Published in final edited form as:

Biochemistry. 2018 July 17; 57(28): 4125–4134. doi:10.1021/acs.biochem.8b00552.

Alanine scanning of YbdZ, an MbtH-like protein, reveals essential residues for functional interactions with its nonribosomal peptide synthetase partner EntF

Rebecca A. Schomer, Hyunjun Park, John J. Barkei, and Michael G. Thomas*

Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin, 53706

Abstract

Nonribosomal peptide synthetases (NRPSs) are megasynthetases that require complex and specific interactions between multiple domains and proteins to functionally produce a metabolite. MbtH-like proteins (MLPs) are integral components of many NRPSs and interact directly with the adenylation domain of the megasynthetases to stimulate functional enzymology. All of the MLP residues that are essential for functional interactions between the MLP and NRPS have yet to be defined. Here we probe the interactions between YbdZ, an MLP, and EntF, an NRPS, from *Escherichia coli* by performing a complete alanine and serine scan of YbdZ. A phenotypic screen identified eleven YbdZ variants that are unable to replace the wild-type MLP, and these YbdZ variants were characterized using a series of *in vivo* and *in vitro* assays in an effort to explain why functional interactions with EntF were disrupted. All of the YbdZ variants enhanced the solubility of overproduced EntF, suggesting they were still capable of direct interactions with the megasynthetase. Conversely, we show that EntF also influences the solubility of YbdZ and its variants. *In vitro* biochemical analyses of EntF function with each of the YbdZ variants found the impact that an amino acid substitution will have on NRPS function is difficult to predict, highlighting the complex interaction between these proteins.

Graphical Abstract

*Corresponding Author: Contact information for the author to whom correspondence should be addressed. Phone: (608) 263-9075 michael.thomas@wisc.edu.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

SUPPORTING INFORMATION

Modeling the location of YbdZ residues P6, D8, L17, W27, W37, I16, L53, P62 (Figure S1), Modeling the location of YbdZ residues involved in forming or potentially influencing the alanine pocket and the location of P62 in YbdZ and PA2412 (Figure S2), Immunoblotting to detect *in vivo* levels of YbdZ (Figure S3), Assessment of *in vivo* complementation by YbdZ variants (Figure S4), Representative data determining the K_d of EntF for YbdZ variants (Figure S5), Representative data for determining the K_m of EntF + YbdZ variants for L-Ser (Figure S6), SDS-PAGE/Coomassie Blue staining to assess P6A or I16A co-purification with EntF (Figure S7), List of primers used in this study (Table S1), List of strains used in this study (Table S2), List of plasmids used in this study (Table S3).

Keywords

Nonribosomal peptide synthetases; MbtH-like proteins; Natural Product Biosynthesis; Nonribosomal peptides

Nonribosomal peptide synthetases (NRPSs) construct complex molecules via repetitive and modular enzymology. Generally, one NRPS module incorporates one amino acid or aryl acid into a nascent nonribosomal peptide (NRP) product. An NRPS module that initiates peptide synthesis contains, at the minimum, an adenylation (A) domain that activates a specific starting substrate and tethers it to a partner peptidyl carrier protein (PCP) domain to form an amino (or aryl) acyl-*S*-PCP intermediate. Extending modules include a third domain, a condensation (C) domain, which catalyzes amide bond formation between two PCP-linked intermediates to grow the peptide in a directional manner. The terminal module commonly includes a thioesterase (Te) domain that catalyzes the release of the final product by hydrolysis or cyclization. These four domains are considered minimum scaffold upon which a NRPS is built.

Recently, our group¹ and others^{2,3} discovered that there are two distinct classes of A domains based on their requirement for an accessory protein. These accessory proteins are members of the MbtH-like protein (MLP) superfamily. MLPs are relatively small proteins of approximately 60–70 amino acids with three highly conserved tryptophan residues being the signature of this protein family. These proteins are often encoded within the gene cluster that contains the MLP-dependent NRPS genes,^{4,5} and to date these proteins have been found only in bacteria. For those NRPSs that are MLP-dependent, the absence of this accessory protein has been observed to result in a decrease in solubility of the NRPSs,^{1,2,6,7} an alteration in the kinetic parameters for substrate activation,^{1,2,8,9} and significantly reduced catalytic turnover of the entire NRPS.¹⁰ MLPs interact with the A domain of an NRPS, forming a 1:1 complex with this domain.^{1,3,11}

Initial structural analysis of the isolated MLPs determined these proteins have a unique fold consisting of a β -sheet formed by three β -strands followed by either one or two α -helices.^{12,13} While these structures did uncover the unique protein fold of MLPs, they did not provide clear insights into how the MLPs influence NRPS enzymology. Interestingly, one of

these structures included electron density of an undefined molecule bound to the protein.¹³ This raised the possibility that the MLPs may bind a small molecule that influences its function with the NRPS. The first structure of an MLP/NRPS complex was of one of the few known proteins where the MLP is fused to the N-terminus of its partner A domain.¹⁴ This structure identified the binding site for the MLP on the A domain, which was far from the amino acid binding site. This observation suggested MLPs must impart a conformational change in the partner A domain to influence substrate binding. Finally, this structure provided some insights into why two tryptophan residues are so highly conserved in MLPs. These two residues form part of a binding pocket for an alanine residue from the A domain. Whether the other A domain is MLP-dependent or not, the corresponding residue in other A domains is predominantly an alanine or a proline residue. Gulick and colleagues subsequently extended the bioinformatic analysis of MLP-dependent and independent A domains and these data suggest many NRPS A domains may retain an ability to interact with partner MLP.⁹ Several other MLP/NRPS structures have been solved.^{9,15–17} Gulick and colleagues were able to solve the structure of EntF, an NRPS involved in the biosynthesis of the siderophore enterobactin by *E. coli*, with and without its MLP partner YbdZ. Surprisingly, these two structures did not have significant conformational differences that explained how MLPs increased NRPS solubility and adenylyating activity.^{9,15} These structures were solved using a substrate inhibitor, serine adenosine vinylsulfamide, (Ser-AVS), which covalently links to the 4'-phosphopantethienyl arm of the PCP domain, trapping the NRPS in the adenylyating conformation. It is possible that use of this inhibitor limits our ability to capture EntF in the less active state and that in solution EntF without YbdZ is found in multiple conformational states. The finding that similar NRPS systems that are involved in the assembly of the siderophore corynebactin (also known as bacillibactin) have altered apparent molecular weights in the presence or absence of the partner MLP supports the hypothesis of an uncaptured conformational state of the uncomplexed NRPS.¹⁶

In addition to structural analyses, several attempts to understand the role of the MLP using mutational studies have been performed.^{1–3,8,14} The signature tryptophan residues of the MLP were targeted by several groups. Felnagle *et al.* exchanged the signature tryptophan residue that is not found in the tryptophan pocket with an alanine and phenylalanine residue.¹ These residue changes resulted in MLP variants that no longer co-purified with their NRPS partner and did not confer functional activity to the A domain. Zhang *et al.* targeted the tryptophan residues in the structural tryptophan pocket.² Exchanging one of the tryptophan residues with an alanine residue resulted in a 50% reduction in A domain stimulation compared to the wild-type MLP. If both tryptophan residues are replaced with alanines, the stimulatory effect of the MLP is abolished. Herbst *et al.* examined the effect of altering the serine residue that falls in between the signature tryptophans, observing that a tyrosine could not be accommodated at this site.¹⁴ To our knowledge, only one study has attempted to target residues of the MLP that are not involved in the tryptophan pocket.⁸ Garneau-Tsodikova and colleagues exchanged three highly conserved prolyl residues with alanine residues individually and in pairs. They argued that the identity of these residues is not critical for the MLP to stimulate A domain activity.

Despite the wealth of prior structural and biochemical data, researchers in the field have not identified the physiologically essential residues of the MLP, nor a clearly defined the role the

conserved residues have in enabling the MLP to function. For such an evaluation, it is essential to use a model NRPS system that provides the necessary *in vivo* and *in vitro* tools to dissect MLP/NRPS interactions. The NRPS that produces the siderophore enterobactin (ENT) is such a system. ENT is produced by *Escherichia coli* using an iterative, two-module NRPS that contains one MLP-dependent A domain.^{1,18} This A domain is within the second module, which is contained by the NRPS EntF (Fig. 1). An *E. coli* strain lacking *ybdZ* cannot grow in iron-limited media (ILM) due to compromised production of ENT.¹ This provides a simple *in vivo* means for assessing physiologically relevant functions between YbdZ and EntF. We have also recently shown we can reconstitute functional YbdZ/EntF complexes *in vitro*¹⁰ enabling us to dissect these interactions and assess the function of the complex. Here we present a full alanine scan of YbdZ and assess the function of protein variants using both *in vivo* and *in vitro* techniques.^{10,19–21} This work provides the most complete structure/function analysis of an MLP/NRPS complex to date. Phenotypically, we identify 11 YbdZ variants that fail to replace YbdZ *in vivo* due to the disruption of important MLP/NRPS interactions or destabilization of critical structural elements of the MLP.

Materials & Methods

Generation of *ybdZ* mutants

The *ybdZ* mutations were generated by PCR-based Quikchange site-directed mutagenesis of pBAD33-*ybdZ*.¹ The primers used to mutate the *ybdZ* codons are listed in Supplementary Table 1. All plasmids were sequenced to confirm the mutation and ensure no additional mutations were generated by PCR amplification. All of plasmid constructs and strains used in this study are listed in Supplementary Tables 2 and 3.

Assessing the *in vivo* function of *ybdZ* mutants

Plasmids expressing the *ybdZ* mutants were transformed into the *E. coli* strain BW27749 *ybdZ*.¹⁰ The resulting strains were assessed for growth in iron-limited media (ILM), which was Teknova M9 minimal media supplemented with 0.4% v/v glycerol, 25 μ M ethylenediamine-N,N'-bis(2-hydroxyphenylacetic acid) (EDDHA) and 34 μ g/mL chloramphenicol as previously described.¹⁰ The final OD₆₀₀ was recorded after 48 hrs of growth shaking at 200 rpm at 37°C. Growth curves of select strains carrying *ybdZ* mutants were also performed in the presence or absence of 0.001% w/v arabinose to the media as previously described.¹⁰

Co-overproduction of EntF with YbdZ and YbdZ variants for solubility assay

BL21(DE3) *ybdZ::acc(IV)*¹ was transformed with pTEV5-*entF* and pACYC-duet-1 containing *ybdZ* or the *ybdZ* coding for the YbdZ variants: P6A, D8A, I16A, L17A, W27A, W37A, S48A, L53A, A55S, P62A and Q69A. The genes coding for the YbdZ variants were cloned in a manner that resulted in the production of un-tagged MLPs. Proteins were overproduced as previously described¹⁰ and cells from 100 mL of each strain were harvested by centrifugation. These cells were resuspended in 5 mL of 20 mM Tris-HCl (pH 8.0 at 4°C) with 300 mM NaCl and sonicated on ice. BSA protein concentration assays were performed and compared to a standard curve to quantify total protein in each cell-free lysate. Seventy-

five micrograms of protein were run separated by SDS-PAGE (10% polyacrylamide) and proteins were visualized by Coomassie blue staining.

Detection of *in vivo* levels of YbdZ and EntF by immunoblotting

A pBAD33 vector containing *ybdZ* or the *ybdZ* coding for the YbdZ variants P6A, D8A, I16A, L17A, W27A, W37A, S48A, L53A, A55S, P62A and Q69A was transformed with pTEV5 or pTEV5-*entF* into BW27749 *ybdZ*. Each strain was then grown in Lysogeny Broth (LB) containing 34 µg/mL chloramphenicol and 100 µg/mL ampicillin for 12 hours and then 100 µL was subcultured into 3 mL of M9 minimal media containing 4% glycerol and both antibiotics. When cultures reached an OD₆₀₀ of 0.5, protein production was induced by adding 60 µM isopropyl β-D-1-thiogalactopyranoside (IPTG) or 0.001% w/v arabinose. Induced cultures were incubated for 2 hours at 37°C before cells were harvested by centrifugation. Harvested cells were treated with 100 µL 1X Bugbuster™ for each 100 µg of cell paste. After a ten-minute incubation gently rocking at 25°C, samples were centrifuged at 16,100 × g for 5 min, generating cell free extracts. Equal volumes of cell free extract were separated by SDS-PAGE (16.5% acrylamide) and transferred to a PDVF membrane.¹⁰

The lysates of BL21(DE3) *ybdZ::acc(IV)* were also separated by SDS-PAGE (16.5% acrylamide) and transferred to PDVF membrane as previously described.¹⁰ Membranes were heated for 1 hour at 42°C until dehydrated and stored at 4°C for 16 hours. Membranes were then rehydrated in phosphate buffer saline (PBS) pH 7.4 containing 1% (v/v) Tween-20 and 1:10,000 dilution of polyclonal antibody raised to YbdZ (Covance Inc.) for 1 hour at 25°C. Membranes were washed for 5 minutes in PBS and 1% (v/v) Tween-20 before transfer into PBS pH7.4 containing 1% (v/v) Tween-20 and 1:10,000 dilution of goat anti-mouse-horseradish peroxidase conjugate. Samples incubated in secondary antibody for 1 hour at 25°C. Membranes were washed in 5 min intervals with PBS and exposed to Supersignal® West Pico chemiluminescent substrate (ThermoScientific).

In vitro assays

EntF, EntF/YbdZ, holo-EntF, EntE, and EntB were overproduced and purified as previously described.¹⁰ Overproduction of the EntF/I16A and EntF/P6A protein complexes followed the same protocol as EntF/YbdZ, except that EntF/I16A and EntF/P6A were concentrated and flash frozen in liquid N₂ following the second Ni-NTA column and not taken through the final purification step.¹⁰ All *in vitro* analyses of these proteins were performed as previously described.¹⁰

Results & Discussion

Eleven YbdZ variants are unable to complement for loss of wild-type YbdZ

We were interested in determining the residues of an MLP that are essential for its functional interactions with its NRPS partner. MLPs are relatively small proteins, making a full alanine and serine scan of an MLP feasible. YbdZ was targeted for this approach due to the extensive *in vivo* and *in vitro* techniques available to assess its functional interactions with EntF. Apart from the starting methionine residue, each of the other 71 individual amino acid

residues of YbdZ were substituted with an alanine, with the exception that the seven naturally occurring alanine residues were substituted with serine. Plasmids coding for these 71 YbdZ variants were introduced into the *E. coli* strain BW27749 *ybdZ* and the resulting strains were analyzed for growth in ILM, providing an *in vivo* assessment of how well the variants replaced the wild-type YbdZ. The strains were all tested in three biological replicates and the final OD₆₀₀ was determined for each strain after incubating at 37°C for 48 hours. A comparison of the variants to an empty plasmid control and a plasmid expressing the wild-type *ybdZ* found eleven YbdZ variants that failed to reach a final OD₆₀₀ above 0.4 (Fig. 2). These eleven variants were: P6A, D8A, I16A, L17A, W27A, W37A, S48A, L53A, A55S, P62A and Q69A.

Amino acid substitutions that that disrupt YbdZ function map to variety of locations on the YbdZ structure

Using the co-structure of EntF and YbdZ,⁹ we mapped the residues identified in the functional screen to generate hypotheses for why each residue substitution resulted in a nonfunctional or poorly functional MLP. All the residues of interest were present in the available structure with the exception of Q69, which mapped to region of the protein that was unresolved in the crystal structure. Our previous observations proposed that overall topology of the MLP interface was critical in forming functional interactions between the MLP and NRPS.¹⁰ Based on that hypothesis, we anticipated that most of the residues found using our *in vivo* screen would be found on the surface of YbdZ that interacts directly with the A domain.⁹ Consistent with this, seven of the eleven residues (P6, D8, L17, W27, W37, L53, and P62) mapped to the interface between the EntF A domain and the MLP (Fig. 3A.) Surprisingly, two of the residues (S48 and A55) mapped to the surface of YbdZ opposite from EntF, while the two remaining residues (L17 and L53) are located in an internal region of the protein. Many of the residues we found to be conserved in 90% of MLPs from a database of 5,393 MLPs found in GenBank¹⁰ were not identified in our screen and may have greater flexibility than conservation might imply (Fig. 3B).

Using the structure as a guide, we began to develop hypotheses for why each amino acid substitution resulted in decreased YbdZ function. Proline 6 is part of an N-terminal α -helix-like structural component that packs into a cleft of the EntF A domain (Fig. S1A). Replacement of this prolyl residue with an alanine may alter the packing of this α -helix-like structure into the cleft, disrupting or altering the YbdZ/EntF interactions. Aspartate 8 forms ionic interactions with R836 of the conserved A7 motif of the A domain (Fig. S1B). Disruption of this ionic interaction may result in a reduced affinity between EntF and YbdZ. Tryptophan 27, W37 and L17 form a small pocket on the MLP that interacts with A825 of EntF (Fig. S1C). This pocket has been implicated as critical in MLP/NRPS interactions, consistent with earlier mutagenesis studies that found substitutions to either tryptophan residue disrupts MLP function.^{14,22} Our finding that an amino acid substitution of a third residue of this pocket disrupts MLP function extends the importance of this structural component. Surprisingly, the third signature conserved tryptophan residue, W57, which is involved in nonpolar interactions with the A domain, was not found to be essential using our phenotypic screen. Instead, the residue immediately above W57 in an α -helix, L53, was identified. L53 and the internal residue I16 appear to contribute significantly to the structure

of YbdZ (Fig. S1E). We hypothesize that these residues play a role in correctly positioning the pocket composed of L17-W27-W37 by occupying essential structural space in the protein directly behind this feature (Fig. S2A). Proline 62 is located in a C-terminal α helix that interacts with the A domain signature sequence A3 of EntF (Fig. S1E). We predict that if Q69 had been resolved in this MLP structure it would be found in this structural element as well. Interestingly, this is the only region on the EntF/YbdZ and EntF/PA2412 co-structures that do not align (Fig. S2B). Many structural studies have identified that the C-terminal regions of MLPs are highly dynamic.^{9,12,13} Disruption of this helix could alter functional interaction. These differences seen in the primary structure and orientation of the C-terminus are a strong candidate for variation in MLP specificity for its partner NRPS. Lastly, the final two residues identified from the growth screen, S48 and A55, map to the surface of the MLP that faces away from the NRPS (Fig. 3A). The presence of these residues on the outward face of the MLP raise questions about the possibility of the MLP binding a small molecule or an alternate protein partner.

The structural analyses of EntF and EntF/YbdZ did not show any distinct conformational changes upon the binding of YbdZ.^{9,15} Because the enzyme must be captured for crystallography using the inhibitor Ser-AVS, it is likely that trapping the enzyme with this inhibitor locks EntF into the same conformation even if physiologically conformational changes occur with MLP-binding. Unfortunately, due to the lack of informative differences between the two structures, it is difficult to ascertain why the residues we identified phenotypically are not capable of influencing NRPS enzymology like the wild-type YbdZ using the structural data available. Therefore, we proceeded to characterize the influence of the YbdZ variants using immunoblotting and biochemical assays to further understand their decreased functionality *in vivo* and test the hypotheses noted above for explaining why the amino acid substitutions disrupted YbdZ function. For the remainder of this work, we will refer to each YbdZ variant protein by the position of the alanine or serine exchange it contains. For example, the YbdZ variant protein containing a P6A exchange will be referred to as P6A.

Analysis of the *in vivo* levels of the YbdZ variants

To begin investigating why these YbdZ variants fail to fully replace the wild-type YbdZ, the *in vivo* level of each variant was analyzed using immunoblotting with polyclonal antibodies raised against YbdZ. Basal levels of expression from the pBAD promoter were sufficient for *in vivo* complementation by wild-type YbdZ, but these levels were not sufficient for protein detection by YbdZ-specific polyclonal antibodies using immunoblotting (Fig. S3). Increasing the expression of *ybdZ* from the pBAD promoter by adding arabinose to the culture medium did enable detection of YbdZ (Fig. S3). Using this induction condition, the *in vivo* level of each YbdZ variant was assessed and compared to wild-type YbdZ (Fig. 4). From this analysis, only D8A was found to be present at levels similar to those observed for wild-type YbdZ. The other YbdZ variants had significantly reduced or undetectable levels of protein. These data suggest that the stability of nearly all of the YbdZ variants was compromised by the amino acid substitution each carries.

Increased levels of EntF result in increased levels of detectable YbdZ and YbdZ variants

The decreased levels of the YbdZ variants compared to the wild-type YbdZ suggested that the instability of these proteins might explain why they failed to functionally replace the wild-type protein. However, when the expression of the variants was induced, it was reasonable to assume the MLPs were in excess relative to the levels of EntF in the cell. It was possible that the YbdZ variants were less stable than the wild-type enzyme because the majority of the protein being detected by immunoblotting was not complexed with EntF. To date, MLPs have been evaluated for their ability to influence the solubility of an NRPS, but the converse has not been investigated. To address this issue, we introduced a plasmid expressing *entF* into the strains expressing *ybdZ* or *ybdZ* variants and assessed whether the presence of increased levels of EntF influenced the detectable levels of the MLPs. Using the wild-type YbdZ as a starting point, increasing the levels of EntF in the cell resulted in increased levels of detectable YbdZ. This is the first evidence that the stability of an MLP can be influenced by its partner NRPS. Similar analyses of the YbdZ variants showed that increased levels of EntF resulted in detectable levels of nine of the eleven YbdZ variants (Fig. 4). Regardless of whether excess EntF was present or not, neither P6A nor I16A were detectable suggesting these MLPs are very unstable *in vivo* and explains why they fail to functionally replace YbdZ. Similarly, Q69A was barely detectable in the presence of increased levels of EntF and was likely to be too unstable *in vivo* to functionally replace wild-type YbdZ. In contrast, eight of the YbdZ variants were detected at levels comparable to the wild-type YbdZ when EntF levels were increased. These data suggest these YbdZ variants were competent for interacting with EntF and these interactions stabilize these MLPs to levels comparable to wild-type YbdZ.

We note that repeated attempts to detect the levels of chromosomally encoded EntF using immunoblotting all failed. For these attempts, chromosomal *entF* was modified to code for an EntF variant with a C-terminal HA or T7 tag to enable the use of commercially available monoclonal antibodies to these tags. Surprisingly, *E. coli* strains producing these tagged EntF variants had a growth defect in ILM, suggesting the C-terminal tagging of EntF somehow disrupts ENT production. Attempts to use a plasmid coding for an N-terminal His-tagged EntF did enable the detection of EntF, but only when the levels of induced EntF were increased to the point that the protein could also be observed by SDS-PAGE/Coomassie blue staining. Due to these detection limitations, we did not make further attempts to analyze the *in vivo* EntF levels to complement the studies analyzing the *in vivo* levels of YbdZ. The impact the YbdZ variants had on the level of soluble EntF when the NRPS is overproduced with these different MLPs was evaluated and is discussed in a later section.

Increased expression of the *ybdZ* variants does not enable full complementation of a *ybdZ* strain

Based on the finding that most of the YbdZ variants were at lower levels in the cell compared to wild-type YbdZ, but when complexed with EntF their levels were comparable to the wild-type MLP, the *in vivo* complementation was re-evaluated using the arabinose induction conditions. We reasoned that by increasing the levels of the YbdZ variants in the cell, the level of YbdZ variant/EntF complexes would be enhanced, possibly improving their ability to functionally replace wild-type YbdZ. Three of the YbdZ variants showed

improvements in complementation by increasing the expression of the corresponding genes (W27A, W37A, and Q69A) (Fig. S4). Even with these improvements in ENT production for the W27A, W37A, and Q69A variants, none of them complemented near the levels seen with wild-type YbdZ. There were two variants (P6A and A55S) where induction slightly reduced the low level of complementation. Based on these complex *in vivo* results, we turned to more *in vitro* studies to gain additional insights into the reason for many of these variants having the observed phenotypes.

The YbdZ variants increase the solubility of overproduced EntF

We have shown that the amount of soluble EntF overproduced in *E. coli* BL21(DE3) *ybdZ::aac(3)IV* is significantly enhanced when YbdZ is co-overproduced.^{1,10} The same solubility experiments were performed with each of the YbdZ variants. Surprisingly, even though these YbdZ variants were identified because they failed to functionally replace the wild-type MLP *in vivo*, they all enabled overproduced EntF to be significantly more soluble than observed when EntF is overproduced in the absence of an MLP (Fig. 5A). These data support our prior finding that solubility and function are separable.¹⁰

It was unexpected that P6A and I16A positively influenced EntF solubility since we failed to detect these variants by immunoblotting even when EntF levels were increased (Fig. 4). We decided to investigate whether these, and the other YbdZ variants, were detectable in these overproducing strains using the same polyclonal antibodies discussed above for immunoblotting (Fig. 5B). When overproduced, P6A was detected at levels very similar to the wild-type YbdZ. At this time it is not clear why there is such a difference between these observations in *E. coli* BL21(DE3) and BW27749 *ybdZ*. The more unusual finding was I16A remained undetectable by immunoblotting but did enable overproduced EntF to be soluble at levels comparable to the presence of the wild-type YbdZ. Further investigations into the failure to detect I16A are discussed in a later section.

Determining the K_d of EntF for the YbdZ variants

At this point in our study, the data indicated that the YbdZ variants were interacting with EntF *in vivo* as YbdZ variant or EntF levels are increased in the presence of their partner protein. To gain additional insight about how these nonfunctional variants interact with EntF, we aimed to characterize the influence of each variant on the adenylation activity of EntF *in vitro*.

To be consistent with our previous report,¹⁰ we will refer to EntF purified alone as EntF, EntF co-purified with YbdZ as EntF/YbdZ and a reconstituted complex of individually purified EntF and YbdZ as EntF+YbdZ. EntF has an affinity for L-Ser that is 10-fold lower than EntF/YbdZ and EntF+YbdZ^{1,10} (Table 1). Previously, we showed that by adding a gradient of exogenously purified MLP back to EntF, we can determine the binding affinity K_d of EntF for the MLP using ATP-PP_i exchange assays and Michaelis-Menton kinetics.¹⁰ Once the K_d of EntF for a specific MLP is determined, excess MLP can be added and an affinity of the EntF+MLP complex for L-Ser can be assessed.¹⁰

We overproduced and purified each YbdZ variants with the exception P6A, which we could not isolate in the absence of its EntF partner. Surprisingly, we were able to isolate I16A,

which was not detected by antibody even when co-overproduced with EntF in the BL21(DE3) *ybdZ::acc(IV)* strain (Figs. 4 and 5). To determine if the variation EntF affinity for each YbdZ variant correlated with the reduced *in vivo* complementation by each mutant, we determined the K_d of EntF for each YbdZ variant. Notably, three of the variants, D8A, W27A and W37A, have a K_d that is significantly higher in comparison to WT YbdZ. For these three variants, it is likely that, in addition to lower protein levels (Fig. 4), the reduced ability to interact with EntF compromises the ability of these MLPs to complement for the loss of YbdZ (Fig. 1). Consistent with this, EntF has the highest K_d for W37A and this variant increases the solubility of EntF at a noticeably reduced level (Fig. 5A). This aligns with previous observations that noncognate MLPs with increased K_d for EntF sometimes have a reduced influence on NRPS solubility.¹⁰ The remaining variants had either unaffected or slightly impaired K_d in comparison to YbdZ, showing that EntF has a similar affinity for each of these variants, but these complexes are functionally different.

L-Ser activation by the EntF+YbdZ variants

To parse apart whether these YbdZ variants fail to replace YbdZ *in vivo* was due to an altered affinity of EntF for L-Ser while complexed with these YbdZ variants, the K_m for L-Ser for each reconstituted NRPS/MLP complex was determined. Our previous study of noncognate MLPs observed that a noncognate MLP was capable of *in vivo* complementation despite a doubling of the K_m for L-Ser that led to a 2-fold reduction in k_{cat}/K_m compared to the cognate MLP.¹⁰ Noncognate MLPs that had a k_{cat}/K_m that was less than half that of EntF + YbdZ were not capable of *in vivo* complementation. Similar analyses of these kinetic parameters were performed on EntF with the YbdZ variants in this study. With the exception of EntF+W27A, all of the tested EntF+YbdZ variant complexes have increased K_m compared to EntF (Table 1). EntF+W27A is similar to EntF lacking an YbdZ partner, aligning with previous results that this highly conserved residue is critical for MLP function.² EntF+I16A, EntF+W37A, EntF+S48A, EntF+L53A and EntF+A55S have a K_m for L-Ser that is more than double the K_m observed for the EntF+YbdZ complex, resulting in turnover rates that are less than half of EntF+YbdZ. These data led to the conclusion that these residue changes fail to enhance the function of EntF significantly enough to allow growth in ILM.

For the remaining YbdZ variants, several altered EntF kinetics for L-Ser activation in a similar manner to the noncognate MLPs that could functionally replace YbdZ *in vivo*. EntF + D8A, EntF+L17A, EntF+P62A, and EntF+Q69A all have approximately two-fold weaker affinity for L-Ser compared to EntF+YbdZ. The kinetic parameters of these complexes are more similar to EntF+YbdZ than those found for the functional noncognate complexes, EntF + MbtH, EntF+PA2412, EntF+MXAN_3118 and EntF/Atu3678.¹⁰ This level of reduction of K_m was shown to be an insufficient alteration of the EntF enzymology to result in a growth deficiency, implying the mechanism of failure for these variants is likely not due to defects in L-Ser activation. The Q69A levels are clearly compromised in relation to YbdZ both in the presence and absence of EntF; therefore, it is likely that reduced levels of this YbdZ variant are responsible for its inability to replace YbdZ *in vivo*. The reduction of L17A and P62A levels does not appear as severe as the reduced levels of Q69A, while D8A was the only variant we observed that maintained protein levels similar to YbdZ under all conditions

(Figs. 4 and 5). It is possible these three YbdZ variants are either failing to fulfill alternative functions of YbdZ in ENT biosynthesis or negatively influencing an additional step in NRPS enzymology.

D8A, L17A and P62A can efficiently replace YbdZ for *in vitro* ENT production

Prior work has shown that even when the affinity of EntF for L-Ser is compensated for by adding an excess concentration of L-Ser, ENT assembly is still compromised, suggesting YbdZ influences more than just amino acid binding by EntF.¹⁰ Using the same *in vitro* ENT biosynthesis reconstitution assays, we probed whether D8A, L17A and P62A influenced ENT production like wild-type YbdZ. While these variants influenced EntF affinity for L-Ser, we hypothesized they may be unable to fulfill alternate functions of YbdZ in ENT biosynthesis. However, in the presence of all three of these YbdZ variants, ENT is produced at levels similar to when wild-type YbdZ is present (Fig. 6). These data suggest that the decreased *in vivo* levels of L17A and P62A were the primary cause for their inability to replace YbdZ *in vivo* (Figs. 1 and 4). Reduced solubility does not account for the inability of D8A to replace WT YbdZ, rather the increased K_d of EntF for the D8A is the only parameter we identified that is interfering with the functionality of this YbdZ variant. When D8A levels are increased *in vitro* to account for the decreased affinity of EntF for D8A, the EntF+D8A complex has kinetic parameters similar to WT YbdZ (Table 1). Clearly, increased expression of D8A did not result in increased levels of D8A to compensate for the affinity difference *in vivo*.

Co-production of EntF with I16A and P6A influences EntF adenylation activity

Since we could not visualize or isolate P6A without coproduction of EntF, we co-overproduced EntF and P6A and attempted to purify the EntF/P6A complex to evaluate how this YbdZ variant influences EntF activity. Additionally, the odd observation that I16A was not detected by immunoblotting even though it enhanced over-produced EntF solubility and that we were able to overproduce and purify I16A led us to co-overproduce EntF/I16A and attempt to purify the complex to evaluate EntF activity.

Purification of EntF that was co-overproduced with YbdZ P6A followed by SDS-PAGE/ Coomassie blue staining determined that these proteins did not co-purify in the usual ratio of approximately 1:1^{1,3,11} (Fig. S7). The EntF/P6A complex had to be concentrated to detect minimal quantities of P6A, leading us to conclude that only a small population of this EntF is complexed with P6A. These data suggest that while this YbdZ variant can interact with EntF and enhance its solubility, the affinity of EntF for YbdZ P6A must not be strong enough for the majority to survive purification in a typical ratio. In contrast, EntF and YbdZ I16A did co-purify (Fig. S7). This finding suggests the amino acid substitution abolishes the recognition site of the antibody, making it undetectable by immunoblotting.

The kinetic parameters for L-Ser activation by EntF co-overproduced with P6A and EntF/I16A were determined (Table 1). Even though the EntF co-produced with P6A lost much of the MLP after purification, the kinetics of L-Ser activation were improved relative to EntF purified in the absence of YbdZ. These results suggest that P6A influenced the conformation of EntF, and this altered form of EntF was retained even after the MLP disassociated from

the NRPS. When EntF/I16A was assayed for L-Ser affinity, the kinetics for L-Ser activation were quite similar to those observed for EntF+YbdZ. However, when we added the EntF/I16A to our *in vitro* ENT biosynthesis reconstitution assay, it was severely defective in producing ENT (Fig. 7). These data suggest that while YbdZ I16A is holding EntF in a more soluble and active conformation for adenylation activity, it fails to hold EntF in a conformation that is compatible with overall function of the NRPS.

It is important to note that while the addition of exogenously produced YbdZ to EntF fully restores the adenylation activity to levels of EntF/YbdZ,¹⁰ the K_m of EntF/I16A for L-Ser and K_m of EntF+I16A for L-Ser are quite different. This is likely due to the presence of EntF influencing the solubility of YbdZ (Fig. 4). This complicates our ability to extract physiologically relevant data from characterization of separately purified and recapitulated MLP/NRPS complexes. While co-production allows us to bypass the issue of misfolded MLPs or NRPSs to some extent, one issue with the analysis of co-purified MLP/NRPS complexes is that they vary in levels of MLP-bound and unbound NRPS, as observed in the overproduction and purification of EntF in the presence of YbdZ P6A or YbdZ I16A (Fig. S7). The complexity of the interdependence of the NRPS and MLP for one another makes these interactions challenging to disentangle. Thus, conclusions drawn from a combination of *in vivo* and *in vitro* analyses of any NRPS/MLP interactions are likely to be more informative.

Conclusions

In summary, we identified 11 residues in YbdZ that influenced the ability of this MLP to have functional *in vivo* interactions with EntF. We found no evidence of a single catalytic residue or of a small molecule binding pocket using this functional screen. Rather, the amino acid substitution scan indicated several structural elements of the MLP that are important for functional interactions or MLP stability. In particular, the N-terminal region, an alanine-binding pocket, and the C-terminal helix were identified as important structural features. P6A and D8A of the N-terminal region have compromised interactions with EntF as seen by an inability to co-purify with EntF and a reduced K_d respectively, indicating this region as important in MLP/NRPS affinity for one another. Variants that disrupt the binding pocket for A825 of EntF (L17A, W27A, W37A) or that potentially disrupt the coordination of this binding pocket (I16A and L53A) have reduced stability as well as a variety of functional incompatibilities *in vitro*. This pocket that has been targeted repeatedly for mutagenesis is obviously functionally relevant to the MLP/NRPS complex. Finally, variants that disrupted the C-terminal helix region (P62A, Q69A) have reduced *in vivo* levels, but were able to stimulate EntF activity nearly as well as YbdZ *in vitro*, indicating that this region is critical for the structural integrity of the MLP.

All of the YbdZ variants, with the exception of D8A and I16A, had reduced levels in the absence of the NRPS partner, showing that MLP stability can be influenced by the NRPS partner. However, beyond influencing solubility, the EntF/YbdZ interactions are dynamic and the functional interactions continue after folding. These data indicate that overall topology and stability of the MLP/NRPS interface plays an important role for functional interactions. While the MLP variants that are structurally compromised may be able to

stabilize EntF (Fig. 5A), topological incompatibilities between the MLP and NRPS result in functionally reduced EntF (Table 1). However, the complex influence of the MLP and NRPS on the stability of one another makes these systems difficult to parse apart using *in vitro* methods as both EntF and the YbdZ variants likely have altered properties when overproduced in the absence of their partners. It is also important to note that a caveat to our approach is that the substitution of an alanine for each residue in most cases removes a significant component of the amino acid side chain. Such changes may cause unwanted voids in the internal regions of the protein, impacting protein stability, or will only identify residues with side chains that make specific interactions needed for function NRPS/MLP interactions for those residues that are surface exposed. Even with the limitations, this approach has been invaluable in providing significant insights into the key residues needed for functional EntF/YbdZ interactions.

Throughout the study of NRPS enzymology, protein-protein interactions between the NRPS components have been shown to be specialized and specific for their appropriate partners.^{23–27} This trend holds true for MLPs as well and is further supported by this structure-function analysis. While some conserved residues (W27, W37 and P62) were found to be critical, the majority of the residues we identified were not highly conserved across the MLP superfamily. Understanding the specificity between MLP/NRPS components will be a significant challenge in engineering NRPSs to produce hybrid natural products. Our data may indicate that rather than target specific residues in the MLP to generate functional noncognate interactions, the generation of chimeric MLPs containing critical regions of contact might be a useful strategy for engineering noncognate MLPs that can interact with multiple NRPS partners. Alternatively, our understanding of this specificity would be improved if we knew why certain NRPSs do not need MLPs. The existence of MLP-independent NRPSs indicates that the role of the MLP can be bypassed, yet the structural and biochemical data accumulated thus far do not explain why. Therefore, directed evolution of an MLP-dependent NRPS to be soluble or active in the absence of an MLP is a logical next step in understanding the physiological role of this superfamily and overcoming the intricacy of MLP/NRPS dependence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding Sources

This work was funded by the National Institutes of Health grant GM100346.

ABBREVIATIONS

A	adenylation
PCP	peptidyl carrier protein
C	condensation

NRPS	nonribosomal peptide synthetase
MLP	MbtH-like protein
ENT	enterobactin
EDDHA	ethylenediamine-N,N'-bis(2-hydroxyphenylacetic acid)
TCEP	Tris-(2-carboxyethylphosphine)
IPTG	Isopropyl β -D-1-thiogalactopyranoside
PAGE	polyacrylamide gel electrophoresis
TEV	tobacco etch virus
DTT	dithiothreitol
CAS	Chrome azurol S
Ser-AVS	serine adenosine vinylsulfamide

References

1. Felnagle EA, Barkei JJ, Park H, Podevels AM, McMahon MD, Drott DW, Thomas MG. MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry*. 2010; 49:8815–8817. [PubMed: 20845982]
2. Zhang W, Heemstra JR, Walsh CT, Imker HJ. Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins. *Biochemistry*. 2010; 49:9946–9947. [PubMed: 20964365]
3. Boll B, Taubitz T, Heide L. Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. *J Biol Chem*. 2011; 286:36281–36290. [PubMed: 21890635]
4. Baltz RH. Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J Ind Microbiol Biotechnol*. 2011; 38:1747–60. [PubMed: 21826462]
5. Quadri LE, Sello J, Keating TA, Weinreb PH, Walsh CT. Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem Biol*. 1998; 5:631–645. [PubMed: 9831524]
6. Reichert J, Sakaitani M, Walsh CT. Characterization of EntF as a serine-activating enzyme. *Protein Sci*. 1992; 1:549–556. [PubMed: 1338974]
7. McMahon MD, Rush JS, Thomas MG. Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *J Bacteriol*. 2012; 194:2809–2818. [PubMed: 22447909]
8. Zolova OE, Garneau-Tsodikova S. Importance of the MbtH-like protein TioT for production and activation of the thiocoraline adenylation domain of TioK. *Medchemcomm*. 2012; 3:950.
9. Miller BR, Drake EJ, Shi C, Aldrich CC, Gulick AM. Structures of a nonribosomal peptide synthetase module bound to MbtH-like proteins support a highly dynamic domain architecture. *J Biol Chem*. 2016; 291:22559–22571. [PubMed: 27597544]
10. Schomer RA, Thomas MG. Characterization of the functional variance in MbtH-like protein interactions with a nonribosomal peptide synthetase. *Biochemistry*. 2017; 56:5380–5390. [PubMed: 28880538]
11. Imker H, Krahn D, Kaiser M, Walsh CT. N-Acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide. *Chem Biol*. 2010; 17:1077–1083. [PubMed: 21035730]
12. Buchko GW, Kim CY, Terwilliger TC, Myler PJ. Solution structure of Rv 2377c-founding member of the MbtH-like protein family. *Tuberculosis*. 2011; 90:245–251.

13. Drake EJ, Cao J, Qu J, Shah MB, Straubinger RM, Gulick AM. The 1.8 Å crystal structure of PA2412, an MbtH-like protein from the pyoverdine cluster of *Pseudomonas aeruginosa*. *J Biol Chem*. 2007; 282:20425–34. [PubMed: 17502378]
14. Herbst DA, Boll B, Zocher G, Stehle T, Heide L. Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylation enzymes. *J Biol Chem*. 2013; 288:1991–2003. [PubMed: 23192349]
15. Drake EJ, Miller BR, Shi C, Tarrasch JT, Sundlov JA, Leigh Allen C, Skinotis G, Aldrich CC, Gulick AM. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature*. 2016; 529:235–238. [PubMed: 26762461]
16. Tarry MJ, Haque AS, Bui KH, Schmeing TM. X-Ray crystallography and electron microscopy of cross- and multi module nonribosomal peptide synthetase proteins reveal a flexible architecture. *Structure*. 2017; 25:783–793. [PubMed: 28434915]
17. Mori S, Pang AH, Lundy TA, Garzan A, Tsodikov OV. Structural basis for backbone N-methylation by an interrupted adenylation domain. *Nat Chem Biol*. 2018; 14:428–430. [PubMed: 29556104]
18. Pettis GS, McIntosh MA. Molecular characterization of the *Escherichia Coli* Enterobactin cistron entF and coupled expression of entF and the fes gene. 1987; 169:4154–4162.
19. Gehring AM, Mori I, Walsh CT. Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase. *Biochemistry*. 1998; 2960:2648–2659.
20. Ehmann DE, Shaw-Reid CA, Losey HC, Walsh CT. The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: Sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *PNAS*. 2000; 97:2509–2514. [PubMed: 10688898]
21. Shaw-Reid CA, Kelleher NL, Losey HC, Gehring AM, Berg C, Walsh CT. Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization. *Chem Biol*. 1999; 6:385–400. [PubMed: 10375542]
22. Walsh CT, Zhang W. Chemical logic and enzymatic machinery for biological assembly of peptidyl nucleoside antibiotics. *ACS Chem Biol*. 2011; 6:1000–1007. [PubMed: 21851099]
23. Lai JR, Fischbach MA, Liu DR, Walsh CT. Localized protein interaction surfaces on the EntB carrier protein revealed by combinatorial mutagenesis and selection. *J Am Chem Soc*. 2006; 128:11002–11003. [PubMed: 16925399]
24. Frueh DP, Arthanari H, Koglin A, Vosburg DA, Bennett E, Walsh CT, Wagner G. Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature*. 2008; 454:903–906. [PubMed: 18704088]
25. Hur GH, Meier JL, Baskin J, Codelli JA, Bertozzi CR, Marahiel MA, Burkart MD. Crosslinking studies of protein-protein interactions in nonribosomal peptide biosynthesis. *Chem Biol*. 2009; 16:372–381. [PubMed: 19345117]
26. Jaremko MJ, Lee DJ, Patel A, Winslow V, Opella SJ, Mccammon JA, Burkart MD. Manipulating protein-protein interactions in nonribosomal peptide synthetase type II peptidyl carrier proteins. *Biochemistry*. 2017; 56:5269–5273. [PubMed: 28920687]
27. Liu H, Gao L, Han J, Ma Z, Lu Z, Dai C, Zhang C, Bie X. Biocombinatorial synthesis of novel lipopeptides by COM domain-mediated reprogramming of the plipastatin NRPS complex. *Front Microbiol*. 2016; 7:1–8. [PubMed: 26834723]

Fig. 1.

Diagram of megasynthetase for ENT biosynthesis. This two-module NRPS acts iteratively three times to generate one molecule of ENT.

Fig. 2.

Analysis of complementation of BW27749 *ybdZ* by YbdZ alanine variants. Data shown is the final OD₆₀₀ after 48 hours of growth in ILM. Variants that reached an OD₆₀₀ of 0.4 or lower are highlighted in red. Data are shown as mean of triplicate biological replicates with standard deviation. Abbreviations: EV, empty vector; YbdZ, wild-type protein.

Fig. 3. (A) Surface renderings of YbdZ (blue) structure in complex with EntF (yellow). The outward facing surface of YbdZ (left) and inward facing surface (right) contain most of the residues of interest (red) that result in a growth defect when altered. (B) Amino acid sequence of YbdZ. Underlined residues are on the interface between YbdZ and EntF, starred residues are highly conserved across MLP superfamily (90% conservation across 5,393 unique predicted MLPs from GenBank) and the residues of interest from the growth screen are highlighted in red.

Fig. 4. Representative immunoblots for the assessment of YbdZ and variants levels in the presence and absence of EntF in BW27749 *ybdZ*.

Fig. 5.

Top: Analysis of levels of overproduced EntF in the presence of YbdZ variants. **Bottom:** Immunoblot of analysis of YbdZ variant levels when overproduced with EntF in BL21(DE3) *ybdZ::acc(IV)*.

Fig. 6.
In vitro ENT production in the presence of several YbdZ variants monitored by CAS absorbance.

Fig. 7.

Detection of ENT production by reconstituted ENT biosynthetic system in the presence of EntF, EntF/YbdZ and EntF/I16A as detected by CAS after 80 minutes at 37°C. Each reaction was run in triplicate and standard error is shown. Statistic analyses indicated EntF/I16A was statistically different from EntF ($p=0.0009$) and EntF/YbdZ ($p<0.0001$).

Table 1Kinetic parameters of EntF with YbdZ variants^a.

Protein:	K_d (nM)	K_m (uM Ser)	K_{cat} (s ⁻¹)	K_{cat}/K_m (M ⁻¹ s ⁻¹)
EntF	-	1029 ± 171	0.84 ± 0.050	816
EntF + WT YbdZ	186 ± 69	116 ± 22	0.88 ± 0.04	7858
EntF + D8A YbdZ	555 ± 150	279 ± 59	0.97 ± 0.08	3476
EntF + I16A YbdZ	109 ± 27	430 ± 83	1.26 ± 0.11	2930
EntF + L17A YbdZ	295 ± 88	159 ± 32	0.93 ± 0.05	5849
EntF + W27A YbdZ	636 ± 78	1013 ± 143	0.98 ± 0.07	848
EntF + W37A YbdZ	1177 ± 338	438 ± 130	1.1 ± 0.11	2255
EntF + S48A YbdZ	41 ± 10	451 ± 59	1.14 ± 0.72	2527
EntF + L53A YbdZ	213 ± 100	453 ± 92	1.29 ± 0.13	2847
EntF + A55S YbdZ	270 ± 88	450 ± 80	1.01 ± 0.08	2244
EntF + P62A YbdZ	229 ± 85	297 ± 60	1.21 ± 0.10	4074
EntF + Q69A YbdZ	374 ± 130	208 ± 74	0.94 ± 0.11	4519
EntF/P6A	-	382 ± 59	0.93 ± 0.04	2434
EntF/I16A	-	129 ± 34	0.78 ± 0.04	5736
EntF + MbtH ¹⁰	54.6 ± 8	229 ± 31	0.74 ± 0.04	3231
EntF + PA2412 ¹⁰	358 ± 95	397 ± 94	1.21 ± 0.06	3047

^aData shown in graphical form in Figs. S5 and S6.