MXAN_3634-36 megasynthase. A) Results from ATP/PP_i exchange assays. Error bars represent the standard deviations between three independent assays. Substrates identified with the one letter amino acid code. B) Schematic of the MXAN_3634-3636 assembly line diagram and amino acid substrate specificity of A domains. Domain abbreviations: AL, acyl-CoA ligase; T, thiolation; C, condensation; ox, oxidation; KS, ketosynthase; AT, acyltransferase; E, epimerase; Te, thioesterase.

not incorporate an amino acid into the mixed nonribosomal peptide-polyketide product. Nine of the 16 remaining A domains activated the amino acid L-valine, suggesting that the final product is a valine-rich natural product. The initiating module of the megasynthase is a didomain module with a homolog of a medium-chain acyl-CoA ligase and a PCP domain. Repeated attempts to

overproduce soluble forms of module 1 with and without MXAN_3118 in *E. coli* failed; thus, at this time we do not know if this module is active.

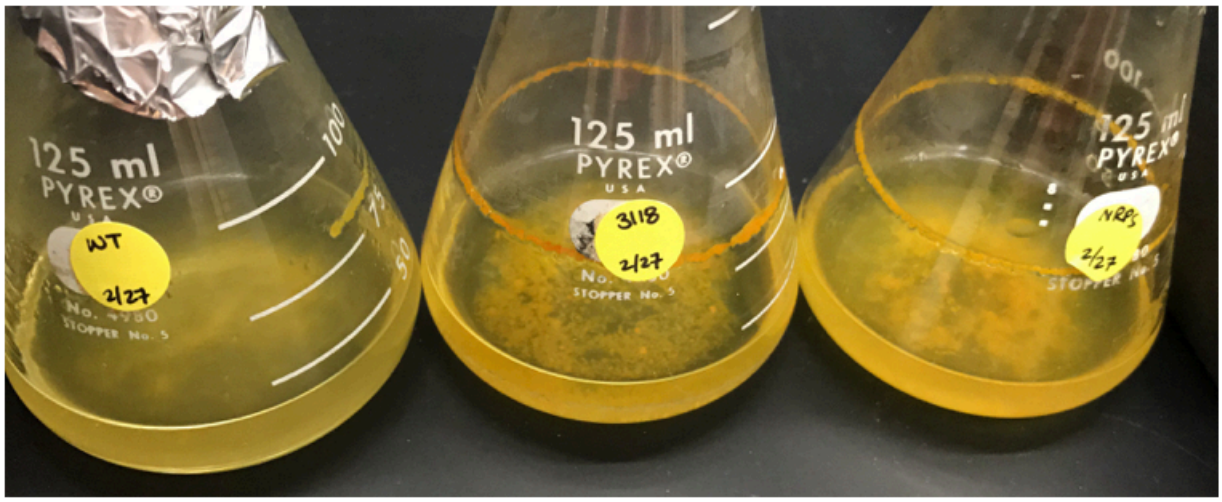


Figure 8. Aggregation phenotype associated with loss of MXAN_3118 or MXAN_3634-3636. Relevant genotype is noted below each flask. Flasks shown are representative results of three biological replicates.

Repeated attempts to detect a metabolite associated with MXAN_3634-3636 were not successful. We note that Müller and colleagues detected components of this megasynthase in proteomic studies but also failed to detect an associated metabolite (37). During our attempts to detect an associated metabolite, we observed that a strain of *M. xanthus* DK1622 lacking the BGC MXAN3634-3636 aggregated when grown in minimal liquid medium, but the wild-type strain did not (Fig. 8). Consistent with an essential role for MXAN_3118 in the function of the associated megasynthase, a strain lacking MXAN_3118 showed the same clumping phenotype (Fig. 8). These data support our protein-protein interaction assay results that MXAN_3118 plays a role in the function of the megasynthase consisting of MXAN_3634-3636.

MXAN_3118 is likely to be expressed constitutively during vegetative growth of M. xanthus DK1622. Our observation that MXAN_3118 interacts with NRPS components coded by two

distinct BGCs suggested that MXAN_3118 may play a broad role in natural product production by *M. xanthus* DK1622. With this in mind, we used RT-PCR to investigate when the mRNA of MXAN_3118 is produced in DK1622. We detected MXAN_3118 mRNA in both rich and minimal media, (Fig. 9), suggesting that MXAN_3118 is constitutively expressed.

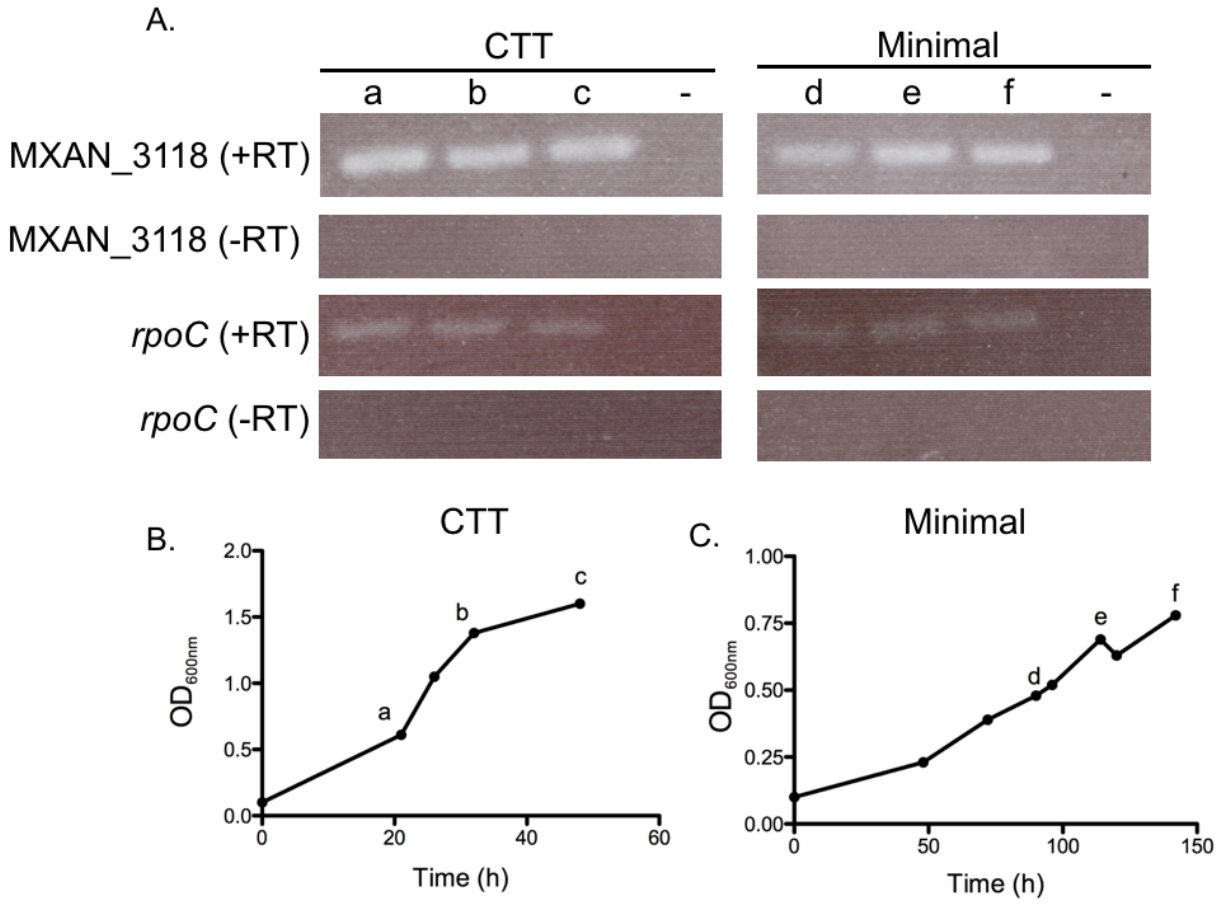


Figure 9. MXAN_3118 mRNA detection during exponential growth of *M. xanthus* DK1622. A) Detection of MXAN_3118 mRNA using RT-PCR analysis. RNA samples were extracted from cells taken from time points during exponential growth in CTT rich media condition (B) and minimal media conditions (C). Samples a, b and c correspond to the time points from CTT media; samples d, e and f to time points from minimal media and (-) correspond to no RT control. Detection of *rpoC* mRNA was used as a positive control and to ensure equal loading of samples. Data shown are representative of three biological replicates.

MXAN_3118 interacts with at least five additional NRPS systems in M. xanthus DK1622. The presence of MXAN_3118 mRNA throughout the growth of *M. xanthus* DK1622 in both rich and minimal media suggests that the coded MLP is also present. Due to this, we investigated whether MXAN_3118 is a "universal" MLP in *M.* xanthus DK1622 that works with any NRPS system that requires this accessory protein. *M. xanthus* DK1622 has 15 BGCs that code for NRPS or NPRS/PKS megasynthases, two of which we already have provided evidence for interactions with MXAN_3118 (Fig. 4 and 5). We targeted one module from each of the remaining megasynthases for analysis by B2H and co-purification assays (Fig. 10A).

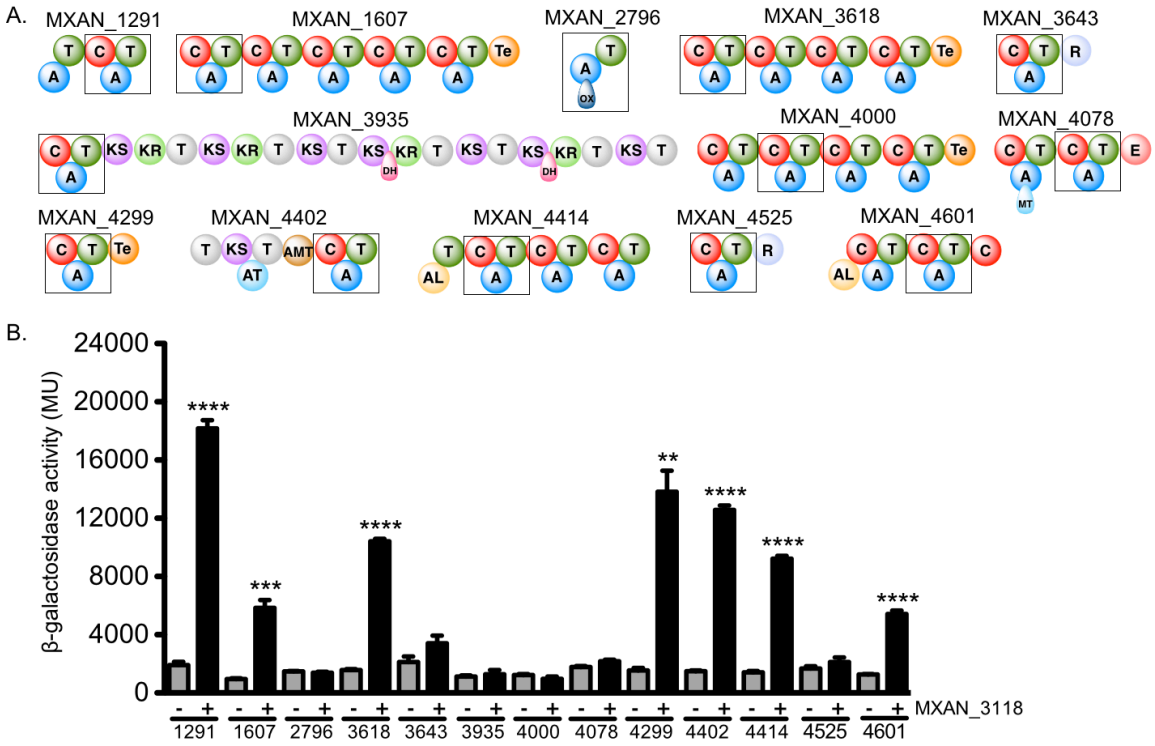


Figure 10. Protein-protein interactions of *M. xanthus* DK1622 NRPS and MXAN_3118 using B2H assay. A) Schematic of the NRPS proteins from the thirteen NRPS encoding gene clusters that were tested. Boxes highlight the module containing the A domain tested. Domain abbreviations: AL, acyl-CoA ligase; T, thiolation; C, condensation; ox, oxidation; KS, ketosynthase; AT, acyltransferase; E, epimerase; Te, thioesterase; R, reductase; KR, ketoreductase; DH, dehydratase; and AMT, aminotransferase. B) Results of a quantitative β-galactosidase assay of the α-subunit-A domain fusions with MXAN_3118-λcI fusion (+) or λcI alone (-). Error bars show standard deviations between three independent cultures. P values were calculated using the Student's t test (* means p<0.01, ** means p<0.001, **** means p<0.0001).

The B2H assay detected statistically significant interactions between MXAN_3118 and A domains from seven NRPS systems (Fig. 10B). Of the seven B2H-positive A domains, five (MXAN_1291, MXAN_1607, MXAN_3618, MXAN_4402, MXAN_4414) had improved solubility and co-purified with MXAN_3118 when heterologous overproduced in *E. coli* (Fig. 11).

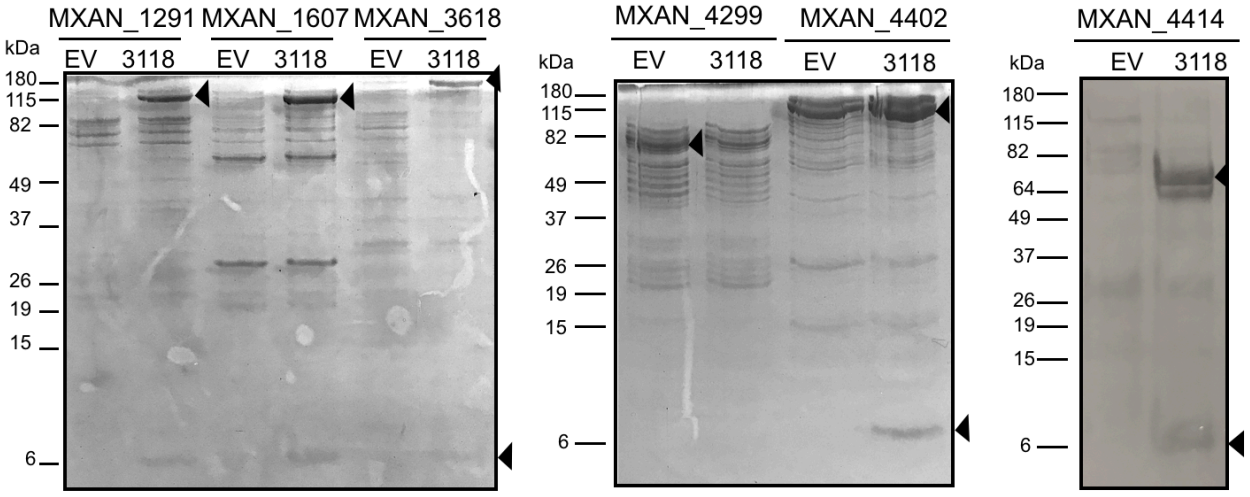


Figure 11. Co-elution of MXAN_3118 with histidine-tagged NRPS components. Tris-tricine (16.8%) polyacrylamide gels and Coomassie blue staining analysis. His-tagged NRPS module eluting from an Ni-NTA column after overproduction without (EV, empty expression vector) or with (3118, expression vector expressing MXAN_3118) untagged MXAN_3118. Top arrow heads point to the NRPS and bottom arrow heads point to MXAN_3118. (10-17 μg of protein were loaded in each lane).

Interestingly, MXAN_4299 was positive in the B2H assay, but did not co-purify with MXAN_3118. We were unable to find an expression construct that produced soluble MXAN_4601 hence we were unable to assess whether it co-purifies with MXAN_3118. In addition to the NRPS systems that were positive for in the B2H assay, we also assessed whether any of the NRPS systems that were negative in the B2H assay co-purified with MXAN_3118 (Fig.12). None of these NRPSs co-purified with MXAN_3118. Thus, based on our requirement for obtaining two independent lines of evidence for MXAN_3118 interactions with an NRPS, we conclude that seven NRPS systems

(Mpc, MXAN_3436-3636, MXAN_1291, MXAN_1607, MXAN_3618, MXAN_4402, and MXAN_4414) in *M. xanthus* 1622 are MLP-dependent.

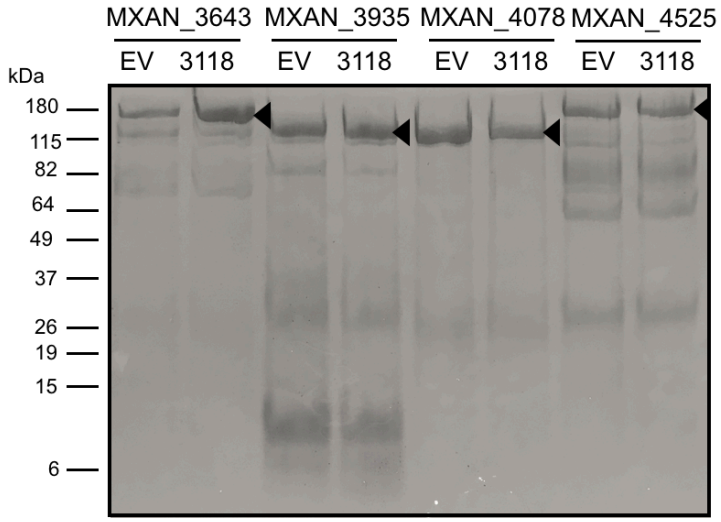


Figure 12. Evaluation of MXAN_3118 impact on NRPS solubility and NRPS co-elution with MXAN_3118. Tris-tricine (16.8%) polyacrylamide gels and Coomassie blue staining analysis of His-tagged NRPSs eluting from an Ni-NTA column after overproduction without (EV, empty expression vector) or with (3118, expression vector expressing MXAN_3118) untagged MXAN_3118. Top arrow heads point to the NRPS proteins. (17 μg of protein loaded in each lane)

Discussion

Members of the MLP superfamily are nearly always encoded within NRPS-associated BGCs or are adjacent to genes involved in the biosynthesis of nonribosomal peptides. This makes is quite clear that the MLP will associate with at least one of the A domains in the associated NRPS. In contrast, here we have identified 50 orphan MLPs found across different bacterial orders that have no clear NRPS partner. In fact, we found that *Wenzhouxiangella marina* KCTC 42284 codes for an MLP but does not contain any NRPS-encoding genes (Table 1). The most highly represented order coding for an orphan MLP is the *Myxococccales* and we targeted a member of this order, *M. xanthus* DK1622, to characterize the function of an orphan MLP encoded by

MXAN_3118. This is the only MLP encoded by the *M. xanthus* DK1622 genome even though this bacterium has 15 BGCs that encode NRPSs (38).

Using a combination of B2H assays and heterologous protein overproduction and purification, we provide evidence that MXAN_3118 is an MLP that naturally interacts with at least seven distinct NRPS systems (Figs. 4-5 and 10, Table 6). It is possible that this MLP interacts with additional NRPS systems. First, while we detected MXAN_3118 interactions with nine NRPS systems using the B2H assay, only seven of these interactions could be confirmed by a solubility or co-purification assay. It is possible these interactions occur, but the solubility or co-purification assays were not sensitive enough to detect them. We have shown previously that some MLP/NRPS interactions are detected by *in vivo* phenotypic assays but fail to be detected by solubility or co-purification and also vice versa (18, 39). Second, we did not test all the A domains in *M. xanthus* DK1622 for interactions with MXAN_3118, and it is possible a different A domain from a targeted NRPS system is MLP dependent as we have reported for the viomycin and capreomycin NRPS systems (5). Regardless, our data support the conclusion that MXAN_3118 interacts with multiple NRPS systems in *M. xanthus* DK1622. In the past, evidence that MLPs cross-react with more than one NRPS system was only observed when the cognate MLP-encoding gene of a BGC was deleted, opening the opportunity for the non-cognate MLP to replace it for metabolite production (15, 16). Our observations with *M. xanthus* DK1622 differ with these prior studies because MXAN_3118 is the only MLP in this bacterium; thus, MLP-dependent NRPS systems are dependent upon the function on this single MLP.

Table 3. Summary of *M. xanthus* DK1622 NRPS-MXAN_3118 interactions.

	Results for NRPS and MXAN_3118 interactions	
NRPS locus tag (A domain) [a]	B2H	Co-purification
MXAN_1291 (A2)	+ [b]	+ [b]
MXAN_1607 (A1)	+ [b]	+ [b]
MXAN_2796	-	ND[c]
MXAN_3618 (A1)	+ [b]	+ [b]
MXAN_3634 to MXAN_3636	+ [b]	+ [b]
MXAN_3643	-	-
MXAN_3935	-	-
MXAN_3779	+ [b]	+ [b]
MXAN_4000 (A2)	-	ND[c]
MXAN_4078	-	-
MXAN_4299	+ [b]	-
MXAN_4402 (A1)	+ [b]	+ [b]
MXAN_4414 (A1)	+ [b]	-
MXAN_4525	-	-
MXAN_4601 (A2)	+ [b]	ND[c]

[a] A lack of parentheses indicates that all A domains of the NRPS(s) were tested.
[b] Categorized as positive for MXAN_3118 interactions.
[c] ND, not determined.

The ability of MXAN_3118 to interact with multiple NRPS systems strongly suggests this protein is a universal MLP in *M. xanthus* DK1622. Our finding that the mRNA for MXAN_3118 is detectable throughout the growth of this bacterium (Fig. 9) is consistent with this MLP playing a role in the function of multiple NRPS systems rather than a single system expressed as a specific stage of growth. With so many required interactions, it is likely that MXAN_3118 has broad NRPS-interaction flexibility. This flexibility may prove invaluable for combinatorial biosynthesis of NRPSs. The construction of hybrid NRPSs will necessitate the use of non-cognate MLP/NRPS pairings. The ability of an MLP to naturally interact with multiple NRPS systems suggests that it is a candidate for a universal MLP to be included in any strain where hybrid NRPSs are being produced. We have previously shown that MXAN_3118 can functionally replace YbdZ, the natural MLP partner of the NRPS involved in enterobactin production in *E. coli*, even though it is the evolutionarily quite distant from YbdZ (18). We propose that MXAN_3118, and potentially other

members orphan MLPs in the *Myxococcales* order, especially those from the *Myxococcus* species, may be a source of universal MLPs. These may prove to be combinatorial biosynthesis tools analogous to the broad range 4'-phosphopantheteinyl transferase Sfp that is a workhorse of combinatorial biosynthesis studies (40).

Materials and Methods

Phylogenetic analysis. A total of 57 *Myxococcales* MbtH-like protein homolog sequences were obtained from a BLAST search using MXAN_3118 as a reference sequence and aligned using MUSCLE (41). Phylogenetic analysis was performed using MrBayes software launched from the Mesquite v3.51 software package (42). The substitution model used was Felsenstein 81 plus Gamma and a Markov Chain Monte Carlo setting for 10,000,000 generations. The MLP from *Desulfobacula phenolica* DSM3384, accession number WP_092230082.1, was used as an outgroup. The phylogenetic tree was visualized using FigTree v1.4.3. (43).

Bacterial strains and plasmid construction. All strains used in this study are listed in Table 4. All overexpression vectors were constructed using a polymerase incomplete primer extension (PIPE)-based method. (44) The primers used to amplify vectors and inserts used in this study are listed in Table 5. The genes coding for the NRPS or PKS components were amplified from *M. xanthus* DK1622 genomic DNA and cloned into pET28b vector with an N-terminal histidine tag. The MXAN_3118 gene was cloned into pACYC Duet-1 in a manner that enabled an untagged MLP to be produced. All plasmids are listed in Table 6.

Table 4. Bacterial strains used in the course of this study.

Strain	Description	Reference
BL21(DE3) ybdZ::acc(3)IV	BL21(DE3) with apramycin resistance gene disrupting *ybdZ*	(5)
FW102 F' placO$_L$2-62-*lacZ*	λCI operator centered at -62 upstream of the lac promoter, *kan* resistant	(45), R.L. Gourse
FW102 Δ*ybdZentF* F' placO$_L$2-62-*lacZ*	Deletion of *ybdZ* and *entF* in FW102 F' placO$_L$2-62-*lacZ*	this study
Myxococcus xanthus		
DK1622	Wild-type strain	(46)
LS3135	Deletion of *mtbH* (MXAN_3118) in DK1622	this study
LS3944	Deletion of MXAN_3634, MXAN_3635, and MXAN_3636 in DK1622	this study

Table 5. Primers used in the course of this study.

Primer name	Sequence (5'-3')
pBR_F	CATGCTACAGCCCTGTGGAG
pBR_R	TAAGGATCCTCTACGCCGGACG
pAC_F	TGACGGAATGTTAATTCTCGTTGAC
pAC_R	TGACCTAGGATCTGCATCGCAGG
pBR1291_F	CCAGAGGCGGCCGCAATGCAGACCTTCATCCT
pBR1291_R	GTCAGGCGCAGGTGTAAGGATCCTCTACGCCG
pBR1607_F	CAGTCCGCGTCCACATTCCTAGGAGATGCGGC
pBR1607_R	CAGAGGCGGCCGCAATGCGCGTGCTGGCGGAT
pBR2796_F	GCTTCGGGGAAGGGAGATTCCTAGGAGATGCGGCC
pBR2796_R	CCAGAGGCGGCCGCAATGAACCCGGAGACCTTCGA
pBR3618_F	GTCCGCGTCCGCGCGCCAATTCCTAGGAGATGCGGC
pBR3618_R	CCAGAGGCGGCCGCAATGGAGCGTCAACAGCTC
pBR3779A4_F	CCAGAGGCGGCCGCAATGGATGAGCAGCACCAG
pBR3779A4_R	CCTGGAGCCGCTTAAGGATCCTCTACGCCGG
pBR3779A10_F	AACCAGAGGCGGCCGCAATGACGGAGGAGCAGCAG
pBR3779A10_R	CGGAGCCAGGAGGCCTAAGGATCCTCTACGCCGG
pBR3935_F	CCAGAGGCGGCCGCAATGGACGACGAAGAGCGC
pBR3935_R	CCTTCTGTAGGCGCCAGATTCCTAGGAGATGCGGC
pBR4078_F	CCAGAGGCGGCCGCAATGGCGCTACTGGGTGCG
pBR4078_R	CGGAAGCGGCTACGCATTCCTAGGAGATGCGG
pBR4414_R	TCCGGAGCCTGTTGCCCATTCCTAGGAGATGC
pBR4414_F	CCAGAGGCGGCCGCAATGGAGGAGGAACGCGCC
pBR4299_F	CCAGAGGCGGCCGCAATGACCCGTCCGGTCCGG
pBR4299_R	TGCGTCGCGCTCCTCATTCCTAGGAGATGCGGC
pBR4402_F	GGCGCGGGCTGGGCCTCATTCCTAGGAGATGCGGCC
pBR4402_R	AAACCAGAGGCGGCCGCAATGGCCGAGCGCCGCAAG
pBR4414_F	CCAGAGGCGGCCGCAATGGAGGAGGAACGCGCC
pBR4414_R	GAGGCCTCGGACAACGGGTAAGGATCCTCTACG
pBR4525_F	CCAGAGGCGGCCGCAATGACGGAAGAGCGGACG
pBR4525_R	CACCCGAGGCCCTGCATTCCTAGGAGATGCGGC
pBR4601_R	GACCTGTCGGGTCTCCGAATTCCTAGGAGATGCGGC
pBR4601_R	AAACCAGAGGCGGCCGCAATGGAGCCGGAGGTGGCGG
pAC3634-A3_F	TTTGGCGCGGCCGCAATGGCGGAGCGGCACCGGG
pAC3634-A3_R	ATGCAGATCCTAGGTCACGCACCACGCCCCTGCAC

pAC3636-A2_F	TTTGGCGCGGCCGCAATGGAGCGGCAGCGACTGG
pAC3636-A2_R	GGCTTCTGCCGCGTGACTGGATCCTAGACGTA
pAC3636-A6_F	TTTGGCGCGGCCGCAATGGAGCAGCATCGACTG
pAC3636-A6_R	GGCTTCCGTCGCGTGACTGGATCCTAGACGTAGC
pACentF_F	TTTGGCGCGGCCGCAATGGATATTATGCTGCCAGGT
pACentF_R	GGCAACGGACTTGACTTCACTGGATCCTAGACGTAGCG
pET_NdeI	CATATGGCTGCCGCGCGGCACCAG
pET_HindIII	AAGCTTGCGGCCGCACTCGAGCAC
R_pET3634_FAAL	CGCGAGCGCGACTGGGAGTTCGAACGCCGGCGTGAG
F_pET3634_FAAL	CCGCGCGGCAGCCATATGCTGCGCGTTCCCCCC
R_pET3634A2	TGCAGCGAGATCTTCGAACGCCGGCGTGAG
F_pET3634A2	CCGCGCGGCAGCCATATGCCGCTGTTGCCT
R_pET3634A3	GCACCACGCATCTTCGAACGCCGGCGTGAG
F_pET3634A3	CCGCGCGGCAGCCATATGGCGGAGCGG
R_pET3634A4	CGCGGGTGGCAGCTTAGCTTCGAACGCCGGCGTGAG
F_pET3634A4	CCGCGCGGCAGCCATATGGCCGAGCGGCGCCGCGTG
R_pET3634A5	GCGCAGGTCGTCCCATCTTCGAACGCCGGCGTGAG
F_pET3634A5	CCGCGCGGCAGCCATATGGCGGATCGTCA
R_pET3635AT	CCGTAGAGGCGCGCGTAGTTCAGCATCTTCGAACGCCGGCGTGAG
F_pET3635AT	CCGCGCGGCAGCCATATGATGACGGATGAGCGACATCGCGGTCATC
R_pET3635A7	GAGTGCGGCCGCAAGCTTCCGCCCCACCCGCACCAC
F_pET3635A7	CCGCGCGGCAGCCATATGCCGCTGGCGGCA
R_pET3635A8	TGCGCTTCCATCTTCGAACGCCGGCGTGAG
F_pET3635A8	CCGCGCGGCAGCCATATGGCGGAGCGCCAG
R_pET3636A10	GCCCTCAGAGCTTTCCTCTTCGAACGCCGGCGTGAG
F_pET3636A10	CCGCGCGGCAGCCATATGGAGGAGCGGCACCGCTTG
R_pET3636A11	GAGCGAGAGGCCGACATCTTCGAACGCCGGCGTGAG
F_pET3636A11	CCGCGCGGCAGCCATATGCCGTTGTCGTTCGCGCAG
R_pET3636A12	TTGACCGGGAGTTCGACTTCGAACGCCGGCGTGAG
F_pET3636A12	CCGCGCGGCAGCCATATGAGCTGGGTATTGCCTCC
R_pET3636A13	GCCGCGTGAATCTTCGAACGCCGGCGTGAG
F_pET3636A13	CCGCGCGGCAGCCATATGCAGCAGCGGGTT
R_pET3636A14	GTCAGGCACCTACGGATCTTCGAACGCCGGCGTGAG
F_pET3636A14	CCGCGCGGCAGCCATATGGCAGTGGAGCAGCGCCGG
R_pET3636A15	CCGGGCTTCCGTCGCGTGTTCGAACGCCGGCGTGAG
F_pET3636A15	CCGCGCGGCAGCCATATGGGCAGTGAGGAGCAG
R_pET3636M16	AGGACAGGCTCGCCCACTTCGAACGCCGGCGTGAG
F_pET3636M16	CCGCGCGGCAGCCATATGAGCGATGCGCTACCGCTG
R_pET3636A17	TTGACCGGGAGTTCGACTTCGAACGCCGGCGTGAG
F_pET3636A17	CCGCGCGGCAGCCATATGAGCTGGGTATTGCCTCC
R_pET3636M18	GGCCGAAGACGGCGCACAAGCTTGCGGCCGCACTC
F_pET3636M18	CCGCGCGGCAGCCATATGGGCAGTGAAGAGCAG
R_pET3636A19	GGCACGGCGAGCGAGAAAAAGCTTGCGGCCGCACTC
F_pET3636A19	CCGCGCGGCAGCCATATGGGCAGTGAGGAGCGG
R_pET3636M20	AGCTGGCCTCTCGGCTCAAGCTTGCGGCCGCACTC
F_pET3636M20	CCGCGCGGCAGCCATATGGACAGGGCGCTTCCCCT
pACYC_F	GACGTTTGGCGCGGCCGCAATGGCATTCAG
pACYC_R	GAGGCACAATGATGACCTAGGATCTGCAT
pACYC_F_MXAN3118	ACAGGAAACAGCGTATGACGGATGAGCGAGA
pACYC_R_MXAN3118	AGAGCTGAAGTCGTAGTGACCTAGGATCTGC
F_rpoC_RT	ATGCTGGACGTGATTCCGGTGATT
R_rpoC_RT	CATCAACCGCAACAACCGTCTGAA
F_MXAN3118_RT	GACACGACCGTCTACAAGG

R_MXAN3118_RT	GAGTGCCTGGAGTACATCAAG
pBJ113_*3118* up *F*	GTTCTTGGATCCGCGTCCCGGTCGATGAGC
pBJ113_*3118* up R	GTTCTTGCTAGCTGATTTCTCTCCTGGAACGGGGC
pBJ113_*3118* down F	GTTCTTGCTAGCTCACGGCACCGTCCGTG
pBJ113_*3118* down R	GTTCTTAAGCTTCGCAGGGACGTCAGGAGTAGC
pBJ113_MXAN3634 up F	GTTCTTGGTACCGTGGCCCTCCGCGTCCTTC
pBJ113_MXAN3634 up R	GTTCTTGCTAGCAAGCTATGAGCACCCACATTCCGG
pBJ113_MXAN3636 down F	GTTCTTGCTAGCACGGAGCCGAACGCCATCTC
pBJ113_MXAN3636 down R	GTTCTTAAGCTTTGAAGCGCGTGCTTCCCATTGG

Table 6. Plasmids used in the course of this study.

Plasmid	Description/Purpose	Reference
pBJ113	*M. xanthus* recombination backbone plasmid	(47)
pTOB50	pBJ113 containing 1000 kb upstream and downstream of MXAN_3118	this study
pTOB51	pBJ113 containing 1000 kb upstream of MXAN_3634 and 1000 kb downstream of MXAN_3636	this study
pMAK705/ybdZentF	Deletion of ybdZ and entF	(5)
pACYC duet-1	T7 co-expression vector	(5)
pACYC duet-NO MCS	co-expression empty vector	(5)
pACYC duet-MXAN_3118	co-expression of MXAN_3118	this study
pET28b	T7 expression vector (N-terminal His tag)	(5)
pET28b-FAAL	expression of FAAL	this study
pET28b-A2	expression of A2	this study
pET28b-A3	expression of A3	this study
pET28b-A4	expression of A4	this study
pET28b-A5	expression of A5	this study
pET28b-A6	expression of A6	this study
pET28b-AT	expression of AT	this study
pET28b-A7	expression of A7	this study
pET28b-A8	expression of A8	this study
pET28b-A10	expression of A10	this study
pET28b-A11	expression of A11	this study
pET28b-A12	expression of A12	this study
pET28b-A13	expression of A13	this study
pET28b-A14	expression of A14	this study
pET28b-A15	expression of A15	this study
pET28b-A16	expression of A16	this study
pET28b-A17	expression of A17	this study
pET28b-M18	expression of M18	this study
pET28b-A19	expression of A19	this study
pET28b-M20	expression of M20	this study
pET28b-MXAN3779-M4	expression of MXAN3779-M4	this study
pET28b-MXAN3779-M10	expression of MXAN3779-M10	this study
pET28b-MXAN1291-M2	expression of MXAN1291-M2	this study
pET28b-MXAN1607-M1	expression of MXAN1607-M1	this study
pET28b-MXAN2796-M1	expression of MXAN2796-M1	this study

pET28b-MXAN3618-M1	expression of MXAN3618-M1	this study
pET28b-MXAN3643-M1	expression of MXAN3643-M1	this study
pET28b-MXAN3935-M1	expression of MXAN3935-M1	this study
pET28b-MXAN4000-M2	expression of MXAN4000-M2	this study
pET28b-MXAN4078-M2	expression of MXAN4078-M2	this study
pET28b-MXAN4299-M1	expression of MXAN4299-M1	this study
pET28b-MXAN4402-M1	expression of MXAN4402-M1	this study
pET28b-MXAN4414-M1	expression of MXAN4414-M1	this study
pET28b-MXAN4525-M1	expression of MXAN4525-M1	this study
pET28b-MXAN4601-M2	expression of MXAN4601-M2	this study
pBR α	α fusion protein	(48), R.L. Gourse
pAC λCI	λCI fusion protein	(48), R.L. Gourse
pBR-1291-A2	expression of α-1291-A2	this study
pBR-1607-A1	expression of α-1607-A1	this study
pBR-2796-A1	expression of α-2796-A1	this study
pBR-3618-A1	expression of α-3618-A1	this study
pBR-3643-A1	expression of α-3643-A1	this study
pBR-3634-A3	expression of α-3634-A3	this study
pBR-3635-A8	expression of α-3635-A8	this study
pBR-3636-A11	expression of α-3636-A11	this study
pBR-3636-A15	expression of α-3636-A15	this study
pBR-3779-A4	expression of α-3779-A4	this study
pBR-3779-A10	expression of α-3779-A10	this study
pBR-3935-A1	expression of α-3935-A1	this study
pBR-4000-A2	expression of α-4000-A2	this study
pBR-4078-A2	expression of α-4078-A2	this study
pBR-4299-A1	expression of α-4299-A1	this study
pBR-4402-A1	expression of α-4402-A1	this study
pBR-4414-A1	expression of α-4414-A1	this study
pBR-4525-A1	expression of α-4525-A1	this study
pBR-4601-A2	expression of α-4601-A2	this study
pBR-entF-A1	expression of α-entF-A1	this study
pAC-MXAN3118	expression of λCI-MXAN3118	this study
pAC-*ybdZ*	expression of λCI-YbdZ	this study

Co-overexpression of NRPSs with MXAN_3118 and protein purification. Overexpression vectors containing various NRPS-encoding genes were transformed into *E. coli* BL21(DE3) *ybdZ*::*acc(IV)* cells containing either pACYC-Duet1-NO-MCS or pACYC-Duet1-MXAN_3118, as previously described (5, 21). Overproducing strains were grown in three liters of LB supplemented with kanamycin (50 μg/mL) and chloramphenicol (34 μg/mL). The cells were grown at 28°C with shaking. After reaching an OD_{600} of 0.5 the temperature was shifted to 15°C

for an hour, then cells were induced with 100 μM of isopropyl β-D-1-thiogalactosidase (IPTG) for 15 hours. Cells were harvested by centrifugation and resuspended in His-tag buffer (300 mM NaCl, 20 mM HEPES (pH 7.5), 10% v/v glycerol). Protein purification using nickel-affinity chromatography was performed as previously described (21). Fractions containing the protein of interest based on SDS-PAGE/Coomassie blue staining were pooled and concentrated (Millipore Centripep YM-3K). Concentrated proteins were flash frozen in liquid nitrogen and stored at -80°C until use. NRPS/MLP co-elution was evaluated using 16.8% acrylamide Tris-tricine gels stained with Coomassie blue. All protein concentrations were determined using the BCA™ Protein Assay Kit (Pierce) using bovine serum albumin as a protein standard.

ATP/PP_i exchange assays. ATP/PP_i exchange assays were performed as previously described (2, 21). In summary, each assay (100 μL) contained 75 mM Tris-HCl (pH 7.5 at 25°C), 10 mM $MgCl_2$, 5 mM dithiothreitol, 3.5 mM ATP (pH 7.0), 1 mM $NaPP_i$, 1 mM [^{32}P]PP_i (0.9 Ci/mol PerkinElmer),1 mM amino acid and 2.2-21 μg of protein. To determine substrate specificity, enzymatic activity was first assessed using pools of amino acids. Briefly, the four pools of L-amino acids consisted of pool one (valine, methionine, cysteine, leucine, isoleucine), pool two (proline, alanine, glycine, serine, threonine), pool three (asparagine, aspartate, glutamate, glutamine, phenylalanine) and pool four (tyrosine, tryptophan, histidine, lysine, arginine). The amino acid components of the pool(s) with the highest activity were further tested to identify the amino acid substrate. All assays were performed at 25°C for 30 minutes. The reaction was quenched with a solution containing perchloric acid (3.5% v/v), $NaPP_i$ (100 mM), and activated charcoal (1.6% w/v). The ATP bound to charcoal was separated by centrifugation, and pellets were washed with quench solution and water before being counted in the scintillation counter. For each amino acid/pool the rate of the reaction (pmol product/min/mg protein) was calculated from three independent assays performed in parallel using GraphPad Prism ver.6.0h.

Examination of the expression of MXAN_3118 by Reverse Transcription-PCR (RT-PCR). *M. xanthus* DK1622 cultures were grown at 30°C in CTT (1% w/v casitone, 8 mM $MgSO_4$, 10 mM Tris-HCl [pH 7.6 at 25°C], 1 mM K_2HPO_4-KH_2PO_4) and A1 minimal media (0.5% w/v potassium aspartate, 0.5% w/v sodium pyruvate, 0.5 mg/mL $(NH_4)_2SO_4$, 8 mM $MgSO_4$, 0.125 mg/ml spermidine, 0.1 mg/ml asparagine, 0.1 mg/ml isoleucine, 0.1 mg/ml phenylalanine, 0.1 mg/ml valine, 0.05 mg/ml leucine, 0.01 mg/ml methionine, 1 µg/ml cobalamin, 10 µM $FeCl_3$, 10 µM $CaCl_2$, 1 mM K_2HPO_4-KH_2PO_4 , and 10 mM Tris-HCl [pH 7.6 at 25°C]) (49). Triplicate cultures were grown until stationary phase. OD_{600} was measured every 6-8 hours and 500 µL of cells where harvested at different time points during vegetative growth. RNA extractions were performed following the manufacturer's instructions (RNeasy Mini Kit; Qiagen). All of the RNA samples were also treated with RQ1 DNase (Promega) to remove any residual genomic DNA, followed by ethanol precipitation, and re-suspended in nuclease-free water. RT-PCR was performed using the Access RT-PCR introductory system (Promega) using primers listed in Table 5. The housekeeping gene MXAN_3078 (*rpoC*) was used as a positive control for each time point. The PCR products were resolved using 3% (w/v) agarose gels.

Bacterial two-hybrid (B2H) assays. The *E. coli* reporter strain FW102 F' $placO_L2$-62-*lacZ*, containing a *lacZ* reporter and an F' epitosome (45), had both *ybdZ* and *entF* genes deleted from its genome using the temperature-sensitive plasmid pMAK705-Δ*ybdZentF* vector as previously described (2, 50). Plasmids encoding B2H fusion proteins (Table 6) were constructed by cloning the portion of the NRPS-encoding gene that covered the adenylation domain into pBRα and MXAN_3118 into pACλCI between the *Not*I and *Bam*HI sites in the vectors using the primers listed in Table 5. The FW102 Δ*ybdZentF* F' $placO_L2$-62-*lacZ* was used as our reported strain and transformed with the pairs of pBRα/NRPS A domain and pAClλCI/MXAN_3118 fusion plasmids. Assays were perform as previously described (36, 51, 52), with minor modifications. Briefly, all

strains were grown in three mL of LB containing carbenicillin (100 μg/mL), kanamycin (50 μg/mL) and chloramphenicol (34 μg/mL) (LB/Carb/Kan/Cm) at 30°C overnight in triplicate. Cultures were diluted 1:100 in LB/Carb/Kan/Cm with IPTG (100 μg/mL) and incubated at 30°C with shaking until OD_{595} reached 0.5. Cultures were then incubated on ice for 20 min, and cells were lysed with PopCulture (Millipore) and r-Lysozyme (Fisher Scientific) for 30 min. The lysed culture (30 μL) was mixed with 170 μL Z-buffer containing 0.8 mg/mL 2-nitrophenyl β-D-galactopyranoside and the formation of *o*-nitrophenol was measured every minute for 60 min by following the increased absorbance at 415 nm (Biotrek microplate spectrophotometer). Correction factors and Miller units were calculated as previously described (52).

***M. xanthus* bacterial strains and plasmids.** All strains and plasmids are listed in Table 4 and 5. Sequences of all PCR primers can be found in Table 4. *M. xanthus* DK1622 strains were grown at 32ºC in CYE (1.0% w/v casitone, 0.5% w/v yeast extract, 10 mM 3-[N-morpholino] propanesulfonic acid (pH 7.6 at 25°C) and 8 mM $MgSO_4$), supplemented with 50 μg/mL kanamycin when appropriate. Deletion strains LS3135 and LS3944 were generated through the double recombination method previously described (47) using the vector pBJ113 (53). Briefly, plasmids were created by first generating PCR fragments 1000 bp immediately upstream of the target deletion and 1000 bp immediately downstream. For LS3135, the DNA fragments had 5' *Bam*HI and 3' *Nhe*I sites and 5' *Nhe*I and 3'*Hin*dIII sites, respectively. The fragments were inserted into the appropriate sites of the plasmid vector pBJ113. For LS3944, the upstream fragment had 5' *Kpn*I and 3' *Nhe*I sites instead. Resulting plasmids were then transformed into electrocompetent *M. xanthus* DK1622 cells with selection for kanamycin resistance after integration by a single homologous crossover. A second recombination event was counter selected for galactose sensitivity with CYE containing 1% w/v galactose and resulted in in-frame deletions that were verified by PCR.

***M. xanthus* growth in minimal media.** *M. xanthus* DK1622, ΔMXAN_3634-6 (LS3944) and ΔMXAN_3118 (LS3135) cultures were grown at 30°C in 25 mL A1 Minimal Media (49). Triplicate cultures were grown for a period of two weeks.

Chapter 3

Characterization of an enterobactin-producing nonribosomal peptide synthetase from yeast.

Karla J. Esquilín-Lebrón and Michael G. Thomas.
Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin.

A manuscript describing this work is in preparation.

The authors contributed to the research in the following ways:

Prof. Michael G. Thomas constructed the mutant strains used in this work.

Karla J. Esquilín-Lebrón performed the remainder of the work described.

Abstract

Yeast strains from the *Wickerhamiella/Starmerella* clade acquired the bacterial gene cluster responsible for the NRPS-dependent synthesis of the siderophore enterobactin (ENT). Over 50-million years of evolution, the original bacterial cluster acquired eukaryotic characteristics (e.g. additional transcriptional regulation, poly-A tails, and mono and polycistronic transcripts) and genes redundant or non-essential for siderophore synthesis in yeast were lost. One of the genes for which a homolog is not found in the yeast ENT biosynthetic gene clusters is *ybdZ*, which codes for an MbtH-like protein that is essential for ENT production in *Escherichia coli.* To date, this is the first example of a functional ENT biosynthetic cluster that does not code for a YbdZ homolog. In this Chapter, I summarize the work I performed to characterize the ENT NRPS assembly line from *Candida vaccinii* NRRL Y-17684. Interestingly, I found that the bacterial YbdZ protein does not influence the solubility of or copurify with the EntF homolog from *C.* vaccinia NRRL Y-17684 ($EntF_{vac}$) when heterologously produced in *E. coli*. The kinetics for amino acid activation by $EntF_{vac}$ were determined and the addition of YbdZ did not influence the kinetics of L-Ser activation by the adenylation domain of $EntF_{vac}$, as occurs with the *E. coli* EntF ($EntF_{eco}$). Further analysis of the ENT biosynthetic enzymes determined homologs of EntB and EntE in yeast were able to functionally interact with $EntF_{eco}$ to produce ENT *in vitro* and the corresponding genes complemented *entB* and *entE* mutations in *E. coli*, providing *in vivo* support for the function of these enzymes. Surprisingly, $EntF_{vac}$ was not able to complement the loss of *entF* but was able to functionally interact with $EntB_{vac}$ and $EntE_{vac}$ to produce ENT (preliminary data).

Introduction

Biosynthetic gene clusters (BGCs) encoding the enzymatic machinery to produce nonribosomal peptide natural products commonly found in bacteria and fungi (2). A recent study revealed the first bacterial BGC encoding a nonribosomal peptide synthase (NRPS) assembly line that was horizontally transferred into a group of closely related yeast taxa (54). This is the first example of a horizontal operon transfer (HOT) of a bacterial natural product BGC into yeast. The HOT event was of the well characterized siderophore enterobactin (ENT) NRPS-encoding BGC. Kominek *et al*. hypothesized that the siderophore BGC was transferred from a member of *Enterobactericeae* to an ancestor of the *Wickerhamiella/Starmerella* (W/S) yeast clade, while sharing an insect gut environment (54). The bacterial operon went through a series of genetic changes, including the loss of the MbtH-like protein (MLP)-encoding gene *ybdZ,* that resulted in an operon with both eukaryotic and bacterial features (Fig.1).

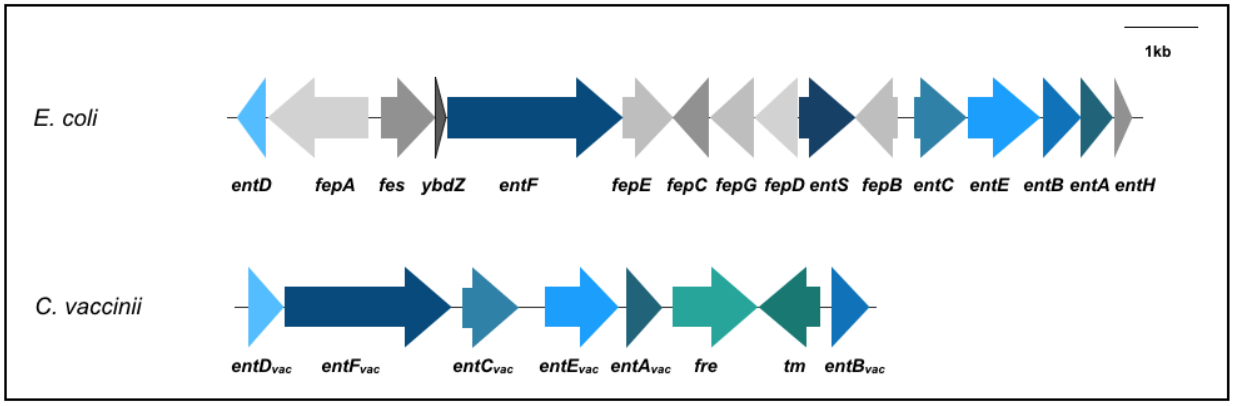


Figure 1. Genetic organization of the ENT biosynthetic cluster in *E. coli* MG1655 and *C. vaccinii* NRRL Y-17684. Similar color identifies homologous genes. The *C. vaccinii* NRRL Y-17684 cluster lost the genes coding for the bacterial outer membrane receptor system (*fepA-G*), the enterobactin esterase (*fes*), thioesterase (*entH*), and exporter (*entS*). The cluster also displays gene rearrangement of *entD* and the integration of the eukaryotic genes for a ferric reductase (*fre*) and a transmembrane protein (*tm*).

I was surprised to find that the gene coding for *ybdZ* was missing in the yeast ENT BGC and that none of the W/S clade members coded for an MLP-homolog in their genomes (C. Hittinger, personal communication). These findings raised the question of whether the yeast ENT system evolved to become MLP-independent. The ENT BGC codes for a two-module NRPS that incorporates 2,3-dihydroxybenzoate (DHB) and L-Serine (L-Ser) three times to assemble the final molecule (Fig. 2) (55). In *E. coli*, $EntF_{eco}$ is the only MLP-dependent module in the NRPS for ENT biosynthesis (5). Previous biochemical studies in our laboratory and others have shown that YbdZ influences EntF protein solubility, kinetic parameters for the activation of L-Ser, and the overall turnover of the NRPS assembly line (5, 6, 8, 18). *In vivo* studies have shown the essentiality of *ybdZ* during growth under iron-limited conditions providing a selection for functional ENT biosynthesis (5, 18). Also, the ENT NRPS system can be reconstituted *in vitro* to test for functional protein-protein interactions and ENT production (12).

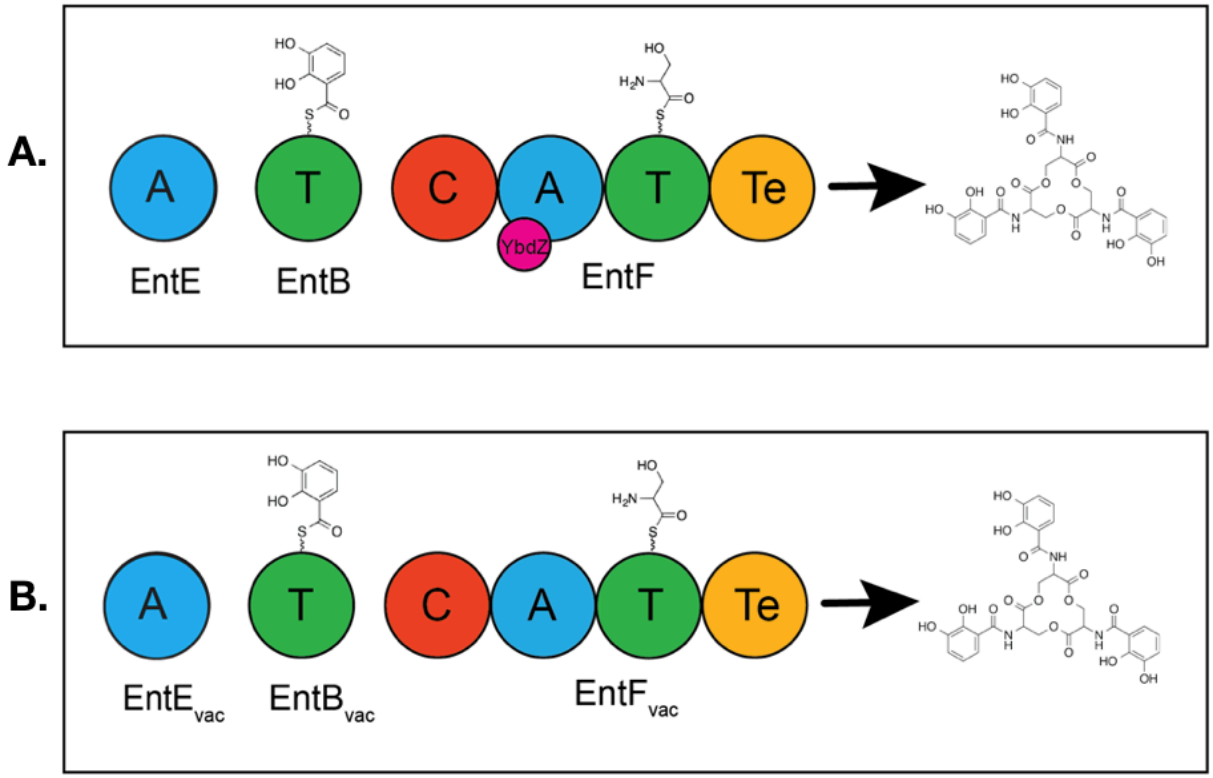


Figure 2. Schematic of the enterobactin NRPS megasynthase from *E. coli* (A) and a proposed scheme for *C. vaccinii* NRRL Y-17684 (B). The NRPS megasynthase is composed of two modules that work iteratively three times to synthesize one molecule of ENT. Domain abbreviations: A, adenylation; T, thiolation; C, condensation; and Te, thioesterase.

Several studies have investigated the specificity MLPs to NRPSs and the mechanisms of these interactions. Structural biology studies identified that YbdZ binds away from the active site of the adenylation (A) domain of $EntF_{eco}$ (29, 30). Surprisingly, the structural studies of YbdZ bound to $EntF_{eco}$ did not detect any conformational changes compared to a structure of $EntF_{eco}$ without YbdZ. Protein sequence analysis showed that there is no specific protein sequence that identifies an MLP-dependent A domain from an independent one, but it was observed that two of the three essential tryptophan residues from YbdZ form a pocket that interacts with an alanine or proline residue from the A domain (6, 26, 27). This observation and further bioinformatic analysis of protein sequences from characterized A domains show that MLP-dependent A domains can also have large non-polar residues; like leucine, isoleucine and valine, in the alanine position (Ala-433 in Slgn1) (27). Even though the ENT system is one of the best studied siderophore systems, we still do not fully understand the role of YbdZ, or other MLPs, in NRPS enzymology (12, 20, 39).

Some of the unanswered questions in the field are to understand what makes an A domain MLP-dependent and how NRPSs function without an MLP partner. The ENT BGCs found in the W/S clade are the first clear example of an NRPS system that evolved to be MLP-independent, providing a system to dissect what changes to the NRPS are needed to become MLP-independent by comparing the *E. coli* and yeast ENT systems. Evidence from Kominek *et al*. shows that this BGC is expressed and required for the production of ENT in yeast (54). My main goal was to determine if the yeast EntF enzymes were MLP-independent and how these enzymes function without the influence of YbdZ. In this Chapter, I characterized the ENT system of *C. vaccinii* NRRL Y-17684, one of the twelve members of the W/S clade, using *in vivo* and *in vitro*

approaches to investigate if this ENT system functions in the same way as *E. coli* ENT and address if YbdZ influences with $EntF_{vac}$ NRPS function.

Results

ENT biosynthetic gene clusters from the Wickerhamiella/Starmerella clade do not contain an MLP-encoding gene. Twelve species of the *W/S* clade were identified to have acquired the ENT biosynthetic gene cluster by HOT from an *Enterobactericeae* strain. Since Kominek *et al.* excluded *ybdZ* in their studies we analyzed the genomes of the W/S clade members using Basic Local Alignment Search Tool (BLAST) to search for homologs of *ybdZ* (31). There were no MLP-encoding genes found in the ENT BGC region nor in the genomes of the W/S clade strains. As far as I am aware, these are the first examples of ENT BGCs that do not include an *ybdZ* homolog. These findings raised the questions about whether the yeast EntF homologs are MLP-independent, if so, how has the NRPS changed to function independent of its MLP partner.

The ENT BGC in *C. vaccinii* NRRL Y-17684 and the other W/S clade members also lost additional genes compared to the ENT cluster from *E. coli* MG1655 (Fig. 1). The *C. vaccinii* NRRL Y-17684 ENT BGC does not include homologs for the genes *fes* and *fepA-G*, *entH*, or *entS*. In *E. coli*, the *fepA-G* genes code for the outer and cytoplasmic membrane transport systems used in bacteria for the recognition and import of ENT into the cell (56). In *S. cerevisiae* cells, there are at least four different membrane uptake systems to acquire siderophores produced by other organisms (57). It has been proposed that the biological redundancy of other transport systems in yeast could explain the loss of these genes (54). Since yeast cells are well known to obtain their essential iron through the reduction of ferric ions on the cell surface or the acquisition of cognate and non-cognate siderophores, and iron complexes from the environment there might not be a

requirement for an ENT specific esterase (*entS*) coded in the BGC (58, 59). The proofreading thioesterase EntH is also missing from the yeast ENT cluster. We (18) and others (12) have shown that this enzyme is not essential for *in vitro* siderophore production. Only under iron-limited conditions supplemented with salicylate, an inhibitor of EntB substrate, was *entH* essential *in vivo* (60). Because the EntH proofreading mechanism is only conditionally essential for ENT biosynthesis in *E. coli*, its loss in the yeast operon might be expected not to affect ENT production. I propose that the ENT cluster in the W/S clade will function as the ENT system from *E. coli*, and that the missing components will be functionally replaced by redundant genes found in the yeast genome.

Gene	Proposed function	Identity (%)[a]	Size (aa)
entD$_{vac}$	Phosphopantetheinyl transferase	40%	256
entF$_{vac}$	NRPS module	48%	1220
entC$_{vac}$	Isochorismate synthase	47%	410
entE$_{vac}$	2,3-dihydroxybenzoate AMP ligase	63%	544
entA$_{vac}$	2,3-dihydroxybenzoate dehydrogenase	67%	253
entB$_{vac}$	2,3-dihydroxybenzoate synthase	58%	282

Table 1. Proposed function of open reading frames in the ENT operon found in *C. vaccinii* NRRL Y-17684. [a] Percent identity determined using the homolog ENT biosynthetic cluster proteins from the *E. coli*.

MLP binding site region is not conserved between the yeast EntF proteins. Since EntF$_{vac}$ functions *in vivo* in the absence of YbdZ or any MLP homolog, I hypothesized that EntF$_{vac}$ evolved to be MLP independent. Multiple protein alignment of the 12 W/S clade EntF proteins with EntF$_{eco}$ were performed using Clustal W to identify conserved and variable regions (41). I focused on the A domain alignment results to highlight the identical (orange), similar (light orange) and variable (green) residues on the structure of EntF$_{eco}$/YbdZ (PDB 5JA1) to generate Figure 3.

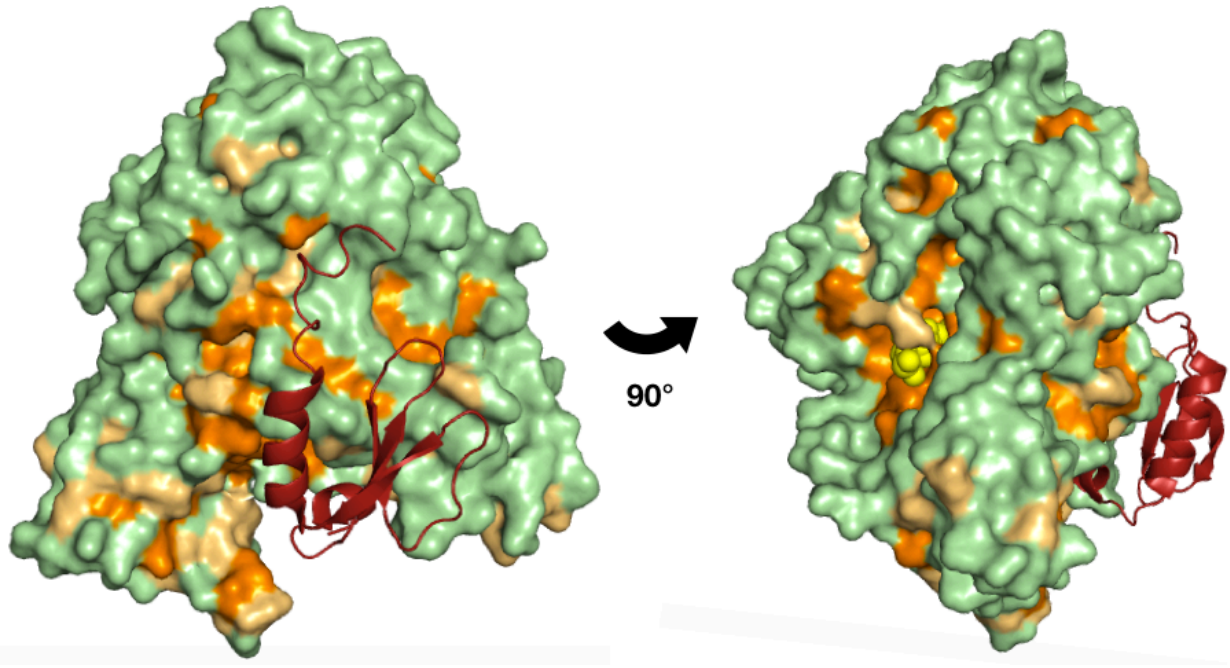


Figure 3. Surface depiction of EntF and cartoon depiction of YbdZ (red). The results of multiple protein alignment of $EntF_{eco}$ with the twelve yeast EntF sequences from the W/S clade are used to highlight the identical residues in orange, conserved and similar in light orange, and variable in green. The inhibitor serine adenosine vinylsulfonamide, in yellow spheres, was used to highlight the substrate. (PDB 5JA1)

I observed that the majority of the residues on the protein surface of the yeast EntF A domains differ from those of $EntF_{eco}$. Structural analyses have shown that there is a specific region in A domains where the MLP binds (26, 27, 29, 30). Therefore, I targeted this region of the $EntF_{vac}$ A domain to analyze if evolution of this region allowed $EntF_{vac}$ to lose the ability to interact with YbdZ (27) (Fig. 4). Of the 23 A domain residues (green dots in Figure 5) previously identified to be involved in MLP-NRPS interactions, only 7 were conserved or similar amongst the group of EntF proteins (28). In Figure 4, I highlight the residues from the A domain-MLP interface that differ from $EntF_{eco}$ and $EntF_{vac}$. The alanine residue from the A domain identified to fit in the pocket formed by two tryptophan residues in the MLP was not conserved in 11 of the 12 yeast EntF homologs. Only the *C. pararugosa* EntF sequence shared this alanine residue, while the other 11 yeast EntF

proteins have a glutamate, lysine, arginine, threonine, or valine residue in this position. Previous research has shown that several MLP-independent domains can have glutamate in the alanine position. Also, biochemical characterization of the MLP-dependent Slgn1 showed that a Ala-433-Glu substitution resulted in an inactive NRPS (26). These findings suggest that the glutamate residue in $EntF_{vac}$, does not fit in the YbdZ tryptophan pocket. From the protein sequence alignment of $EntF_{eco}$ and $EntF_{vac}$ (Fig. 5). I conclude that there is no specific region in the A domain or EntF specific sequence that can be identified to confer the changes needed for EntF to be MLP-independent.

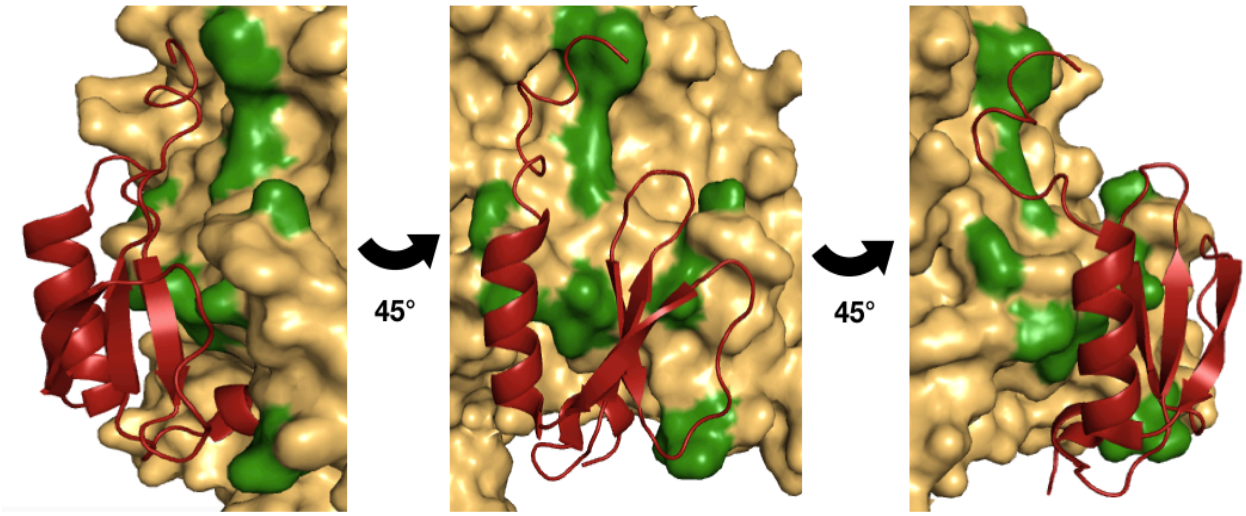
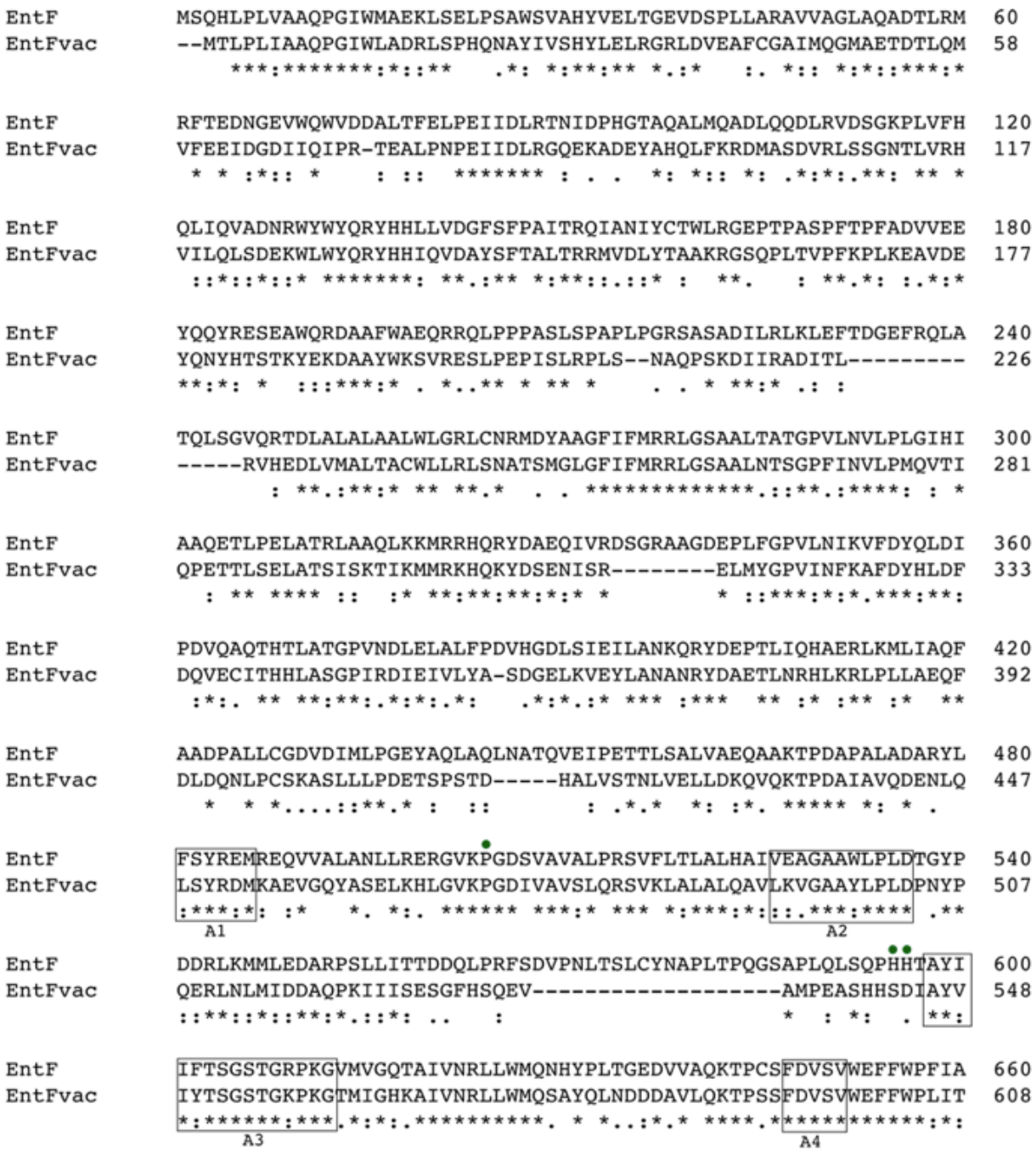


Figure 4. Surface depiction of $EntF_{eco}$ (light orange) and cartoon depiction of YbdZ (red). Residues in the MLP surface region that differ between $EntF_{eco}$ and $EntF_{vac}$ are highlighted in green. (PDB 5JA1).


EntF MSQHLPLVAAQPGIWMAEKLSELPSAWSVAHYVELTGEVDSPLLARAVVAGLAQADTLRM 60
EntFvac --MTLPLIAAQPGIWLADRLSPHQNAYIVSHYLELRGRLDVEAFCGAIMQGMAETDTLQM 58
RFTEDNGEVWQWVDDALTFELPEIIDLRTNIDPHGTAQALMQADLQQDLRVDSGKPLVFH 120
VFEEIDGDIIQIPR-TEALPNPEIIDLRGQEKADEYAHQLFKRDMASDVRLSSGNTLVRH 117
QLIQVADNRWYWYQRYHHLLVDGFSFPAITRQIANIYCTWLRGEPTPASPFTPFADVVEE 180
VILQLSDEKWLWYQRYHHIQVDAYSFTALTRRMVDLYTAAKRGSQPLTVPFKPLKEAVDE 177
YQQYRESEAWQRDAAFWAEQRRQLPPPASLSPAPLPGRSASADILRLKLEFTDGEFRQLA 240
YQNYHTSTKYEKDAAYWKSVRESLPEPISLRPLS--NAQPSKDIIRADITL--------- 226
TQLSGVQRTDLALALAALWLGRLCNRMDYAAGFIFMRRLGSAALTATGPVLNVLPLGIHI 300
-----RVHEDLVMALTACWLLRLSNATSMGLGFIFMRRLGSAALNTSGPFINVLPMQVTI 281
AAQETLPELATRLAAQLKKMRRHQRYDAEQIVRDSGRAAGDEPLFGPVLNIKVFDYQLDI 360
QPETTLSELATSISKTIKMMRKHQKYDSENISR--------ELMYGPVINFKAFDYHLDF 333
PDVQAQTHTLATGPVNDLELALFPDVHGDLSIEILANKQRYDEPTLIQHAERLKMLIAQF 420
DQVECITHHLASGPIRDIEIVLYA-SDGELKVEYLANANRYDAETLNRHLKRLPLLAEQF 392
AADPALLCGDVDIMLPGEYAQLAQLNATQVEIPETTLSALVAEQAAKTPDAPALADARYL 480
DLDQNLPCSKASLLLPDETSPSTD-----HALVSTNLVELLDKQVQKTPDAIAVQDENLQ 447
FSYREMREQVVALANLLRERGVKPGDSVAVALPRSVFLTLALHAIVEAGAAWLPLDTGYP 540
LSYRDMKAEVGQYASELKHLGVKPGDIVAVSLQRSVKLALALQAVLKVGAAYLPLDPNYP 507
A1
A2
DDRLKMMLEDARPSLLITTDDQLPRFSDVPNLTSLCYNAPLTPQGSAPLQLSQPHHTAYI 600
QERLNLMIDDAQPKIIISESGFHSQEV------------------AMPEASHHSDIAYV 548
IFTSGSTGRPKGVMVGQTAIVNRLLWMQNHYPLTGEDVVAQKTPCSFDVSVWEFFWPFIA 660
IYTSGSTGKPKGTMIGHKAIVNRLLWMQSAYQLNDDDAVLQKTPSSFDVSVWEFFWPLIT 608
A3
A4

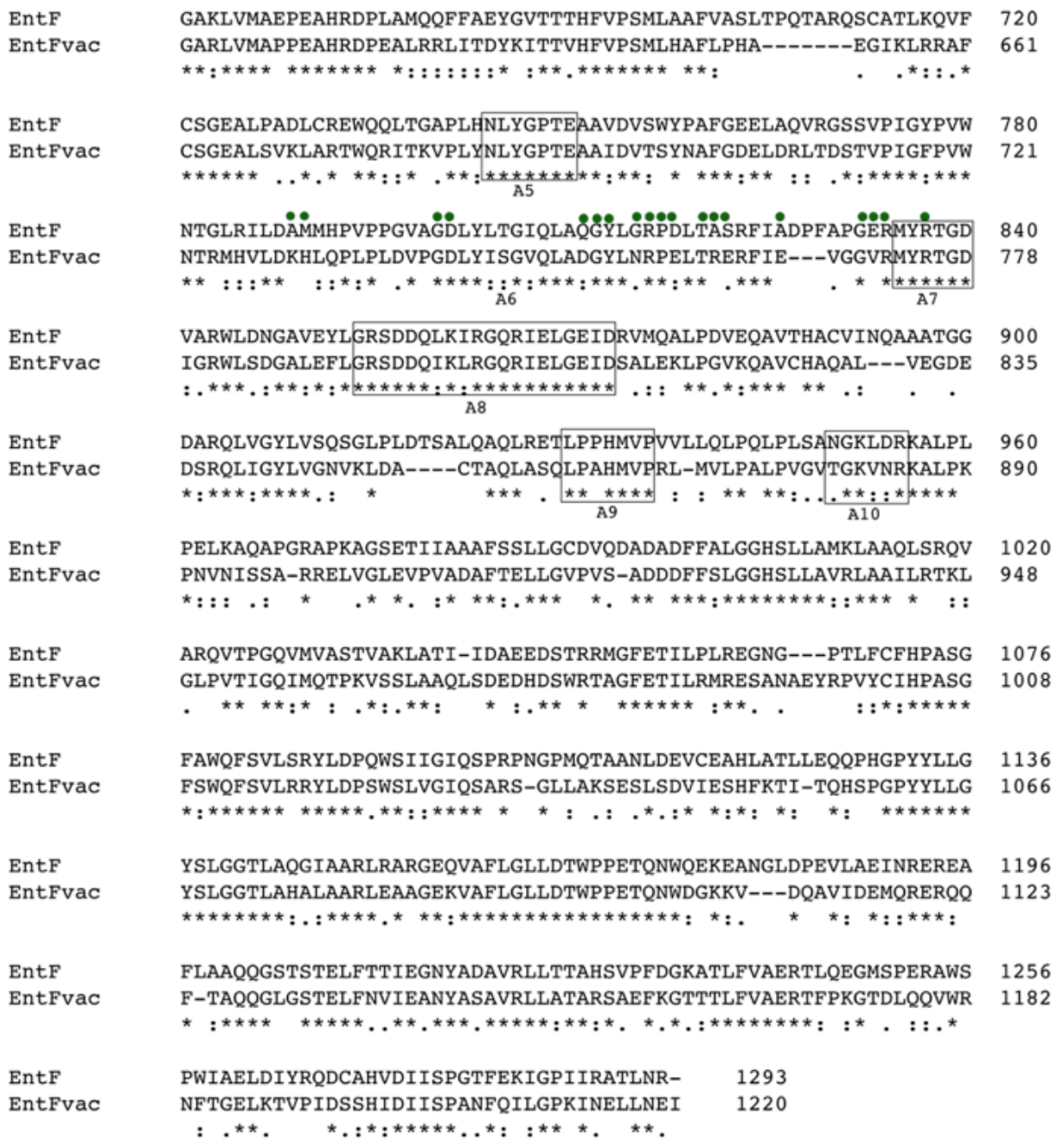


Figure 5. Multiple protein alignment of the sequences of $EntF_{eco}$ (ADB98044.1) and $EntF_{vac}$. The alignment was performed using Clustal Omega (1.2.1) (EMBL-EBI). The identical amino acids between sequences are indicated by an asterisk (*), conserved residues with a colon (:) and similar residues with a period (.). The residues corresponding to the conserved adenylation domain motifs A1-A10 are identified with a solid box. The residues from the MLP-NRPS region are identified with a green circles.(28)

Heterologous co-expression of EntFvac with untagged YbdZ in E. coli does not influence protein solubility or result in EntFvac/YbdZ co-purification. Our lab and others have shown that MLP-dependent A domains often co-purify with their cognate and noncognate MLPs (5–7, 21, 23). I was interested in investigating whether $EntF_{vac}$ had any interaction with YbdZ, therefore I probed whether $EntF_{vac}$ co-purifies with YbdZ and if YbdZ influences NRPS solubility. Co-expression of His-tagged $EntF_{vac}$ in the presence or absence of untagged YbdZ did not influence $EntF_{vac}$ solubility (Figure 6A). Even though YbdZ was not detected in purified yeast $EntF_{vac}$ samples using Coomassie-blue staining of the Tris-tricine gel, I opted to test for its presence using anti-YbdZ polyclonal antibody (39). To confirm the absence of YbdZ, I used 10 µg purified YbdZ as a control. This is a concentration of YbdZ that is below the detection limit of Coomassie blue-staining but detectable by immunoblotting. In Figure 6B, the immunoblot shows that YbdZ was not detected in the $EntF_{vac}$/EV or $EntF_{vac}$/YbdZ samples. These findings show that $EntF_{vac}$ does not copurify with YbdZ or influence its solubility as is the case for $EntF_{eco}$ and other MLP-dependent NRPSs.

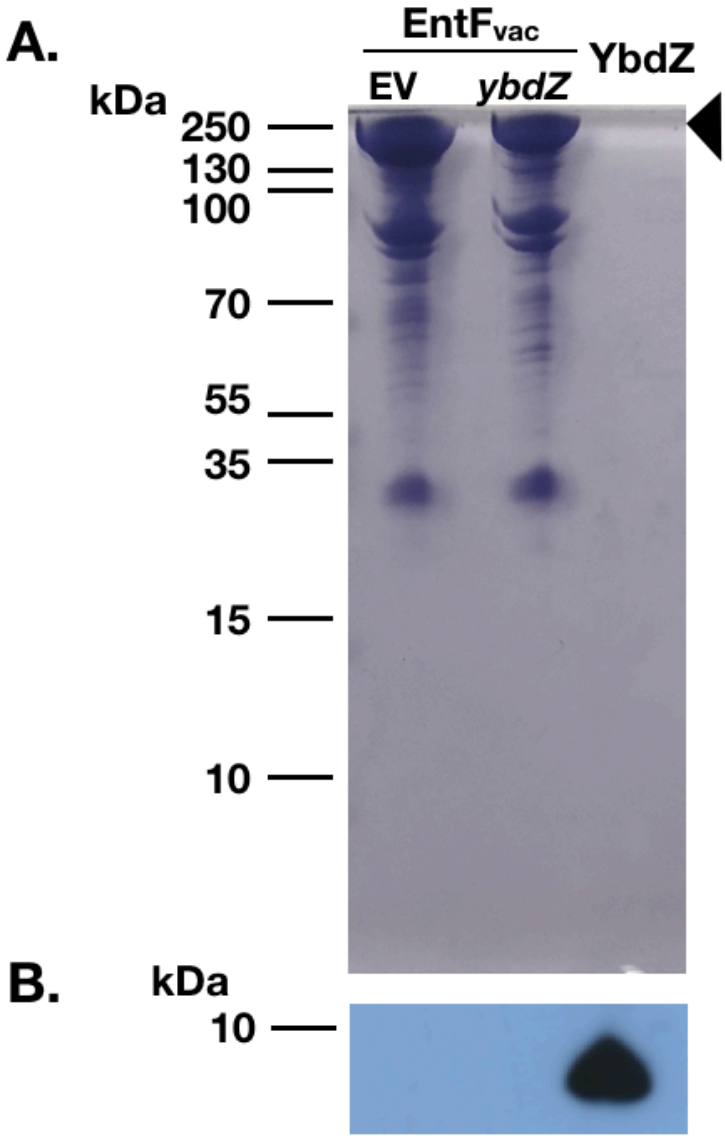


Figure 6. $EntF_{vac}$ does not copurify with YbdZ. Seventeen µg of purified $EntF_{vac}$/EV and $EntF_{vac}$/YbdZ were loaded in 16.8% Tris-tricine gel. Purified YbdZ was used as positive control. (A) Gel stained with Coomassie blue stain. (B) Immunoblot to detect the presence of YbdZ.

Co-production of EntFvac with YbdZ has a negative impact on NRPS function. Interestingly, co-production of $EntF_{vac}$ with YbdZ resulted in two populations of $EntF_{vac}$ after elution from a size exclusion chromatography column. Protein analysis using a Tris-tricine gel/Coommassie-blue staining (data not shown) revealed that both samples are approximately the same molecular

weight, and immunoblotting analysis showed that YbdZ did not co-purify with either of the fractions. In comparison, when $EntF_{vac}$ was overproduced in the absence of YbdZ and then purified, only a single population of $EntF_{vac}$ eluted from size-exclusion chromatography (Fig.7).

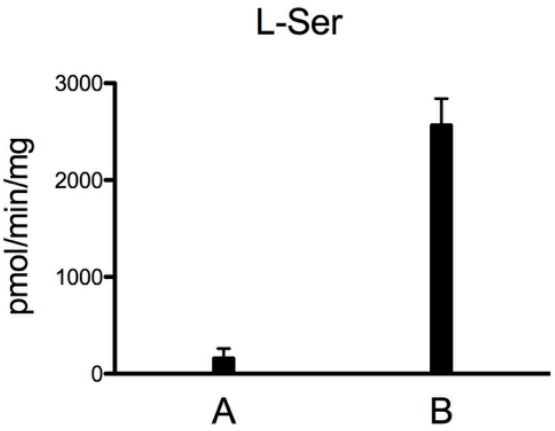


Fig 7. L-Ser activation of $EntF_{vac}$/YbdZ populations A and B. Values are shown as the mean of triplicate assays for each enzyme, error bars represent the standard deviation from three independent assays.

To compare all $EntF_{vac}$ populations I used the ATP/PP_i exchange assay to test how well these proteins activated L-Ser. First, $EntF_{vac}$ overproduced and purified in the absence of YbdZ ($EntF_{vac}$/EV: EV indicates $EntF_{vac}$ overproduction in the presence of an empty vector lacking *ybdZ*) activated L-Ser but had a significantly lower affinity for L-Ser than observed for $EntF_{eco}$ (Table 2). Analysis of the two population of $EntF_{vac}$ that were observed when the NRPS was co-overproduced with YbdZ ($EntF_{vac}$/YbdZ) determined that the population eluting earlier from the size-exclusion chromatography column was inactive. In contrast, the second population activated L-Ser and had an improved affinity for L-Ser compared to the $EntF_{vac}$/EV (Fig.7 and Table 2). A comparison of the V_{max}/K_m values for the two active forms of $EntF_{vac}$ determined the $EntF_{vac}$/EV had a 3-fold higher activity compared to $EntF_{vac}$/YbdZ. In summary, co-expression of $EntF_{vac}$ with YbdZ has a negative impact on NRPS function.

	Substrate	K_m [a] (mM)	V_{max} [a] (pmol/min/mg protein)	V_{max}/K_m [a] (pmol/min/mM)
$EntF_{vac}$/EV	L-Ser	14.6 ± 3.2	9768 ± 1200	6690
$EntF_{vac}$/YbdZ	L-Ser	5.4 ± 2.1	11730 ± 220	2172
$EntF_{eco}$ [b]	L-Ser	1.03 ± 0.17	373 ± 18	
$EntF_{eco}$/YbdZ [b]	L-Ser	0.15 ± 0.06	401 ± 58	

Table 2. Kinetic parameters of ENT enzymes. [a] Values and standard errors are shown for assays run in triplicates. [b] Values calculated by Schomer & Thomas (2018) (18).

Addition of YbdZ does not influence $EntF_{vac}$ kinetic parameters. Previous work from our lab showed that the addition of YbdZ to $EntF_{eco}$, purified in the absence of the MLP, can restore the activity of the enzyme to the levels of copurified $EntF_{eco}$/YbdZ (18). I hypothesized that if YbdZ has a negative impact on $EntF_{vac}$ enzymatic activity I might also be able to detect it by adding exogenous purified YbdZ. I added increasing concentrations of YbdZ (0, 0.25, 2.5, 5 and 10 μM) to both $EntF_{vac}$/EV and $EntF_{vac}$/YbdZ and analyzed L-Ser activation with ATP/PP_i exchange assays (Fig. 8).

]

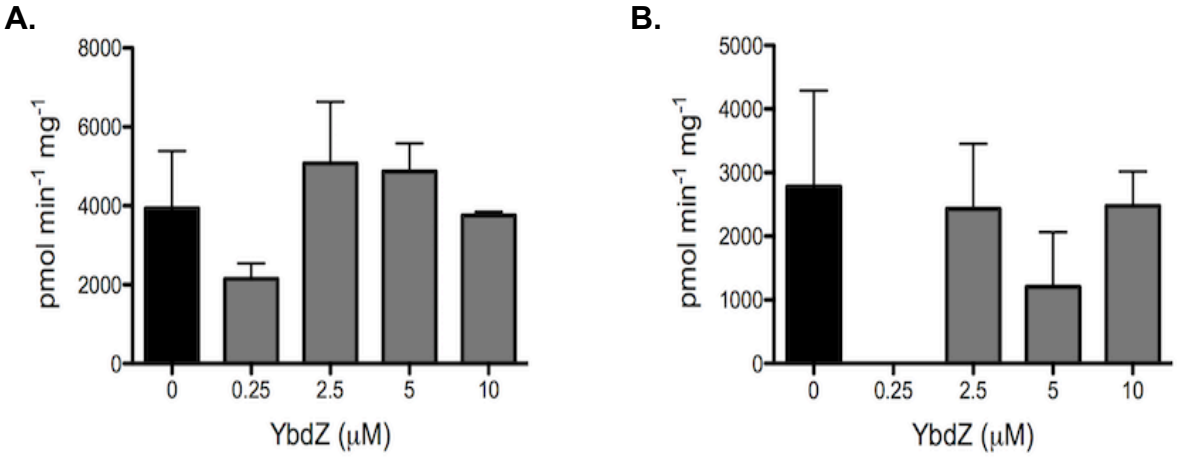


Figure 8. Enzyme activity for L-Ser activation of $EntF_{vac}$/EV (A) and $EntF_{vac}$/YbdZ (B) with varying concentrations of exogenously purified YbdZ added. Values are shown as the mean of triplicate assays for each YbdZ concentration tested and error bars show standard deviation from three independent assays. Assays were performed in the linear range of enzyme concentration and less than 10% substrate-to-product conversion.

There was no statistically significant impact on L-Ser activation of $EntF_{vac}$/EV or $EntF_{vac}$/YbdZ when YbdZ was added in concentrations up to 10 μM. These results show that addition of YbdZ did not influence $EntF_{vac}$ activity, suggesting that the impact of YbdZ on $EntF_{vac}$ activity must occur during heterologous expression and cannot be reconstituted *in vitro*.

Co-expression of entFvac with sfp does not result in functional complementation of MG1655 ΔentF for growth in iron limited media (ILM). I tested if $entF_{vac}$ was able to functionally replace the loss of $entF_{eco}$ in MG1655 Δ*entF* cells in ILM. MG1655 Δ*entF* cells were transformed with pET28b empty vector (EV), pET28b containing $entF_{vac}$ or $entF_{eco}$, and these strains were grown in ILM to assess functional replacement of $EntF_{vac}$ for $EntF_{eco}$. After 72 hours, $entF_{vac}$ did not complemented the growth of MG1655 Δ*entF* in ILM. I hypothesized that one reason why $EntF_{vac}$ was not able to complement the Δ*entF* mutation was because the phosphopantetheinyl transferase $EntD_{eco}$ was unable to post-translationally modify $EntF_{vac}$. Sfp, expressed from the

pSU20 vector, was used to test if a broad substrate specificity phosphopantetheinyl transferase could modify $EntF_{vac}$ to enable growth in ILM (40). I transformed pSU20 empty vector (EV) or pSU20 containing *sfp* into the MG1655 Δ*entF* strains containing the previously described *entF*-expression plasmids. MG1655 Δ*entF* cells containing the pair of pET28b- and pSU20-based plasmids were grown in ILM supplemented with 100 μM IPTG, to induce $entF_{vac}$ expression. All strains grew in M9 and M9+IPTG, but only strains with wild-type $entF_{eco}$ were able to grow in ILM (Fig. 9) and produce ENT (Fig. 10). These results show that neither induction of $entF_{vac}$ nor coexpression with *sfp* results in functional $EntF_{vac}$ able to complement MG1655 Δ*entF* cells for ENT production. Control strains, MG1655 Δ*entF* EV/EV and MG1655 Δ*entF* EV/*sfp*, only grew in M9 media and were not able to produce ENT (Fig. 9 and 10).

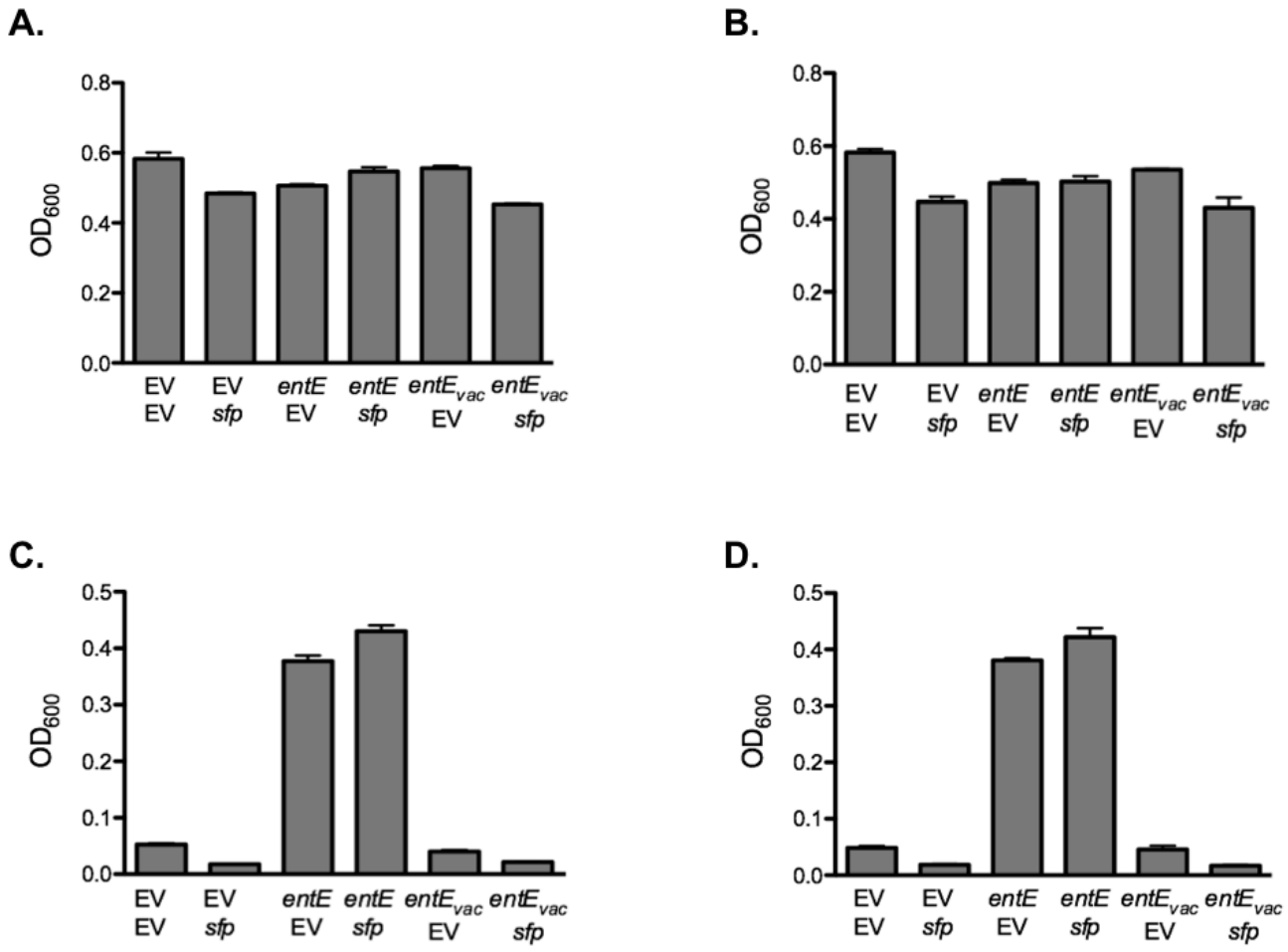


Figure 9. $EntF_{vac}$ does not complement MG1655 Δ*entF* for growth under iron-limited conditions. Data represents end-point OD_{600nm} of triplicate cultures grown in iron-limited media; (A) M9 (B) M9+IPTG, (C) M9+EDDHA and (D) M9+EDDHA+IPTG after 48 hours. Error bars show standard deviation from three independent cultures. The iron chelator EDDHA was used to reduce iron availability in the media. IPTG was used to induce expression of *entF* and $entF_{vac}$. MG1655 Δ*entF* was co-transformed with pET28b plasmids containing empty vector (EV), *entF*, or $entF_{vac}$, and pSU20 plasmids containing empty vector (EV) or *sfp*, as indicated.

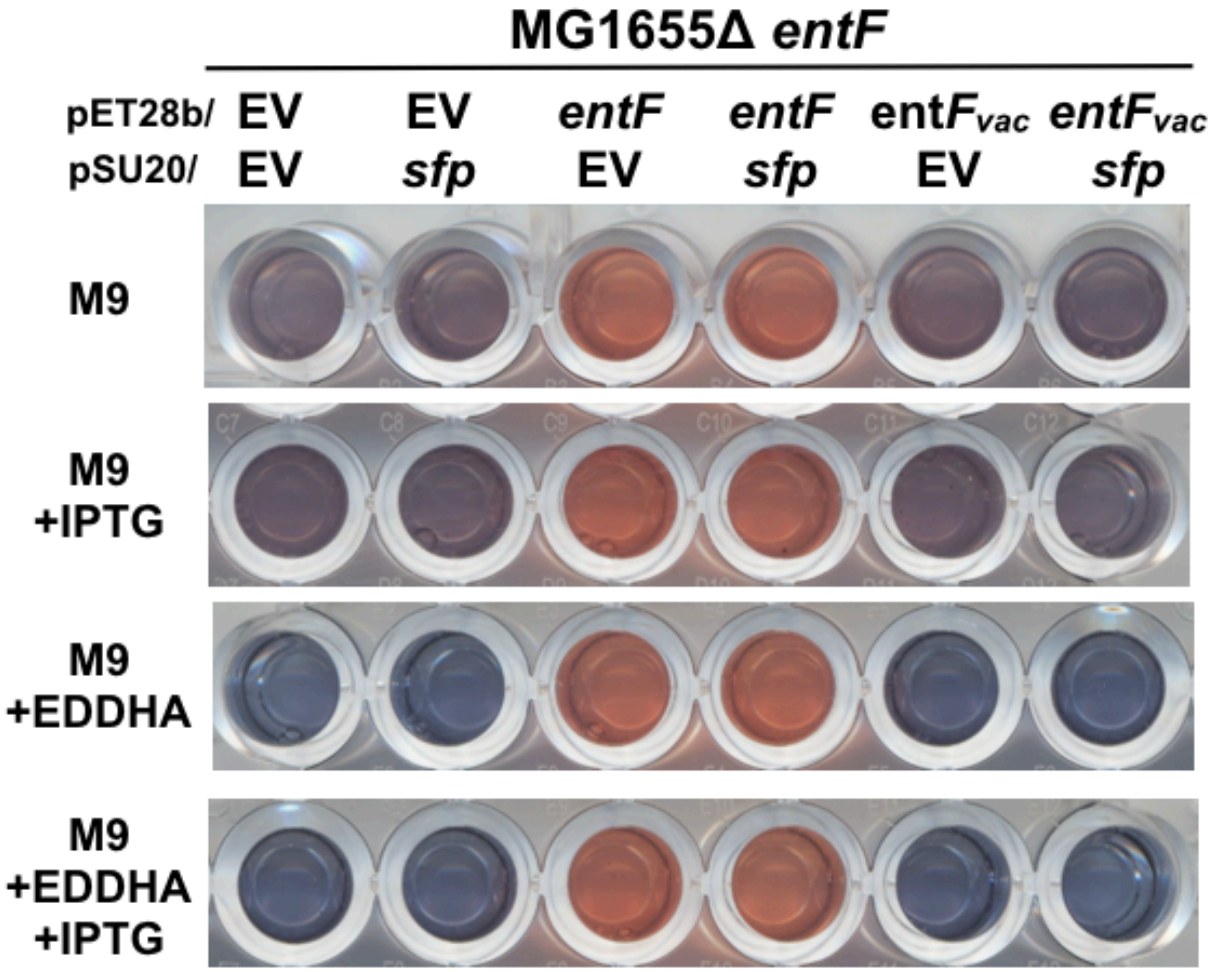


Figure 10. $EntF_{vac}$ does not complement MG1655 Δ*entF* for siderophore production. CAS assay for siderophore production. Assays shown are representative of triplicate cultures. Positive CAS detection of siderophore is considered when is changed to orange.

Characterization of $EntE_{vac}$ kinetic parameters. Since one of our goals is to determine if YbdZ influences ENT production *in vivo* and *in vitro* we set out to purify and characterize the other NRPS enzymes in ENT. EntE is a MLP-independent A domain responsible for the activation and loading of DHB molecule in the ENT megasynthase and is essential for ENT biosynthesis (5, 12). I compared $EntE_{vac}$ protein sequence with $EntE_{eco}$ and found that the enzymes are very well conserved and have a 63% protein identity (Fig.11). Preliminary results show that $EntE_{vac}$

activates DHB as a substrate. Ongoing experiments are being done to characterize the kinetics parameters, I hypothesized that $EntE_{vac}$ has similar kinetic parameters as $EntE_{eco}$ (61).

```
EntE        MSIPFTRWPEEFARRYREKGYWQDLPLTDILTRHAASDSIAVIDGERQLSYRELNQAADN  60
EntEvac     MSIEFTRWPEALVAKYRAKGYWIDEPLCDIISRQSKNTNIAVVDGDRQISYSELDRRSSL  60
            *** ****** :. :** **** * ** **::*:: . .***:**:**:** **:: :.

EntE        LACSLRRQGIKPGETALVQLGNVAELYITFFALLKLGVAPVLALFSHQRSELNAYASQIE  120
EntEvac     LASNLHSQGIRRGDTALVHLGNTAELYVVYFGLLKLGVILVNALFSHRQLELDSYTRQIQ  120
            **..*: ***: *:****:***.****:.:*.******  * *****:: **::*: **:

EntE        PALLIADRQHALFSGDDFLNTFVTEHSSIRVVQLLNDSGEHNLQDAINHPAEDFTATPSP  180
EntEvac     PKLLIGDRLNPLFKDDQYLSHLRTLSPELKVAVFNNDNLAKLSQ---EGSIVEAALTPTP  177
            * ***.** : **..*::*. : *   .::*. : **.  :  *   :    : : **:*

EntE        ADEVAYFQLSGGTTGTPKLIPRTHNDYYYSVRRSVEICQFTQQTRYLCAIPAAHNYAMSS  240
EntEvac     ADEVAFFNLSGGSTGTPKLIPRTHNDYYYSVRRSDEICGFDESTRYLCAIPAAHNFPMSS  237
            *****:*:****:********************** *** * :.************: ***

EntE        PGSLGVFLAGGTVVLAADPSATLCFPLIEKHQVNVTALVPPAVSLWLQALIEGES-----  295
EntEvac     PGALGVFYAGGQVVLAPDPSASSCFPLIAKHKVNVAALVPPAASLWLQEVLSGREVAGTS  297
            **:**** *** **** ****: ***** **:***:******.***** ::.*..

EntE        -RAQLASLKLLQVGGARLSATLAARIPAEIGCQLQQVFGMAEGLVNYTRLDDSAEKIIHT  354
EntEvac     PKSALESLKLLQVGGARLSPTLAARLVEHLNVKLQQVFGMAEGLVCYTRLDDDEESILTT  357
             :: * ************* *****:  .:. :************ ******. *.*: *

EntE        QGYPMCPDDEVWVADAEGNPLPQGEVGRLMTRGPYTFRGYYKSPQHNASAFDANGFYCSG  414
EntEvac     QGLPMCEDDEVWAADLEGKRVGPGEVGVLMTQGPYTFRGYYKASEHNAKSFDANGFYCSG  417
            ** *** *****.** **: :  **** ***:**********: :***.:**********

EntE        DLISIDPEGYITVQGREKDQINRGGEKIAAEEIENLLLRHPAVIYAALVSMEDELMGEKS  474
EntEvac     DLIAIAEDGRITIHGRQKDQINRGGEKISTEEIEHLLLRHPQVVEAGLVAMDDELMGEKS  477
            ***:*  :* **::**:**********::****:****** *: *.**:*:********

EntE        CAYLVVKEPLRAVQVRRFLREQGIAEFKLPDRVECVDSLPLTAVGKVDKKQLRQWLASRA  534
EntEvac     VAFIMTTSPVKPVQLRKFLREQGVAEYKLPDKFTILKEMPLTAVGKVDKKELRILLKQQE  537
             *:::...*:: **:*:******:**:****:.  :..:***********:**  * .:

EntE        SA-----     536
EntEvac     VTAQAAA     544
             :
```

Figure 11. Multiple protein alignment of the sequences of $EntE_{eco}$ (ADB98052.1) and $EntE_{vac}$. The alignment was performed using Clustal Omega (1.2.1) (EMBL-EBI). The identical amino acids between sequences are indicated by an asterisk (*), conserved residues with a colon (:) and similar residues with a period (.).

$EntE_{vac}$ can complement MG1655 ΔentE growth under ILM. I tested if *$entE_{vac}$* was able to complement the loss of *$entE_{eco}$* from MG1655 Δ*entE* cells under iron-limited conditions. Strains of MG1655 Δ*entE* cells containing a plasmid for *$entE_{vac}$, $entE_{eco}$,* or empty vector (EV) were grown in iron-limited media (ILM) in triplicate. ILM is defined as M9 minimal media (Teknova) supplemented with 2 μM EDDHA. Under these conditions, MG1655 Δ*entE*/EV is not able to grow or produce ENT under ILM. End-point growth (OD_{600nm}) and a liquid CAS assay were used to determine siderophore production. $EntE_{vac}$ was able to functionally complement MG1655 Δ*entE* for ENT production *in vivo.* In Figure 12, I observed that the levels of growth and ENT production are similar between strains complemented by *$entE_{eco}$* or *$entE_{vac}$*.

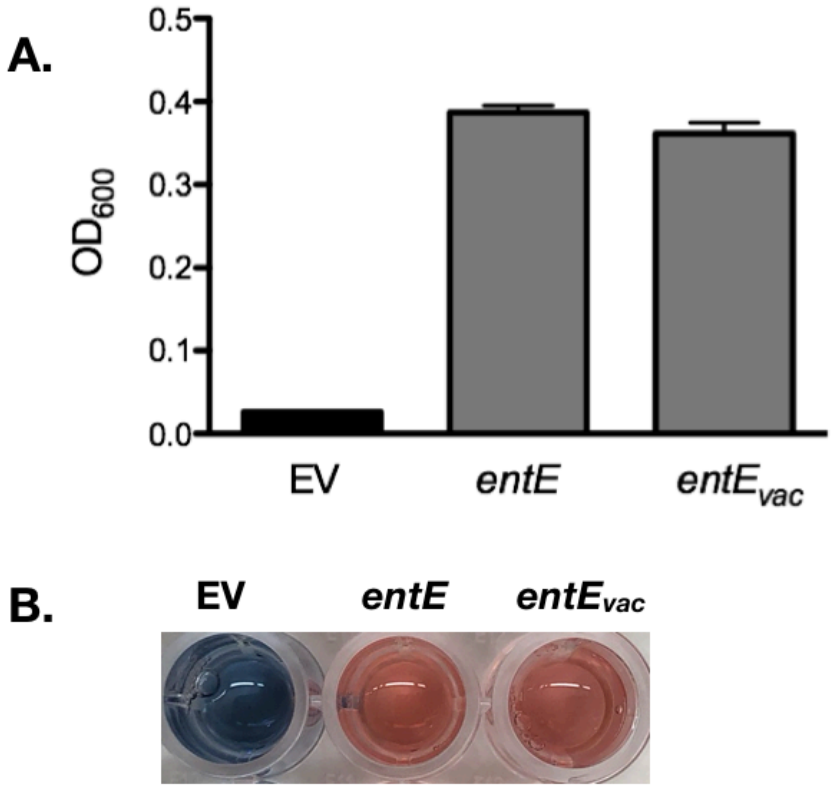


Figure 12. $EntE_{vac}$ complements MG1655 Δ*entE* under iron-limited conditions. (A) Data represents end-point OD_{600nm} of triplicate cultures grown in iron-limited media (M9+2 μM EDDHA) after 72 hours. Error bars show standard deviation from three independent cultures. MG1655

Δ*entE* complemented with empty vector (EV), *entE*$_{eco}$, or *entE*$_{vac}$. (B) CAS assay for siderophore production. Assays shown are representative of triplicate cultures.

EntB*$_{vac}$ *is predicted to be a functional T domain. EntB is the thiolation domain of the first module in ENT biosynthesis (Fig. 2). EntB$_{vac}$ protein sequence share 58% protein similarity with the EntB$_{eco}$ (Fig.13). I tested if EntB$_{vac}$ could be phosphopantothenylated by Sfp into its active form, using a transacylase assay with ^{14}C-malonyl CoA as substrate. Our preliminary results showed that EntB$_{vac}$ was activated *in vivo* when overproduced with Sfp in the BAP1 cells (data not shown). These findings suggested that EntB$_{vac}$ was in its active form for *in vitro* ENT reconstitution assays. Ongoing efforts in the laboratory are focus on deleting *entB*$_{eco}$ from MG1655 cells to test *entB*$_{vac}$ complementation.

```
EntB        MAIPKLQAYALPES-HDIPQNKVDWAFEPQRAALLIHDMQDYFVSFWGENCPMMEQVIAN  59
EntBvac     -MIPQLKSYSLPSSAKELPENKVNWAIDPTRAALLIHDMQEYFLNYWGQDSELVAKLVQN  59
              **:*::*:**.* :::*:***:**::* **********:**:.:**::. :: ::: *

EntB        IAALRDYCKQHNIPVYYTAQPKEQSDEDRALLNDMWGPGLTRSPEQQKVVDRLTPDADDT  119
EntBvac     IKVLKALCKSQGIPVFYTAQPVNQRPEDRALLNDMWGAGIDKRPELRNIVSELSPEADDR  119
            * .*:  **.:.***:***** :*  ********** *: : ** :::*..*:*:***

EntB        VLVKWRYSAFHRSPLEQMLKESGRNQLIITGVYAHIGCMTTATDAFMRDIKPFMVADALA  179
EntBvac     QLVKWRYSAFARSSFENDLKELKRDQLIICGVYAHIGCLTTATEAFMKDIQPFMVADALA  179
             ********* ** :*: ***  *:**** ********:****:***:**:*********

EntB        DFSRDEHLMSLKYVAGRSGRVVMTEELLPAPIPASKAALREVILPLLDESDEPFDDDNLI  239
EntBvac     DFSRSQHDITLQFAAECSSRVELVSGILRQIVEHD--VRKTIANALEDDEDWPEDDENLI  237
            ****.:* ::*::.*  *.** :.. :*   :  .  . : :   * *:.* * **:***

EntB        DYGLDSVRMMALAARWRKVHGDIDFVMLAKNPTIDAWWKLLSREVK 285
EntBvac     DYGLDSVKIMALAGRWRNVFPEVDFVKLAEKPTIDNWHSVLLRSN- 282
            *******::****.***:*. ::*** **::**** * .:* *.
```

Figure 13. Multiple protein alignment of the sequences of EntB (ADB98053.1) and EntB$_{vac}$. The alignment was performed using Clustal Omega (1.2.1) (EMBL-EBI). The identical amino acids between sequences are indicated by an asterisk (*), conserved residues with a colon (:) and similar residues with a period (.).

Discussion

In this Chapter, I characterized the recently identified ENT NRPS system acquired through HOT from bacteria into the yeast W/S clade (54). I decided to study this system because none of the W/S cluster members code for the MLP-encoding gene *ybdZ* within their cluster or an MLP-homolog elsewhere in their genomes, raising the question of whether this NRPS system evolved to be MLP-independent (Fig.1). Previous research in our lab has focused on understanding the role of MLPs in NRPS enzymology (5, 18, 21, 39). One of the goals in the field is to overcome the MLP challenge by understanding what makes an A domain MLP-dependent and how we can change an MLP-dependent A domain to be MLP-independent. Here, we have a NRPS assembly line that appears to have lost the MLP-encoding gene and possibly evolved to become MLP-independent; providing a system to ask why YbdZ is not required. To ask this question I focused my studies on the ENT BGC from *C. vaccinii* NRRL Y-17684 as a representative of the W/S clade (Fig. 1).

I needed to investigate whether the yeast NRPS enzymes function in a similar manner to their *E. coli* ENT homologs using a combination of *in vivo* and *in vitro* experiments. For my *in vivo* studies, I investigated whether the yeast genes could complement *E. coli* MG1655 mutant strains for ENT production and growth under ILM. The gene coding for $EntE_{vac}$, the stand-alone A domain responsible for the recognition and incorporation of DHB, was able to functionally replace the loss of $entE_{eco}$ for growth under ILM and ENT production *in vivo* (Fig. 13). Preliminary findings shows that $EntE_{vac}$ activates DHB (data not shown). In contrast, $entF_{vac}$ was not able to functionally complement the loss of *entF* (Fig. 10). I tested if co-expression of $entF_{vac}$ with the phosphopantetheinyl transferase *sfp* was required to obtain functional enzyme in *E. coli* cells. $EntF_{vac}$/Sfp positive cells were not able to restore growth under ILM or ENT production in MG1655 Δ*entF*.

To investigate why *entF*$_{vac}$ was not able to replace *entF*$_{eco}$, I heterologously overproduced and purified *entF*$_{vac}$ in the presence and absence of YbdZ to characterize the NRPS. In contrast to EntF$_{eco}$/YbdZ interactions, YbdZ did not influence the solubility of EntF$_{vac}$ nor did the proteins copurify (Fig. 6). Interestingly, co-production of EntF$_{vac}$ with YbdZ resulted in two populations of the NRPS, with only one of the populations of EntF$_{vac}$/YbdZ retaining its activity for L-Ser activation (Fig. 7). When EntF$_{vac}$ was heterologously overproduced without YbdZ, only one population of EntF$_{vac}$ was observed and this form of the NRPS was active for L-Ser activation. The kinetic parameters of the active enzymes showed that EntF$_{vac}$/YbdZ (5.4 $\pm$ 2.1 mM) was a slightly better enzyme than EntF$_{vac}$/EV (14.6 $\pm$ 3.2 mM) in terms of affinity for L-Ser (K_m,) but its V_{max} is negative impacted over 8-fold, resulting in a reduced enzyme efficiency (V_{max}/K_m) for L-Ser activation (Table 2). Our previous work on EntF$_{eco}$, the impact of YbdZ as a K_m that is is 10-fold lower, but there was no significant effect on enzyme turnover (5, 18, 21). Thus, it appears the EntF$_{vac}$ has evolved to be a more efficient enzyme in the absence of YbdZ.

One argument for YbdZ-dependence of EntF$_{eco}$ was the enzyme having a higher affinity for L-Ser when YbdZ was present. This is physiologically important because the intracellular concentration of L-Ser in *E. coli* was found to be comparable to the L-Ser K_m of EntF$_{eco}$/YbdZ (5). In contrast, the intracellular concentration of L-Ser in yeast was recently determined to be in the 1-5 mM range (62), which is comparable to the L-Ser K_m of EntF$_{vac}$/EV (Table 2). Based on these observations, I proposed that one of the factors that enabled EntF$_{vac}$ to become MLP-independent is that the physiology of the organism does not require the higher affinity for L-Ser, in contrast to *E. coli* that YbdZ is essential to lower the K_m of EntF$_{eco}$ to a value in the range of intracellular L-Ser (0.1mM) (63). These findings may also provide a reason for why EntF$_{vac}$ was unable to functionally complement the loss of *entF*$_{eco}$ in *MG1655* Δ*entF* cells (Fig. 9).

I hypothesized that YbdZ was able to negatively impact $EntF_{vac}$ when co-expressed of YbdZ because there are still some residues conserved in the MLP/A domain interface (Fig.6). Interaction of YbdZ with $EntF_{vac}$ could have caused changes in the protein structure of $EntF_{vac}$ resulting in an inactive protein, as observed for one of the population of enzymes (Fig.7). However, we were still interested in understanding if the yeast ENT system needed YbdZ for ENT production. Preliminary *in vitro* ENT reconstitution assays showed that YbdZ might not require for *C. vaccinii* NRRL Y-17684 ENT production. Also, exogenous addition of YbdZ had no significant impact on $EntF_{vac}$ enzyme activity, as seen in Fig.8, these findings suggest that YbdZ was able to interact with $EntF_{vac}$ when co-produced but not when added to purified enzymes. Further investigations are required to elucidate these potential interactions.

In conclusion, our findings confirm the predicted functions of the NRPS components for the yeast ENT system from *C. vaccinii* NRRL Y-17684 and suggest that $EntF_{vac}$ is YbdZ-independent. We hypothesize that $EntF_{vac}$ evolved to become MLP-independent by a combination of factors including the higher intracellular L-Ser concentration that alleviated the requirement for a lower K_m (Table 3) and a lower ENT production requirement for survival under ILM since yeast cells have other mechanisms for iron acquisition (59), and the relieve of MLP-interactions resulted in changes to the protein surface and MLP interface regions (Fig. 3). $EntF_{vac}$ and possibly other W/S clade ENT systems are the first NRPS-encoding gene cluster characterized to have evolved MLP-independence.

Materials and Methods

Strains and plasmid construction. All strains used in this study are listed in Table 3. All overexpression vectors were constructed using a polymerase incomplete primer extension (PIPE)-based method (44). The primers used to amplify vectors and inserts used in this study are listed in Table 4. The genes coding for the NRPS components were amplified from *C. vaccinii* NRRL Y-17684 genomic DNA, kindly provided by the C. Hittinger laboratory, and cloned into a pET28b vector with an N- or C-terminal histidine tag as indicated. All plasmids are listed in Table 5.

Strain	Description	Reference
BL21(DE3)		(5)
BL21(DE3) ybdZ::acc(3)IV	BL21(DE3) with apramycin resistance gene disrupting *ybdZ*	(5)
BAP1	BL21(DE3) Δ*prp*RBCD::T7prom-*sfp*,T7prom-*prpE*	(64)
MG1655 Δ*entF*		this study
MG1655 Δ*entE*		this study

Table 3. Bacterial strains used in the course of this study.

Primer name	Sequence (5'-3')
N-terminal tagged	
F_pET_entB_vac	CCGCGCGGCAGCCATATGATCCCGCAACTGAAGTCG
R_pET_entB_vac	TGCGGCCGCAAGCTTTCAATTCGACCGCAGCAAAAC
R_pET_entD_vac	CGGCCGCAAGCTTTCATCATGCAGGGTTATTTTCAG
F_pET_entD_vac	CGCGCGGCAGCCATATGTTTCGGCGAGACTCAGA
R_pET_entF_vac	GAGTGCGGCCGCAAGCTTTCAAATTTCATTTAGTAACTC
F_pET_entF_vac	GCCGCGCGGCAGCCATATGACCCTTCCTCTTATTGCG
R_pET_entE_vac	GGCCGCAAGCTTTCAGGCTGCGGCCTGCGCCGT
F_pET_entE_vac	GCCGCGCGGCAGCCATATGAGCATTGAATTCACACGC
pET28B_kuoF	ACTTTAAGAAGGAGATATACCATGAACCCGTACCCCCATTTAATTATTACG
pET28B_kuoR	GAGTGCGGCCGCAAGCTTTTGTATGTAGGGTTTCAGTAGCGC
pET28B_geoF	ACTTTAAGAAGGAGATATACCATGGCCCTTCCTCTCATCGC
pET28B_geoR	GAGTGCGGCCGCAAGCTTCTCAATGCTTTTAAGTAAATCGTTAATAAGTG
pET28B_bomF	ACTTTAAGAAGGAGATATACCATGGCGCTTCCGTTGATCG
pET28B_bomR	GAGTGCGGCCGCAAGCTTACTGATGAACTTGCGCAGC
pET28B_sorF	ACTTTAAGAAGGAGATATACCATGGCCCTTCCTCTAATAGCAG
pET28B_sorR	GAGTGCGGCCGCAAGCTTTTCTATTTTTTTAAGCAGCTCGTTGATTTGC
pET28B_verF	ACTTTAAGAAGGAGATATACCATGGCATTGCCCCTTGTGGC

pET28B_verR	GAGTGCGGCCGCAAGCTTAGCCAGCTTTGATGCCTCTTC
pEU_F	TAAACTAGCCCAAACGAATTCG
pEU_R	GGCTGTAGTTGTAGAATGTAAA
pEU_Fvac_F	CATTCTACAACTACAGCCCATCATCATCATCATCAC
pEU_Fvac_R	ATTCGTTTGGGCTAGTTTAAATTTCATTTAGTAACTC
R_pEU_NentBvac	TTCGTTTGGGCTAGTTTAATTCGACCGCAGCAAAAC
R_pEU_NentEvac	TCGTTTGGGCTAGTTTAGGCTGCGGCCTGCGCCG
pEU_NotI	GCGGCCGCCCGCGTACGTCGCGACC
pEU_BamHI	GGATCCATATATAGGGCCCGGGTTA
F_pEU_ybdZ	CATTCTACAACTACAGCCATGGCATTCAGTAATCCC
R_pEU_ybdZ	CGAATTCGTTTGGGCTAGTTTATTGTGCCTCCTGCA
F_pEU_Nterminal C-terminal tagged	ACGTACGCGGGCGGCCGCCATCATCATCATCATCAC
pEU_Fvac_R_Cter	GCGACGTACGCGGGCGGCCGCATGGCATTGCCCCTTGTGGC
pEU_Fvac_F_Cter	CGGGCCCTATATATGGATCCTCAGTGGTGGTGGTGGTGG

Table 4. Primers used in the course of this study.

Plasmid	Description/Purpose	Reference
pET28b	T7 expression vector (N-terminal His tag)	(5)
pACYC duet-NO MCS	co-expression empty vector	(5)
pACYC duet-*ybdZ*	co-expression of *ybdZ*	(5)
pET28b-entF$_{vac}$	expression of entF$_{vac}$	this study
pET28b-entF$_{eco}$	expression of entF	(5)
pET28b-entB$_{eco}$	expression of entB	(5)
pET28b-entE$_{vac}$	expression of entE$_{vac}$	this study
pET28b-entB$_{vac}$	expression of entB$_{vac}$	this study

Table 5. Plasmids used in the course of this study.

Growth under iron-limited media and CAS assay. *E. coli* strains MG1655 and MG1655 Δ*entF* were transformed with pET28b empty vector (EV) or pET28b plasmids containing N-terminal histidine tag *entB$_{vac}$* or C-terminal histidine tag *entF$_{vac}$*. Each strain was grown in Luria Bertani (LB) medium with 50 μg/mL kanamycin for 8 hours at 30 °C in triplicate, as previously described (18). Cells were washed in iron-limited media, M9 minimal media (Teknova), and 30 μL were inoculated into 3 mL fresh M9 minimal media supplemented with 0.4% (v/v) glycerol, 50 μg/mL kanamycin, and 2 μM ethylenediamine-*N*,*N*′-bis 2-hydroxyphenylacetic acid (EDDHA) and grown at 37 °C in triplicate.

E. coli strain MG1655 Δ*entF* was transformed with pSU20 empty vector (EV) or pSU20/*sfp* plasmids in combination with pET28b empty vector (EV), pET28b plasmids containing $entF_{eco}$ or $entF_{vac}$. Cultures were prepared as previously described with the addition of 34 µg/mL chloramphenicol. Induction of the $entF_{eco}$ and $entF_{vac}$ genes was achieved using 100 µM of isopropyl beta-D-1-thiogalactosidase (IPTG) added to the M9 minimal media. Growth was stopped after 72 hours and end-point OD_{600nm} was measured. CAS assays were performed to detect siderophore production using 50 µL of culture with 100 µL of CAS reagent, as previously described (18).

Co-overexpression and protein purification of $EntF_{vac}$ with and without YbdZ. Overexpression vector containing $entF_{vac}$ gene was transformed into *E. coli* BL21(□DE3) *ybdZ*::*acc(IV)* cells containing either pACYC-Duet1-NO-MCS or pACYC-Duet1-*ybdZ*, as previously described (5, 21). Overproducing strains were grown in 12 L of LB supplemented with kanamycin (50 µg/mL) and chloramphenicol (34 µg/mL) at 28 °C with shaking. After reaching an OD_{600nm} of 0.5, the temperature was shifted to 15 °C for an hour, then cells were induced with 100 µM of IPTG for 15 hours. Cells were harvested by centrifugation for 30 minutes at 6,000 rpm. Cell pellet was resuspended in His-tag buffer (300 mM NaCl, 20 mM HEPES (pH 7.9), 10% v/v glycerol) and lysed after two passes through a French press. Cell debris was removed after centrifugation for 30 minutes at 15,000 rpm. The cell-free extract was mixed with 5 mM imidazole and Ni-NTA (Thermo Scientific) resin for 2 hours. Protein purification using nickel-affinity chromatography was performed as previously described (21). Fractions containing $EntF_{vac}$ based on SDS-PAGE/Coomassie blue staining were pooled and dialyzed overnight in Buffer A (25 mM NaCl, 20 mM HEPES at pH 7.9) at 4 °C. Dialyzed protein was loaded onto a 5 mL HiTrap Q Sepharose FastFlow column (Sigma Aldrich). A fast protein liquid chromatography (FPLC) system was used to elute the protein during a gradient of 100% Buffer A-0% Buffer B to 0% Buffer A-100% Buffer B over 100 minutes at a flow rate of 1 mL/minute. Buffer B contained 500 mM

NaCl, 20 mM HEPES, pH 7.9. Fractions containing $EntF_{vac}$ based on SDS-PAGE/Coomassie blue staining were pooled into a snake skin dialysis tubing (3.5 kDa molecular weight cutoff, ThermoFisher) and dialyzed overnight in Buffer A at 4 °C. Dialyzed proteins were concentrated, using Millipore Centripep YM-3K concentrator, to 1 mL and loaded into a HiPrep 16/60 Sephracryl S-200 column (Sigma Aldrich). The size exclusion chromatography was achieved using a FPLC system with a 0.5 mL/minute flow rate. SDS-PAGE/Coomassie blue staining was used to identify the fractions with $EntF_{vac}$. Fractions were pooled and concentrated with Centripep YM-3K, flash-frozen using liquid nitrogen and stored at -80 °C.

The proteins co-overproduced with YbdZ will be referred to as $EntF_{vac}$/YbdZ and the proteins co-overproduced with empty vector pACYCduet NO MCS as $EntF_{vac}$/EV throughout the manuscript.

YbdZ detection using Immunoblotting. Seventeen micrograms of purified $EntF_{vac}$/EV or $EntF_{vac}$/YbdZ were loaded in a 16.8% acrylamide Tris-tricine gels. A sample of exogenously produced YbdZ was used as control, 10 µM loaded (previously purified by BS). After protein separation one gel was stained with Coomassie blue stain. A second gel was equilibrated in semidry transfer buffer (50 mM Tris-HCl, 40 mM glycine, 0.037% (w/v) SDS, 20% (v/v) MeOH, pH 8.3), as previously described (39). In summary, the equilibrated gel was transferred onto a polyvinylidene fluoride (PDVF) membrane using a semidry transfer apparatus for 30 min at 15 V. The membrane was dried at 42 °C for 1 hour, then incubated for 1 hour in primary antibody buffer containing phosphate buffered saline (PBS pH 7.4) with 2.5% (w/v) milk, 1% (v/v) Tween-20, and 1:5,000 dilution of 1:10,000 dilution of polyclonal antibody raised to detect YbdZ (Covance Inc.). The membrane was washed for 5 minutes with PBS, then incubated in secondary antibody buffer containing PBS with 2.5% (w/v) milk, 1% (v/v) Tween-20, and 1:5,000 dilution of 1:10,000 dilution of goat anti-rabbit antibody conjugated with horseradish peroxidase. This was followed by two washes with PBS for 5 minutes each, then exposure to SuperSignal West Pico chemiluminescent substrate (ThermoScientific). The membrane was developed using an X-ray film.

Overproduction and purification of $EntB_{vac}$ and $EntE_{vac}$ proteins. Overexpression vector containing N-terminal histidine tagged *$entB_{vac}$* was transformed into *E. coli* BAP1 cells and the overexpression vector containing N-terminal Histidine tagged *$entE_{vac}$* was transformed into *E. coli* BL21DE(3) cells. Overproducing strains were grown in 3 liters of LB supplemented with kanamycin (50 μg/mL). The cells were grown at 28 °C with shaking and after reaching an OD_{600nm} of 0.5, the temperature was shifted to 15 °C for an hour and cells were induced with 100 μM of IPTG for 15 hours. Protein purification using nickel-affinity chromatography was performed as previously described. In summary, fractions containing the protein of interest based on SDS-PAGE/Coomassie blue staining were pooled and dialyzed overnight in Buffer A at 4 °C using a snake skin dialysis tubing (10 kDa molecular weight cutoff, ThermoFisher). Dialyzed protein was loaded onto a 5 mL HiTrap Q Sepharose FastFlow column (Sigma Aldrich). An FPLC system was used to elute the protein during a gradient of 100% Buffer A-0% Buffer B to 0% Buffer A-100% Buffer B over 100 minutes at a flow rate of 1 mL/minute. Fractions containing the protein of interest, based on SDS-PAGE/Coomassie blue staining, were pooled and dialyzed overnight in Buffer A at 4 °C. Dialyzed proteins were concentrated, using Millipore Centripep YM-10K concentrator and flash-frozen in liquid nitrogen.

Determination of protein concentration. The protein concentration of all enzymes purified in this work was calculated using bicinchoninic acid (BCA) protein assay kit (Pierce) with bovine serum albumin as a protein standard.

ATP/PP_i exchange assays. ATP/PP_i exchange assays were performed as previously described (5, 21). In summary, each assay (100 μL) contained 75 mM Tris-HCl (pH 7.5 at 25 °C), 10 mM $MgCl_2$, 5 mM dithiothreitol, 3.5 mM ATP (pH 7.0), 1 mM $NaPP_i$, 1 mM [^{32}P]PP_i (1.0 Ci/mol PerkinElmer), with varying concentrations of L-Serine and enzyme as noted in the figure legends.

All assays were performed in linear range substrate-to-product conversion at 25 °C for 30 minutes. To determine the substrate affinity (K_m) of $EntF_{vac}$ for L-Ser varying concentrations of substrate were used (0, 0.25, 0.5, 0.75, 1, 1.5, 2, 4, 6 and 10 μM). To determine the effect of YbdZ on $EntF_{vac}$ and $EntF_{vac}$/YbdZ kinetic parameters, increasing concentrations of YbdZ (0, 0.25, 2.5, 5 and 10 μM) were added to the reaction mixture without L-Ser substrate for 10 minutes. The reaction was initiated with the addition of 1 mM L-Ser. The reaction was run for 30 minutes and then quenched with a solution containing perchloric acid (3.5% v/v), Na-PP_i (100 mM), and activated charcoal (1.6% w/v). The reactions were centrifuge for 6 minutes at 21,000 rcf to separate the ATP bound to charcoal from the reaction mixture. Pellets were washed with quench solution without activated charcoal and with water before analysis in the scintillation counter. From 3 independent assays, the kinetic parameters were determined using the Michaelis-Menten equation (nonlinear regression analysis) in the GraphPad Prism ver.6.0h.

***In vitro* reconstitution of enterobactin (ENT) production.** Production of ENT *in vitro* was performed as previously published (18) with some modifications. First, $EntF_{eco}$, $EntF_{vac}$/EV, and $EntF_{vac}$/YbdZ were phosphopantetheinylated *in vitro* using 50 mM Tris-HCl (pH 7.5), 10mM MgCl2, 1 mM TCEP, 500 μM CoASH, 5 mM ATP, and 1 μM Sfp during one hour at 37 °C. $EntB_{vac}$ was phosphopantetheinylated *in vivo* by co-overproduction with Sfp in BAP cells. Each ENT *in vitro* reaction contained 20 mM HEPES (pH 7.9), 10 mM ATP, 5 μM TCEP, 15 μM $MgCl_2$, 300 nM EntE, 5 nM holo-EntF, and 5 μM holo-EntB. YbdZ was added to 10 μM. As a background control, a reaction without EntB was used. The reaction was initiated by the addition of 1 mM L-Ser and 10 mM 2,3-DHB. The progress of the reaction was followed using the CAS assay and reported as the inverse of the absorbance at 630 nm.

Chapter 4

Conclusions and Future Directions.

The findings presented in this thesis were aimed to contribute to overcoming the MLP challenge in NRPS combinatorial biosynthesis. The repetitive nature of the core enzymology of NRPS systems supports the idea of using a combination of modules from different systems to achieve changes in the structure of the nonribosomal peptide products (1). However, the majority of medically relevant NRPS systems require the presence of an MLP for functionality adding a complexity to the NRPS combinatorial biosynthesis approach (9). Recent findings from the Thomas laboratory showed that noncognate MLP interactions can be detrimental to NRPS function (20), raising issues in how compatible, non-cognate NRPS/MLP interactions can be controlled. Since there is no clear understanding of how to identify a functional noncognate MLP/NRPS pair I hypothesized that two approaches can be used to solve MLP challenge: 1) identify or evolve a "universal" MLP that is able to functionally interact with any NRPS, or 2) understand of how we can change an MLP-dependent NRPS to be independent of the MLP. In Chapter 2 I identified a potential "universal" MLP that has the natural promiscuity to positively interact with cognate and noncognate NRPS systems. In Chapter 3, I characterized an ENT NRPS system that naturally evolved to become MLP independent in yeast. In this Chapter I discuss the conclusions from the previous Chapters and outline the unanswered questions that can lead us to achieve our ultimate goals.

MXAN_3118 as a molecular tool for NRPS.

One of the proposals to overcome the MLP challenge in NRPS enzymology is to identify or evolve a universal MLP that can interact with multiple NRPSs from different systems and organisms. Our findings in Chapter 2 lead us to identify a proposed naturally occurring "universal" MLP. We hypothesize that orphan MLPs, MLPs encoded outside NRPS-encoding BGCs or genes involved in the synthesis of nonribosomal peptides, have a potential to being developed as "universal" MLPs. We identified the orphan MLP MXAN_3118 from *Myxococcus xanthus* DK1622 naturally

interacts with multiple NRPS systems from this bacterium. Concurrently findings from our lab also identified that MXAN_3118 was able to interact with the ENT noncognate system from *E. coli.* We propose that MXAN_3118, and other members of the orphan MLP group, can be develop as a molecular tool for heterologous expression of NRPSs, similar to the broad substrate specificity phosphopantetheinyl transferase Sfp from *Bacillus subtilis*. Sfp is widely used for its essential role in the activation of the T domains from inactive apo-form to active holo-form proteins by incorporating a 4'-phosphopantetheine moiety to the conserved serine on the T domain of NRPSs and PKSs. We propose that MXAN_3118 can be heterologously expressed as another tool to contribute to NRPS activity and stability to overcome the problem of having to introduce multiple MLPs in the case of combining hybrids MLP-dependent NRPSs. However, further experiments need to be done to confirm the use of MXAN_3118 as a molecular tool for NRPS engineering. To test MXAN_3118 potential as a molecular tool I propose to test the MLP-interactions of known MLP-dependent NRPSs in the presence of their cognate MLP versus MXAN_3118.

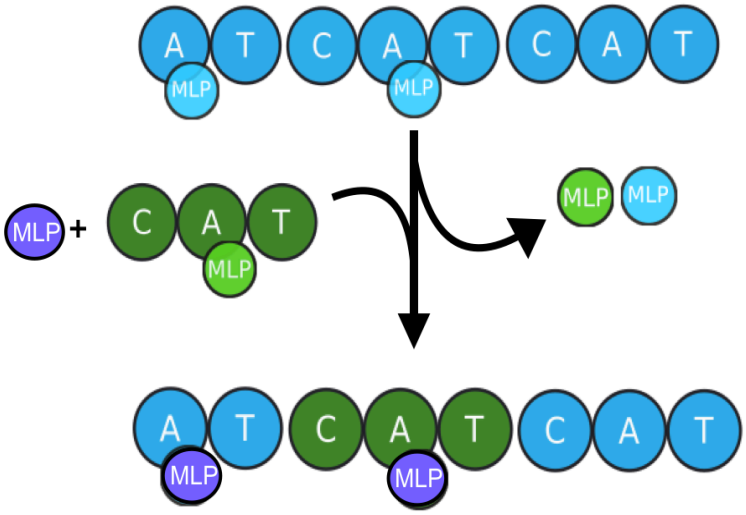


Figure 1. Diagram representing how a universal MLP can solve the combinatorial biosynthesis challenges. MXAN_3118 (purple circle) as a universal MLP might be the key to successful combinatorial biosynthesis efforts.

Bacterial-two hybrid system as a tool for MLP-NRPS interactions.

The complexity and variability of MLP/NRPS interactions have made it difficult to find one assay that can determine if an NRPS is MLP-dependent. In Chapter 2 for the first-time a bacterial-two hybrid (B2H) system was used to detect MLP/NRPS interactions (Fig. 2). I constructed a strain lacking the *E. coli* MLP *ybdZ* to avoid interactions with the introduced MLP/NRPS systems. In summary, plasmids coding the B2H fusion proteins were cloned with the A domain fused to the α-subunit of RNA polymerase and MLP was fused to λcI. Positive interactions between these two proteins resulted in increased expression of the *lacZ* reporter and measured by a β-galactosidase activity assay. The interactions detected through the assay can be confirmed and characterized by the known interactions an MLP can have in a NRPS; influence NRPS solubility, co-purification, influence substrate activation or aminoacylation reaction. I propose the use of B2H assays as a method to screen for MLP-dependence of NRPSs that code for MLPs in their BGC and for NRPS modules used for combinatorial biosynthesis efforts. After identifying a NRPS can interact with an MLP further experiments will be required for the characterization of that interaction.

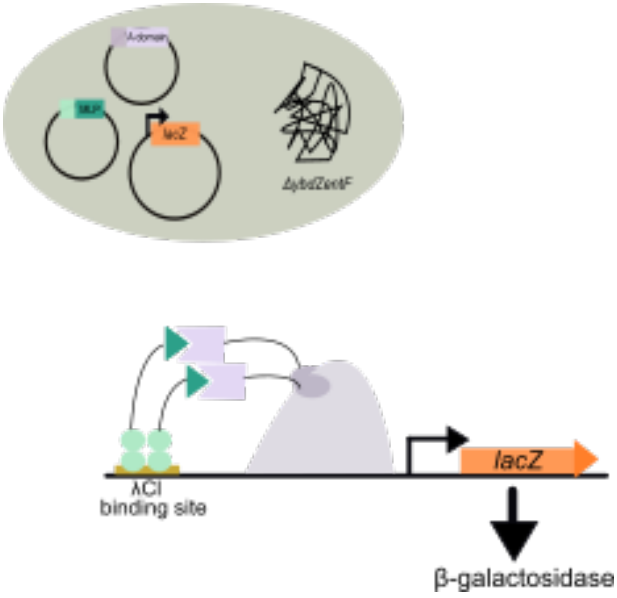


Figure 2. Diagram of Bacterial-two hybrid system to test for MLP/NRPS interactions.

Further characterization of the MLP-independent ENT system from *C. vaccinii* NRRL Y-17684.

A second proposal to overcome the MLP challenge is to evolve MLP-independent A domains to eliminate the MLPs from the engineering of hybrid NRPSs. The Thomas laboratory has been interested in this approach to not only identify how an A domain can become MLP-independent but also to further our understanding about the requirements of NRPSs for MLPs and their evolution. Our findings in Chapter 3 present a NRPS system that evolved to become MLP-independent. The yeast ENT system shows that MLP-independence was achieved by a combination of factors and not just a combination of mutations that resulted in a functional enzyme. I hypothesized that the physiology of the organisms (L-Ser concentration and ENT production) with the changes in the protein structure contributed to the enzyme to not been able to interact with YbdZ, during the assays tested in Chapter 3.

Some of the unanswered questions in Chapter 3 that we are interested in investigating are to test the functionality of $EntB_{vac}$ and $EntF_{vac}$ in strains missing Δ$entB_{eco}$ and Δ$ybdZentF_{eco}$. These pair of strains were missing from our assays to test for functionality of the yeast ENT enzymes *in vivo*. As future directions of this project $EntB_{vac}$ should be tested for complementation of the loss of $entB_{eco}$ and I expect that, same as $entE_{vac}$, $entB_{vac}$ will be able to functionally interact with the other NRPS components from *E. coli* and complement the loss of $entB_{eco}$. Also, I propose to test if $EntF_{vac}$ can complement cells missing both *ybdZ* and $entF_{eco}$ in a double mutant *E. coli* strain. Our findings in Chapter 3 showed that $entF_{vac}$ cannot complement the loss of $entF_{eco}$. I expect to have the same results as with Δ$entF_{eco}$ expressing $entF_{vac}$. However, since I observed a negative effect when $EntF_{vac}$ was co-expressed with YbdZ it is also important to test if the chromosomal levels of *ybdZ* during complementation assay in Δ$entF_{eco}$ strains had any impact. Preliminary results show that YbdZ is not required for ENT_{vac} production *in vitro*. Further experiments need to be performed to determine the effect of YbdZ in ENT_{vac} production using the ENT *in vitro* reconstitution assay.

I investigated the option of overexpressed yeast ENT NRPS using a cell free extract (CFE) but the low levels of proteins were not enough for biochemical assays (data not shown). I propose that the use *S. cerevisiae* as a heterologous host will provide a native eukaryotic environment for the enzyme production and a possible way to address metabolite production *in vivo*. *C. vaccinii* NRRL Y-17684 ENT enzymes should be heterologously expressed in a yeast host to confirm their activity are similar as the one I observed in Chapter 3. In addition, the possibility that there is another factor in yeast cells that can act as an MLP cannot be excluded. The proteins purified from *S. cerevisiae* can be used to determine if a factor from yeast cells copurifies with them or a pulldown assay using *C. vaccinii* NRRL Y-17684 cell-free extract to determine if there is a MLP-like molecule in yeast cells.

Conclusion.

The enigmatic role of MLPs has been a constant question since its discovery as an essential component of NRPS enzymology for the last decade. Recent findings from our lab support that our further understanding of MLPs is essential for successful combinatorial biosynthesis efforts. It is critical that MLPs are taken into consideration during NRPS engineering because incorrect combination of noncognate MLP/NRPS systems can be inhibitory for enzyme function (20). The focus of the work in this thesis had the goal to overcome the MLP challenge using two unusual proteins from NRPS enzymology; the orphan MLPs and the ENT NRPS system from *C. vacinnii* NRRL Y-17684. Our findings in Chapter 2 present a new class of MLPs that we called orphan MLPs since they are not coded within NRPS BGCs or genes predicted to be involved in NRPS enzymology. The orphan MLP MXAN_3118 from *M. xanthus* DK1622 was identified to interact with seven-NRPSs from different clusters and it was constitutively expressed during growth of the organism. The flexibility of this MLP to naturally interact with multiple NRPSs lead us to hypothesize that MXAN_3118 may be used as a “universal” MLP during the construction of

functional hybrid NRPSs. The development of MXAN_3118 as a molecular tool for NRPS engineering might be the key needed for successful nonribosomal peptide hybrid construction. In Chapter 3, our studies of the enterobactin NRPS system from *C.vaccinii* NRRL Y-17684 lead us to hypothesize that MLP-independence was achieved by a combination of factors, including overall protein changes but mainly attributed to the physiology of yeast cells, and that further studies need to be done to determine how we can change a MLP-dependent NRPS to become MLP-independent.

REFERENCES

1. Winn M, Fyans JK, Zhuo Y, Micklefield J. 2016. Recent advances in engineering nonribosomal peptide assembly lines. Nat Prod Rep.
2. Felnagle EA, Jackson EE, Chan YA, Podevels AM, Berti AD, Mcmahon MD, Thomas MG. 2008. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. Mol Pharm 5:191–211.
3. Finking R, Marahiel MA. 2004. Biosynthesis of nonribosomal peptides. Annu Rev Microbiol 58:453–88.
4. Strieker M, Tanović A, Marahiel MA. 2010. Nonribosomal peptide synthetases: Structures and dynamics. Curr Opin Struct Biol.
5. Felnagle EA, Barkei JJ, Park H, Podevels AM, McMahon MD, Drott DW, Thomas MG. 2010. MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. Biochemistry 49:8815–7.
6. Zhang W, Heemstra JR, Walsh CT, Imker HJ. 2010. Activation of the pacidamycin PacI adenylation domain by MbtH-like proteins. Biochemistry 49:9946–9947.
7. Imker HJ, Krahn D, Clerc J, Kaiser M, Walsh CT. 2010. N-Acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF. Chem Biol 17:1077–1083.
8. Boll B, Taubitz T, Heide L. 2011. Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. J Biol Chem 286:36281–36290.
9. Baltz RH. 2011. Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. J Ind Microbiol Biotechnol 38:1747–60.
10. Reichert J, Sakaitani M, Walsh CT. 1992. Characterization of EntF as a serine-activating enzyme. Protein Sci 1:549–556.
11. Quadri LEN, Sello J, Keating TA, Weinreb PH, Walsh CT. 1998. Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. Chem Biol 5:631–645.
12. Gehring AM, Mori I, Walsh CT. 1998. Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. Biochemistry.
13. Stegmann E, Rausch C, Stockert S, Burkert D, Wohlleben W. 2006. The small MbtH-like protein encoded by an internal gene of the balhimycin biosynthetic gene cluster is not required for glycopeptide production. FEMS Microbiol Lett 262:85–92.
14. Ochsner UA, Wilderman PJ, Vasil AI, Vasil ML. 2002. GeneChip® expression analysis of the iron starvation response in *Pseudomonas aeruginosa*: Identification of novel pyoverdine biosynthesis genes. Mol Microbiol 45:1277–1287.
15. Wolpert M, Gust B, Kammerer B, Heide L. 2007. Effects of deletions of mbtH-like genes on clorobiocin biosynthesis in *Streptomyces coelicolor*. Microbiology 153:1413–23.
16. Lautru S, Oves-Costales D, Pernodet J-LL, Challis GL. 2007. MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in Streptomyces coelicolor M145. Microbiology 153:1405–12.
17. Drake EJ, Cao J, Qu J, Shah MB, Straubinger RM, Gulick AM. 2007. The 1.8 Å crystal structure of PA2412, an MbtH-like protein from the pyoverdine cluster of *Pseudomonas aeruginosa*. J Biol Chem 282:20425–20434.
18. Schomer RA, Thomas MG. 2017. Characterization of the Functional Variance in MbtH-like Protein Interactions with a Nonribosomal Peptide Synthetase. Biochemistry 56:5380–5390.
19. Heemstra JR, Walsh CT, Sattely ES. 2009. Enzymatic tailoring of ornithine in the biosynthesis of the Rhizobium cyclic trihydroxamate siderophore vicibactin. J Am Chem

Soc 131:15317–15329.
20. Schomer RA, Thomas MG. 2017. Characterization of the Functional Variance in MbtH-like Protein Interactions with a Nonribosomal Peptide Synthetase. Biochemistry.
21. McMahon MD, Rush JS, Thomas MG. 2012. Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. J Bacteriol 194:2809–18.
22. Davidsen JM, Bartley DM, Townsend CA. 2013. Non-ribosomal propeptide precursor in nocardicin A biosynthesis predicted from adenylation domain specificity dependent on the MbtH family protein NocI. J Am Chem Soc 135:1749–1759.
23. Zolova OE, Garneau-Tsodikova S. 2012. Importance of the MbtH-like protein TioT for production and activation of the thiocoraline adenylation domain of TioK. Med Chem Comm 3:950.
24. Huang T, Li L, Brock NL, Deng Z, Lin S. 2016. Functional Characterization of PyrG, an Unusual nonribosomal peptide synthetase module from the pyridomycin biosynthetic pathway. ChemBioChem 1421–1425.
25. Buchko GW, Kim CY, Terwilliger TC, Myler PJ. 2010. Solution structure of Rv2377c-founding member of the MbtH-like protein family. Tuberculosis 90:245–251.
26. Herbst DA, Boll B, Zocher G, Stehle T, Heide L. 2013. Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes. J Biol Chem 288:1991–2003.
27. Miller BR, Drake EJ, Shi C, Aldrich CC, Gulick AM. 2016. Structures of a nonribosomal peptide synthetase module bound to MbtH-like proteins support a highly dynamic domain architecture. J Biol Chem 291:22559–22571.
28. Mori S, Green KD, Choi R, Buchko GW, Fried MG, Garneau-Tsodikova S. 2018. Using MbtH-like proteins to alter substrate profile of a nonribosomal peptide adenylation enzyme. ChemBioChem 2186–2194.
29. Tarry MJ, Haque AS, Bui KH, Schmeing TM. 2017. X-Ray crystallography and electron microscopy of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture. Structure 25:783–793.e4.
30. Drake EJ, Miller BR, Shi C, Tarrasch JT, Sundlov JA, Leigh Allen C, Skiniotis G, Aldrich CC, Gulick AM. 2016. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. Nature.
31. Coordinators NR. 2017. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 45:D12–D17.
32. Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. 2012. IMG: The integrated microbial genomes database and comparative analysis system. Nucleic Acids Res 40.
33. Lee SH, Choe H, Kim BK, Nasir A, Kim KM. 2015. Complete genome of the marine bacterium *Wenzhouxiangella marina* KCTC 42284T. Mar Genomics 24:277–280.
34. Weissman KJ, Müller R. 2010. Myxobacterial secondary metabolites: bioactivities and modes-of-action. Nat Prod Rep 27:1276–95.
35. Cortina NS, Krug D, Plaza A, Revermann O, Müller R. 2012. Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. Angew Chem Int Ed Engl 51:811–6.
36. Dove SL, Joung JK, Hochschild A. 1997. Activation of prokaryotic transcription through arbitrary protein- protein contacts. Nature 386:627–630.
37. Schley C, Altmeyer MO, Swart R, Mu R, Huber CG, Müller R, Huber CG. 2006. Proteome analysis of *Myxococcus xanthus* by off-line two-dimensional chromatographic separation using monolithic poly- (styrene-divinylbenzene) columns combined with ion-trap tandem mass spectrometry. J Proteome Res 5:2760–2768.

38. Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, Eisen JA, Eisen J, Ronning CM, Barbazuk WB, Blanchard M, Field C, Halling C, Hinkle G, Iartchuk O, Kim HS, Mackenzie C, Madupu R, Miller N, Shvartsbeyn A, Sullivan SA, Vaudin M, Wiegand R, Kaplan HB. 2006. Evolution of sensory complexity recorded in a myxobacterial genome. Proc Natl Acad Sci U S A 103:15200–5.
39. Schomer RA, Park H, Barkei JJ, Thomas MG. 2018. Alanine scanning of YbdZ, an MbtH-like protein, reveals essential residues for functional interactions with its nonribosomal peptide synthetase partner EntF. Biochemistry 57:4125–4134.
40. Quadri LEN, Weinreb PH, Lei M, Nakano MM, Zuber P, Walsh CT. 1998. Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carder protein domains in peptide synthetases. Biochemistry.
41. Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res.
42. Maddison WP, Maddison DR. 2015. Mesquite: a modular system for evolutionary analysis. Version 2.75. 2011. URL http//mesquiteproject org.
43. Rambaut A. 2016. FigTree v1.4.3. Mol Evol phylogenet ics Epidemiol.
44. Klock HE, Lesley SA. 2009. The polymerase incomplete primer extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. Methods Mol Biol 498:91–103.
45. Deaconescu AM, Chambers AL, Smith AJ, Nickels BE, Hochschild A, Savery NJ, Darst SA. 2006. Structural basis for bacterial transcription-coupled DNA repair. Cell 124:507–520.
46. Kaiser D. 1979. Social gliding is correlated with the presence of pili in *Myxococcus xanthus.* Proc Natl Acad Sci.
47. Ueki T, Inouye S, Inouye M. 1996. Positive-negative KG cassettes for construction of multi-gene deletions using a single drug marker. Gene 183:153–157.
48. Leibman M, Hochschild A. 2007. A σ-core interaction of the RNA polymerase holoenzyme that enhances promoter escape. EMBO J.
49. Bretscher AP, Kaiser D. 1978. Nutrition of *Myxococcus xanthus*, a fruiting Myxobacterium. J Bacteriol 133:763–768.
50. Hamilton CM, Aldea M, Washburn BK, Babitzke P, Kushner SR. 1989. New method for generating deletions and gene replacements in *Escherichia coli.* J Bacteriol 171:4617–4622.
51. Banta AB, Chumanov RS, Yuan AH, Lin H, Campbell EA, Burgess RR, Gourse RL. 2013. Key features of σS required for specific recognition by Crl, a transcription factor promoting assembly of RNA polymerase holoenzyme. Proc Natl Acad Sci 110:15955–15960.
52. Thibodeau SA, Fang R, Joung JK. 2004. High-throughput β-galactosidase assay for bacterial cell-based reporter systems. Biotechniques 36:410–415.
53. Julien B, Kaiser a D, Garza a. 2000. Spatial control of cell differentiation in *Myxococcus xanthus*. Proc Natl Acad Sci U S A 97:9098–9103.
54. Kominek, Jacek; Doering, Drew; Opulente, Dana; Shen; Xing-Xing; Zhou, Xiaofan; DeVirgilio, Jeremy; Hulfachor, Amanda; Kurtzman, Cletus; Rokas, Antonis; Hittinger C. 2018. Eukaryotic acquisition of a bacterial operon.(Personal communication)
55. Walsh CT, Liu J, Rusnak F, Sakaitani M. 1990. Molecular studies on enzymes in chorismate metabolism and the enterobactin biosynthetic pathway. Chem Rev.
56. Pettis GS, McIntosh MA. 1987. Molecular characterization of the *Escherichia coli* enterobactin cistron *entF* and coupled expression of *entF* and the *fes* gene. J Bacteriol.
57. Heymann P, Gerads M, Schaller M, Dromer F, Winkelmann G, Ernst JF. 2002. The siderophore iron transporter of *Candida albicans* (Sit1p/Arn1p) mediates uptake of ferrichrome-type siderophores and is required for epithelial invasion. Infect Immun.
58. Potrykus J, Ballou ER, Childers DS, Brown AJP. 2014. Conflicting interests in the

pathogen-host tug of war: fungal micronutrient scavenging versus mammalian nutritional immunity. PLoS Pathog.
59. Lesuisse E, Blaiseau PL, Dancis A, Camadro JM. 2001. Siderophore uptake and use by the yeast *Saccharomyces cerevisiae*. Microbiology.
60. Leduc D, Battesti A, Bouveret E. 2007. The hotdog thioesterase EntH (YbdB) plays a role in vivo in optimal enterobactin biosynthesis by interacting with the ArCP domain of EntB. J Bacteriol.
61. Sikora AL, Wilson DJ, Aldrich CC, Blanchard JS. 2010. Kinetic and inhibition studies of dihydroxybenzoate-AMP ligase from *Escherichia coli*. Biochemistry.
62. Hans MA, Heinzle E, Wittmann C. 2003. Free intracellular amino acid pools during autonomous oscillations in *Saccharomyces cerevisiae*. Biotechnol Bioeng.
63. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli.* Nat Chem Biol.
64. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C. 2001. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. Science (80).