Functional analysis of MbtH-like proteins in the biosynthesis of thaxtomin A in *Streptomyces scabiei* 87.22

By

© Yuting Li

A thesis submitted to the School of Graduate Studies in partial fulfillment for the requirements for the degree of

Doctor of Philosophy

Department of Biology

Memorial University of Newfoundland

March 2021

St. John's, Newfoundland and Labrador

Abstract

Streptomyces scabiei is an important causative agent of potato common scab disease. The main pathogenicity factor produced by this organism is thaxtomin A, a phytotoxin that is also a promising bioherbicide for agricultural applications. The biosynthesis of thaxtomin A involves the nonribosomal peptide synthetases (NRPSs) TxtA and TxtB, both of which contain an adenylation (A-) domain that recruits and activates the amino acid substrate to be integrated into the product. A small gene, *txtH*, encoding a predicted member of the MbtH-like protein (MLP) family, is also present in the thaxtomin (Txt) biosynthetic gene cluster. MLPs are normally required for proper folding of the A-domain(s) and/or for stimulating the enzymatic activity of the domain. In addition, some MLPs can interact with NRPSs from different biosynthetic pathways, though the mechanism behind this is not fully understood.

Here, I investigated the role of MLPs during thaxtomin A biosynthesis in *S. scabiei.* The results showed that TxtH likely functions as a chaperone protein for ensuring the proper folding of the TxtA and TxtB A-domains, and that this function is essential for thaxtomin biosynthesis in *S. scabiei*. Using site-directed mutagenesis, I identified amino acid residues within TxtH that are important for the function of the protein. I also showed that two other MLPs encoded in the *S. scabiei* genome can promote thaxtomin production in the absence of TxtH, indicating that they can exhibit functional cross-talk with TxtH. A survey of various MLPs from diverse phylogenetic lineages revealed that most of these MLPs can exhibit functional cross-talk with TxtH to varying degrees, though two MLPs were identified that could not replace TxtH in the assays performed. *In silico* analysis revealed that the conservation of key residues at the Txt MLP-NRPS interacting interface

may determine the ability of an MLP to interact with the Txt NRPSs. I additionally attempted to assess the impact of TxtH and other MLPs on the enzymology of the Txt A-domains *in vitro*; however, my assays were unsuccessful despite testing different conditions. Overall, this study is the first to investigate the function of MLPs during thaxtomin A biosynthesis in *S. scabiei*.

Acknowledgements

The journey of pursuing the PhD degree has been full of excitement and also challenges for me. I am so grateful that so many people were around helping me and walking the journey with me. I would like to take the chance to acknowledge and thank them here.

First and foremost, I would like to thank my supervisors Dr. Dawn R.D. Bignell and Dr. Kapil Tahlan. During these years, they have been unrelenting in their support and guidance for my project. Dawn, thank you for always being available when I needed your help and guidance, thank you for trusting me and giving me the opportunity to try out any experiment, thank you for taking tremendous time and providing the editorial input for my thesis, and thank you for going beyond the academic field, caring for my life and being a friend. Kapil, thanks very much for your technical support as you always gave me valuable advice for my project, thank you for the jokes and laughs that brightened my days of distress. Your passion for research and diligent work style are inspiring to me for becoming a researcher like you. It is my privilege to work with both of you. I would also like to thank my supervisory committee member, Dr. Andrew Lang, for the feedback and help that he has provided. Thank you, Andrew, for always being so kind and supportive. I am honored to have you as my committee member. Additionally, I would like to acknowledge the School of Graduate Studies and the Department of Biology at Memorial University of Newfoundland for providing me the opportunity to pursue my degrees. The financial support was gratefully provided by the President's Doctoral Student Investment Fund at Memorial University and Natural Sciences Engineering Research Council of Canada. I would also like to acknowledge all of the prior publications included in this thesis.

It wouldn't be easy for a PhD project if there was no group of people helping and working with me. I would like to express my appreciation to them. Lancy Cheng, for your advice on the protein work. Luke Bown, for helping with the HPLC and LC-MS analyses. Jingyu Liu, for sharing my struggles and laughs. Gustavo Diaz Cruz, for being my bench mate and having lovely conversations with me. Hannah Perry, for bringing a lot of joy. Nader Fareid Abusara, for being a kind friend. Brandon Piercey, for your assistance on the statistical analysis and figure making. Arshad Shaikh, for your technical suggestions. Damilola Adekunle, for joining the lab and working together with me.

My family has been unconditional in supporting me mentally and financially. Particularly, I would like to thank my grandpa, who initiated the idea of pursuing my degree abroad. This means a lot to me and it changed my life. Special thanks to Zhiyu Chen, for being my husband, best friend, chef in the kitchen and my training coach. Last but not the least, I would also like to give thanks to my church friends for all your prayers, encouragement and all the time we spent together to grow in Christ.

Table of Contents

List of Figures

List of Tables

List of Symbols, Abbreviations and Nomenclature

A-domain: adenylation domain

ABC: ATP-binding cassette

Ala: L-alanine

AMP: adenosine monophosphate

Amp: amplitude

Amp^R: ampicillin resistance

ANL: Acyl-CoA synthetases, NRPS adenylation domains, and Luciferase enzymes

ANOVA: analysis of variance

$Apra^R$: apramycin resistance

ATP: adenosine triphosphate

AS: acid scab

BGC: biosynthetic gene cluster

BLASTP: protein basic local alignment search

bld: bald

BSA: bovine serum albumin

cbs: CebR-binding site

CCR: carbon catabolite repression

CDA: calcium-dependent antibiotic

cDNA: complementary DNA

C-domain: condensation domain

Cml^R: chloramphenicol resistance

CS: common scab

DCW: dry cell weight

DHB: 2,3-dihydroxybenzoate

DMSO: dimethyl sulfoxide

DNA: deoxyribonucleic acid

dNTP: deoxynucleoside triphosphate

DTT: dithiothreitol

EBI: European bioinformatics institute

ECL: enhanced chemiluminescence

EDTA: ethylenediaminetetraacetic acid

EMBL: European molecular biology laboratory

ENT: enterobactin

E-value: expect value

FRT: flip recombinase recognition sites

GTP: guanosine triphosphate

GMQE: global model quality estimation

HGT: horizontal gene transfer

HIS_6: 6×histidine

HPLC: high performance liquid chromatography

HRESIMS: high resolution electrospray ionization mass spectra

HRP: horseradish peroxidase

Hyg^R: hygromycin B resistance

IAA: indole-3-acetic acid

ID: identifier

IPTG: isopropyl β-D-thiogalactopyranoside

ISP-4: international *Streptomyces* project medium 4

iTOL: interactive tree of life

K_m: Michaelis constant

Kan^R: kanamycin resistance

LB: Luria-Bertani

LC: liquid chromatography

LC-MS: liquid chromatography-mass spectrometry

LC-HRESIMS: liquid chromatography-high resolution electrospray ionization mass spectrometry

M-domain: methyltransferase domain

MEGA: molecular evolutionary genetics analysis

MLP: MbtH-like protein

MS: mass

MW: molecular weight

MWCO: molecular weight cut-off

NA: nutrient agar

n/a: not applicable

NCBI: national center for biotechnology information

NTA: nitrilotriacetic acid

NMR: nuclear magnetic resonance

NRP: non-ribosomal peptide

NRPS: non-ribosomal peptide synthetase

OBA: oat bran agar

OBAC: oat bran agar containing 0.35% w/v cellobiose

OBB: oat bran broth

OBBC: oat bran broth containing 0.35% w/v cellobiose

OD: optical density

*ori*T: origin of transfer

PAI: pathogenicity island

PBS: phosphate-buffered saline solution

PCA: phenazine-1-carboxylic acid

PCP: peptidyl carrier protein

PCR: polymerase chain reaction

PDB: protein data bank

Pfam: protein families

PGPS: Plant growth promoting *Streptomyces*

Phe: L-phenylalanine

P_i: inorganic phosphate

PISA: proteins, interfaces, structures, assemblies software

PMA: potato mash agar

Ppant: 4'-phosphopantetheine

PP_i: inorganic pyrophosphate

PP_iase: pyrophosphatase

QMEAN: qualitative model energy analysis

RNA: ribonucleic acid

rpm: revolutions per minute

RT-PCR: reverse transcription-PCR

Sac: *Streptomyces acidiscabies*

Seu: *Streptomyces europaeiscabiei*

SD: Shine-Dalgarno

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SFM: soy flour mannitol

SFMA: soy flour mannitol agar

SOB: super optimal broth

SOC: super optimal broth with catabolite repression

sp.: species (singular)

spp.: species (plural)

Ssc: *Streptomyces scabiei*

Stu: *Streptomyces turgidiscabies*

syn.: synonym

TBE: tris-borate-EDTA

T-domain: thiolation domain

Te-domain: thioesterase domain

$Thio^R$: thiostrepton resistance

TMA: trimethylamine

TSB: trypticase soy broth

txt: thaxtomin

T1PKS: type one polyketide synthase

VC: vector control.

WAG+G: Whelan and Goldman plus gamma

YMS: yeast extract-malt extract-starch

YMS_m: YMS modified

Δ: deletion

ΔΔΔ: $\Delta mlp_{lipo}/\Delta txtH/\Delta mlp_{sca}$

CHAPTER 1: Introduction and Overview

1.1 General Features of *Streptomyces*

There are over 900 described species of *Streptomyces*, which is a genus of Gram-positive bacteria from the phylum Actinobacteria (Hwang et al., 2019). These bacteria are filamentous, obligate aerobes found abundantly in heterogeneous terrestrial niches (Hodgson, 2000), though some have also been isolated from marine sources, such as fish, sponges, seaweeds and sediment samples (Dharmaraj, 2010; Pathom-aree et al., 2006). In these diverse niches, most *Streptomyces* spp. live as saprophytes by degrading and feeding on decaying organic materials, while some species are plant and animal pathogens (Flärdh and Buttner, 2009).

Streptomyces species have high GC content genomes composed of a single linear chromosome, the size of which can vary from 6 to 11 Mb depending on the species (Hopwood, 2006; Wang et al., 2010; Zaburannyi et al., 2014). Many members of the *Streptomyces* also contain large linear and/or circular plasmids (Ventura et al., 2007). *Streptomyces* chromosomes are known to be unstable due to abundant transposable elements in the terminal regions (Chen et al., 2002). Moreover, genes involved in DNA duplication and horizontal gene transfer are commonly found in *Streptomyces* genomes, and this might contribute to the genomic diversification that is observed within the genus (Zhou et al., 2012). A notable attribute of the *Streptomyces* genome is the presence of a large number (30 on average) of biosynthetic gene clusters (BGCs) for the production of specialized metabolites (Lee et al., 2019). This reflects the tremendous metabolic potential of these bacteria, which will be discussed in greater detail later in the section 1.3.

1.2 The Life Cycle of *Streptomyces* spp.

A feature that distinguishes the *Streptomyces* from other bacteria is the ability to undergo morphological differentiation in a manner similar to that of filamentous fungi (Chater, 2006). The life cycle (Figure 1.1) initiates with a dormant spore germinating when *Streptomyces* encounter a suitable source of nutrients in the surrounding environment. One or more germ tubes emerge and grow by tip extension and branching to form a dense network of hyphae called the vegetative or substrate mycelium (Flärdh and Buttner, 2009). Septation rarely takes place in the vegetative hyphae; rather, cell wall synthesis occurs at the hyphal tips and chromosomes replicate without cell division, giving rise to individual cells with long compartments containing multiple chromosomes (Elliot et al., 2007; Flärdh and Buttner, 2009; Flärdh et al., 2012). In response to nutrient starvation or other signals, such as a downshift in nitrogen and carbon sources as well as guanosine triphosphate (GTP), *Streptomyces* initiate morphological differentiation (Hamedi et al., 2017). At this stage of life, nonbranching aerial hyphae are formed that grow up into the air away from the vegetative hyphae (Flärdh and Buttner, 2009). At the same time, the vegetative hyphae undergo autolysis and programmed cell death to release cellular components as surrogates for depleted nutrients, and specialized metabolism is also initiated (Filippova and Vinogradova, 2017; Flärdh and Buttner, 2009; Miguélez et al., 2000; Yagüe et al., 2012). Once the aerial hyphae obtain enough biomass through cell elongation, they undergo a single, synchronous round of septation and chromosomal segregation. The resulting "prespore" compartments, each containing one copy of the chromosome, continue to develop into mature spores, culminating with the accumulation of a gray pigment on the

spore surface (Flärdh and Buttner, 2009; Hamedi et al., 2017; McCormick and Flärdh, 2012). Finally, the dormant spores are dispersed to new environments, where they will germinate when they encounter suitable conditions (Flärdh and Buttner, 2009).

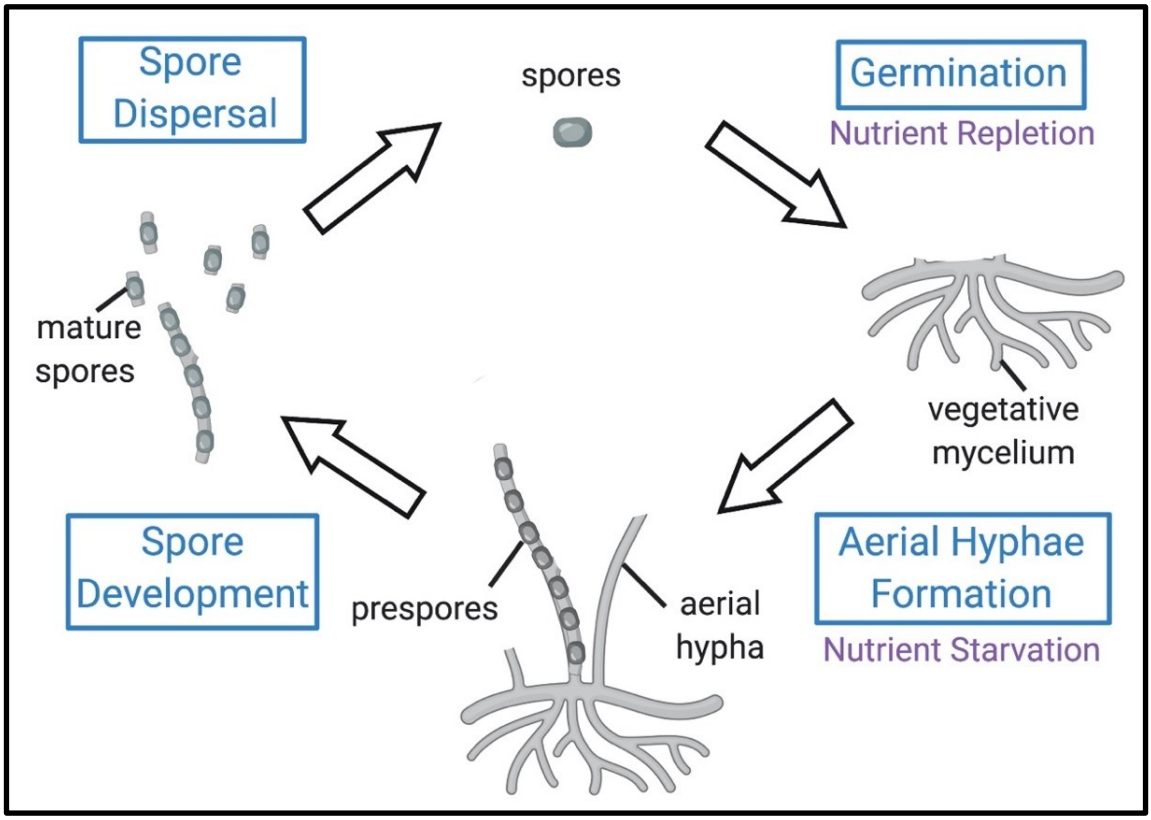


Figure 1.1 The life cycle of *Streptomyces* species.

Recently, a new form of growth called "exploratory growth" has been described for some *Streptomyces* species (reviewed by Jones and Elliot, 2017). Exploratory growth is named after the ability of nonbranching vegetative hyphae called "explorer cells" to rapidly traverse biotic and abiotic surfaces, and it was first discovered when *Streptomyces venezuelae* and other species were co-cultured with the yeast *Saccharomyces cerevisiae* (Jones et al., 2017). Exploratory growth of *S. venezuelae* is initiated by depletion of glucose

in the medium in the immediate vicinity of the *S. venezuelae* and yeast colonies, and by production of the volatile compound trimethylamine (TMA) by *S. venezuelae*, which elevates the pH of the medium. Subsequent studies showed that the elevated pH caused by TMA production reduces iron solubility and availability, and that exploration by *S. venezuelae* can be enhanced by iron depletion by other microorganisms or by the addition of iron chelators to the culture medium, suggesting that low iron availability is the actual trigger for exploratory growth (Jones et al., 2019). It is noteworthy that in a screen of different *Streptomyces* spp., only ~10% were found to exhibit exploratory growth under laboratory conditions (Jones et al., 2017), and a study from our lab was unable to demonstrate this mode of growth for a number of different species (O'Neill, 2019). Thus, it remains unclear how widespread this mode of growth is within the genus.

1.3 Specialized Metabolism in *Streptomyces*

The capability to produce numerous specialized metabolites (also known as secondary metabolites) is a hallmark of the genus *Streptomyces*. Specialized metabolites are small (MW < 3000 Da), chemically diverse molecules that are associated with various biological activities (Berdy, 2005). Specialized metabolites produced by microorganisms include carbohydrates, lipids, peptides, polyketides, steroids, alkaloids, and terpenoids (O'Brien and Wright, 2011), and the building blocks for these molecules are often derived from primary metabolism (Fernández-Martínez and Hoskisson, 2019). Unlike primary metabolites, specialized metabolites are not required for cell growth and survival under laboratory conditions, but instead are thought to function as molecules of adaptation that

confer a selective advantage to the producing organism in its specific ecological niche (Berdy, 2005; Keulen and Dyson, 2014; O'Brien and Wright, 2011). *Streptomyces* bacteria are recognized as a source of many clinically important specialized metabolites, including antibiotics with antibacterial, antifungal, antiviral and antiparasitic activities, as well as antitumor and immunosuppressive agents. In addition, *Streptomyces* spp. can produce compounds with useful applications in agriculture such as insecticides, pesticides and herbicides (Harir et al., 2018). Over the years, much of the research on *Streptomyces* specialized metabolites has focused on the identification of new molecules and their applications rather than on the role of these molecules for the producing organisms. While the production of compounds with antimicrobial activity is consistent with the viewpoint that these molecules can function as agents of inter-microbial warfare to provide an advantage for the producer in nutrient poor environments, there is currently little evidence that these molecules are produced in natural environments at levels that would be inhibitory to other microbes (O'Brien and Wright, 2011).

Other proposed biological roles for specialized metabolites include them functioning as signaling molecules for intra- and inter-generic communication with other microorganisms, or serving as mediators of symbiotic interactions with eukaryotes such as animals, fungi and plants (Berdy, 2005; O'Brien and Wright, 2011; Seipke et al., 2012; Yim et al., 2007). As an example of the latter, a *Streptomyces* strain named '*Candidatus* Streptomyces philanthi' has been found to maintain a mutualistic symbiotic relationship with beewolf digger wasps (*Philanthus* spp.). This endosymbiotic *Streptomyces* is cultivated in specialized antennal glands of female wasps and is transmitted to the brood cell of insect larvae and later to their cocoons, where the *Streptomyces* provides protection

against a broad range of entomopathogenic fungi and bacteria by producing a "cocktail" of different bioactive compounds (Kaltenpoth et al., 2006, 2010; Kroiss et al., 2010). In turn, the *Streptomyces* engaging in the mutualistic interaction can exploit energy sources present on the cuticle or in the excretions of the insects, and this enables the bacteria to survive under unfavourable environmental conditions (Kaltenpoth et al., 2010). Marine bacteria, including *Streptomyces* species, are also found to be involved in stable symbiotic associations with sponges and can take up a large portion of the sponge biomass (Selvin et al., 2009; Taylor et al., 2007; Webster and Blackall, 2008). *Streptomyces* isolates from marine sponges are reservoirs of bioactive compounds with antibacterial and antifungal activities (Dharmaraj and Sumantha, 2009), and it is reasonable to speculate that *Streptomyces* spp. associated with these marine animals can offer protection against pathogenic microorganisms (Seipke et al., 2012). Although many *Streptomyces* are thought to exhibit antagonistic interactions with fungi by producing antifungal compounds or secreting chitinolytic enzymes, some species are known to promote the growth of fungi, especially mycorrhizal fungi (Seipke et al., 2012). For example, Reidlinger and colleagues noticed improved mycelial growth and mycorrhization rate in the fly agaric (*Amanita muscaria*) when it was co-cultured with *Streptomyces* sp. AcH 505, and this was attributed to the production of a novel metabolite called auxofuran by the *Streptomyces* (Riedlinger et al., 2006). Additionally, a wide variety of *Streptomyces* spp. are known to establish beneficial symbiotic relationships with plants, and the specialized metabolites produced by these bacteria may have evolved to mediate such interactions. For example, the plant growth-promoting *Streptomyces* (PGPS) are able to produce phytohormones and iron-chelating siderophores that promote growth and improve the fitness of plant hosts

(Olanrewaju and Babalola, 2019; Sadeghi et al., 2012). Antimicrobial compounds produced by PGPS also provide protection to plant hosts from bacterial and fungal phytopathogens (Park et al., 2011; Wang et al., 2008). However, not all plant-associated *Streptomyces* confer beneficial traits to the host, as a small number of species are pathogenic to plants (phytopathogenic) and contribute to the development of disease in their host.

1.4 Phytopathogenic *Streptomyces* and Common Scab Disease

The success of phytopathogenic *Streptomyces* spp. is due in part to their filamentous lifestyle that enables the bacteria to colonize and subsequently penetrate plant tissues in order to gain entry into the host (Bignell et al., 2010; Loria et al., 2003). The most important disease caused by phytopathogenic *Streptomyces* spp. is common scab (CS) of potato (*Solanum tuberosum* L.), which is characterized by the formation of brown, corky-like lesions on the surface of potato tubers (Figure 1.2). The lesions can be superficial, erumpent or they can extend deep into the tuber tissue, and this is thought to depend on factors such as the pathogen aggressiveness, host susceptibility and environmental conditions (Dees and Wanner, 2012). CS is limited to rapidly expanding plant tissues, and thus once tuber expansion in the soil has ceased, potato plants are no longer susceptible to the disease, and lesion development stops (Loria et al., 2006). This is likely due to the mode of action of the thaxtomin A phytotoxin, which will be discussed later in section 1.5.2. The above ground tissues are normally not affected and remain healthy, unless the transfer of water or nutrients between plant roots and stems is hindered (Dees and Wanner, 2012). CS is mainly considered a “cosmetic disease” due to the reduced quality and market value of the affected potato crop (Wanner and Kirk, 2015). Losses due to CS are not well

documented, though a study by Hill and Lazarovitz estimated that potato growers in Canada lost between $15.3 and 17.3 million Canadian dollars due to CS during the 2002 growing season (Hill and Lazarovits, 2005). There is also evidence from greenhouse studies that CS can reduce the size of the tubers produced as well as the overall yield of the crop (Hiltunen et al., 2005; Wanner and Kirk, 2015). CS pathogens have a broad host range and can also cause scab disease symptoms on taproot crops such as beet, radish, carrot and parsnip (Goyer and Beaulieu, 1997). In addition, seedlings of both monocot and dicot plants can be infected by CS-causing pathogens in laboratory studies, resulting in root and shoot stunting, root swelling and tissue necrosis (Loria et al., 1997).

Figure 1.2 Common scab disease symptoms on a potato tuber. Image courtesy of Dawn R.D. Bignell.

Pathogenic *Streptomyces* species associated with CS disease include *Streptomyces scabiei* (syn. *S. scabies*), *Streptomyces turgidiscabies*, *Streptomyces stelliscabiei*, *Streptomyces niveiscabiei*, *Streptomyces luridiscabiei*, *Streptomyces puniciscabiei,* and *Streptomyces europaeiscabiei.* Of these species, *S. scabiei* was the first described CS pathogen and is the best-characterized, and it has a worldwide distribution (Bignell et al., 2014b). Another pathogen, *Streptomyces acidiscabies,* is associated with a related disease called acid scab (AS), which is identical to CS except that it occurs in low pH (pH $\leqslant$5.2) soils where the growth of CS pathogens is suppressed (Loria et al., 1997).

CS disease is difficult to control as there are currently no strategies that can effectively and consistently manage the disease (Wanner and Kirk, 2015). Reduction of soil pH is one of the traditional strategies used for CS disease control, but this strategy has limited application as many crops do not tolerate low pH soil conditions very well, and it can promote the development of AS by *S. acidiscabies* (Loria et al., 1997). Field irrigation during tuber initiation and expansion has also been used to manage CS disease since high moisture condition has been shown to lessen the disease symptom development (Lapwood and Hering, 1970). However, this strategy often fails and can promote the development of other diseases (Adams et al., 1987). Chemical fumigation is among the main strategies used for the control of soil-borne diseases, including CS, but it is very costly and is not ecofriendly (Dees and Wanner, 2012). The most desired method to control this disease is the use of resistant potato cultivars, but the cultivars currently available exhibit varying degrees of tolerance to CS, and no cultivars show complete resistance to the disease (Hiltunen et al., 2005). The utilization of biological control agents is considered a

promising alternative to traditional physiochemical approaches. A successful biocontrol agent, *Streptomyces violaceusniger* AC12AB, which produces the plant hormone indole-3-acetic acid (IAA) and the antimicrobial antibiotic azalomycin, was shown to exhibit a dual action of promoting potato growth and alleviating disease severity caused by *S. scabiei* strains both in greenhouse and under field conditions (Sarwar et al., 2019). In addition, biological control of CS disease has been demonstrated using the plant beneficial bacterium *Pseudomonas synxantha* (formerly *Pseudomonas fluorescens*) LBUM223 both under controlled and natural field conditions (Arseneault et al., 2013, 2015). The production of phenazine-1-carboxylic acid (PCA) by *P. synxantha* LBUM223 is considered an essential factor in its ability to reduce CS symptoms (Arseneault et al., 2013). A recent whole-transcriptome analysis of *S. scabiei* exposed to *P. synxantha* LBUM223 or PCA revealed that genes involved in several cellular processes including virulence, mycelia formation, siderophore production and oxidative stress in *S. scabiei* were differentially expressed (Arseneault et al., 2020). However, the exact mode of action of these biocontrol agents, especially in natural soil environments, remains poorly understood and requires further investigation.

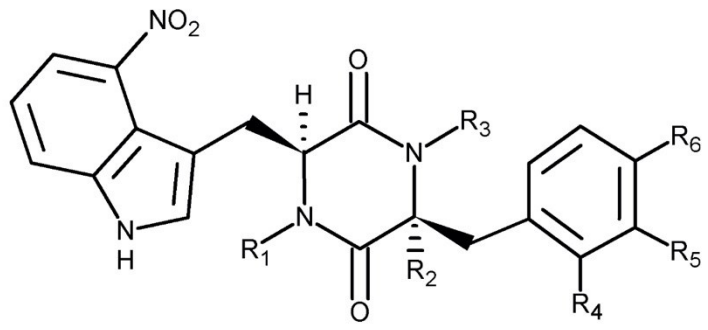
1.5 Virulence Factors in Plant Pathogenic *Streptomyces*

The ability of plant pathogenic *Streptomyces* species to cause disease is dependent on the production of virulence factors that enable host colonization and disease symptom development. A number of known or putative virulence factors are produced by pathogenic *Streptomyces* species, including phytotoxins, secreted proteins and phytohormones, and are

reviewed elsewhere (Li et al., 2019). Here, the principle pathogenicity determinant associated with CS disease will be discussed in detail in the following sections.

1.5.1 General features of thaxtomins

Thaxtomins are a family of cyclic dipeptides (2,5-diketopiperazines) that are derived from the condensation of L-phenylalanine and 4-nitro-L-tryptophan (Bignell et al., 2014b; King and Calhoun, 2009). Eleven different analogues have been characterized, and they vary in the presence or absence of hydroxyl and/or *N*-methyl groups at specific locations on the thaxtomin backbone (Figure 1.3). Thaxtomins are produced by several different plant pathogenic *Streptomyces* species, including *S. scabiei*, *S. turgidiscabies*, *S. acidiscabies*, *S. europaeiscabiei, S. niveiscabiei* and *S. stelliscabiei* (Bignell et al., 2014b; King and Calhoun, 2009). Thaxtomin A is the most prominent analogue produced by these organisms, though *S. scabiei* and possibly other species can produce other derivatives in low amounts in infected plant tissues (King and Calhoun, 2009). Thaxtomin C, a nonhydroxylated thaxtomin analogue (Figure 1.3), is main the metabolite produced by *Streptomyces ipomoeae,* which is the causative agent of sweet potato soil rot disease (King et al., 1994).



	R1	R2	R3	R4	R5	R6
Thaxtomin A	CH_3	OH	CH_3	H	OH	H
Thaxtomin B	CH_3	OH	CH_3	H	H	H
Thaxtomin C	CH_3	H	H	H	H	H
Thaxtomin D	CH_3	H	CH_3	H	H	H
Others	CH_3	OH	CH_3	OH	H	H
	CH_3	OH	CH_3	H	H	OH
	CH_3	OH	CH_3	H	OH	OH
	CH_3	OH	H	H	OH	H
	CH_3	OH	H	H	H	H
	H	OH	CH_3	H	OH	H
	H	H	H	H	H	H

Figure 1.3 Chemical structures of thaxtomin A, B, C, D and other naturally occurring analogues.

1.5.2 Biological activities of thaxtomins

The involvement of specialized metabolites in *Streptomyces* plant pathogenicity was first established when thaxtomin A and B were isolated and characterized from scab lesions of potato tubers infected by *S. scabiei* (King et al. 1989). A positive correlation between the pathogenicity of *Streptomyces* species and their ability to produce thaxtomin A was subsequently reported by several research groups (Goyer et al., 1998; Healy et al., 2000; King et al. 1991; Kinkel et al., 1998; Loria et al., 1995). Thaxtomins have the ability

to induce necrosis on excised potato tuber tissue (Loria et al. 2006) and to cause scab-like lesions on excised minitubers (Lawrence et al. 1990). In addition, monocot and dicot seedlings treated with nanomolar concentrations of thaxtomin A exhibit stunting, hypocotyl and root swelling, cell hypertrophy and tissue necrosis. These symptoms resemble the seedling disease symptoms caused by *S. scabiei* and *S. acidiscabies,* suggesting that the seedling pathogenicity of these organisms is primarily mediated by thaxtomin production (Leiner et al., 1996; Loria et al., 1997).

Various physiological responses have been reported to occur in thaxtomin A-treated plants, including activated expression of several defence-responsive genes, induction of programmed cell death, a disturbance in ion flux, and production of the antimicrobial compound scopoletin (Duval et al., 2005; Errakhi et al., 2008; Lerat et al., 2009; Tegg et al., 2005; Tegg et al., 2016). In addition, Fry and Loria noted that thaxtomin A inhibits cytokinesis of onion root tip cells and disrupts normal cell elongation in tobacco protoplasts (Fry and Loria, 2002). These physiological effects imply that the plant host cell wall is a potential target of thaxtomin A, and more than one line of evidence indicates that the phytotoxin functions primarily as a cellulose synthesis inhibitor (Bischoff et al., 2009; Duval and Beaudoin, 2009; Fry and Loria, 2002; Scheible et al., 2003). Scheible and colleagues demonstrated that thaxtomin A inhibits the incorporation of ^{14}C-glucose into the cellulosic fraction of the *Arabidopsis thaliana* cell wall (Scheible et al., 2003). Moreover, thaxtomin A depletes cellulose synthase complexes from the plasma membranes of *A. thaliana* seedlings, and it affects the expression of cell wall synthesis genes in a manner reminiscent of the known cellulose synthesis inhibitor isoxaben (Bischoff et al., 2009). *A. thaliana* cells have been shown to display a similar transcriptional pattern in response to

thaxtomin A and isoxaben, suggesting a common mode of action between the two (Duval and Beaudoin, 2009). However, the mode of action of thaxtomin A in plant cells is not fully understood, and its direct molecular target(s) remains to be discovered (Li et al. 2019).

1.5.3 Thaxtomin biosynthesis

The ability to produce thaxtomins is a unique feature of plant pathogenic *Streptomyces* spp., and the biosynthetic gene cluster (Figure 1.4) is highly conserved in these species (Huguet-Tapia et al., 2016). In scab-causing pathogens such as *S. scabiei*, *S. acidiscabies* and *S. turgidiscabies*, the thaxtomin (Txt) biosynthetic gene cluster contains seven genes, of which *txtA*, *txtB*, *txtC*, *txtD* and *txtE* encode biosynthetic proteins, *txtR* encodes a pathway-specific regulator, and *txtH* encodes a predicted member of the MbtH-like protein family (Bignell et al., 2014b; Huguet-Tapia et al. 2016; Zhang et al., 2016). The biosynthesis of thaxtomin A (Figure 1.4) commences from the production of nitric oxide (NO) from L-arginine by the nitric oxide synthase TxtD (Kers et al., 2004). Then, a novel cytochrome P450 monooxygenase, TxtE, nitrates L-tryptophan using the NO produced by TxtD, generating the intermediate 4-nitro-L-tryptophan (Barry et al., 2012). Two non-ribosomal peptide synthetases (NRPSs), TxtA and TxtB, utilize L-phenylalanine and 4-nitro-L-tryptophan, respectively, to produce the cyclic dipeptide intermediate thaxtomin D (Jiang et al. 2018; Johnson et al., 2009). TxtA and TxtB are megasynthetases that consist of three core enzymatic domains (Figure 1.4) that catalyze the adenylation, thiolation and condensation reactions, as well as a methylation domain that adds the *N*-methyl groups onto both the nitrotryptophyl and phenylalanyl moieties of thaxtomin D

(Healy et al., 2000; Johnson et al., 2009). At the final step of thaxtomin A biosynthesis, two hydroxyl groups are introduced onto the phenylalanyl moiety of thaxtomin D by the cytochrome P450 monooxygenase TxtC (Alkhalaf et al., 2019; Healy et al. 2002). The function of TxtH in the thaxtomin biosynthetic pathway is the focus of the current thesis, and a detailed discussion about MbtH-like proteins and their roles in NRP biosynthesis is provided in section 1.6.3.

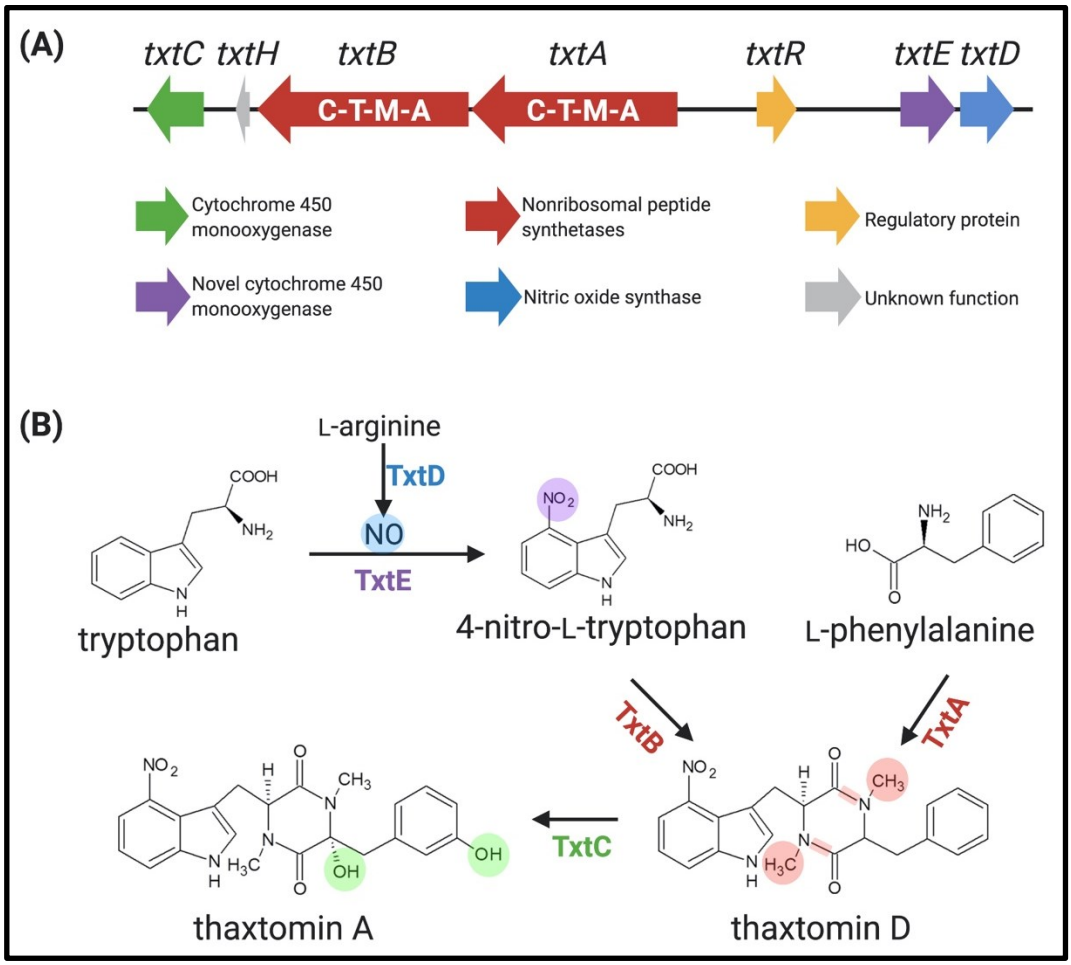


Figure 1.4 (A) The thaxtomin (Txt) biosynthetic gene cluster in *S. scabiei*. Biosynthetic genes *txtC* (green), *txtD* (blue), *txtE* (purple), and the regulatory gene *txtR* (yellow) are indicated in arrows with different colors. The NRPSs encoded by the *txtA* and *txtB* genes (red) consist of four enzymatic domains: adenylation (A-) domain, methylation (M-)

domain, thiolation (T-) domain and condensation (C-) domain. A gene of unknown function is highlighted in gray color. The direction of the arrow indicating the direction of transcription. (B) The proposed thaxtomin A biosynthetic pathway in *S. scabiei*. Modification by each biosynthetic gene is highlighted in each step. Nitric oxide is indicated by NO.

1.5.4 Regulation of thaxtomin production

The *txtR* gene (Figure 1.4) that is embedded in the *txt* gene cluster in *S. scabiei*, *S. turgidiscabies*, *S. acidiscabies* and *S. ipomoeae* encodes an AraC/XylS family transcriptional regulator (Guan et al., 2012; Joshi et al., 2007). Proteins belonging to this family typically contain two helix-turn-helix DNA-binding motifs and regulate processes involved in regulation of metabolic processes, adaptation responses, stress response and pathogenesis (Egan, 2002; Gallegos et al., 1997; Ibarra et al., 2008). Deletion of *txtR* in *S. scabiei* reduces the transcriptional levels of *txtA*, *txtB*, *txtC* and *txtD* and nearly eliminates thaxtomin production, and the virulence phenotype of the mutant is severely compromised as compared to the wild-type strain (Joshi et al., 2007). Additionally, the *txtR* gene contains a single TTA codon (Figure 1.5), which is a rare codon in the GC-rich genomes of *Streptomyces* species (Bignell et al., 2014a). The corresponding UUA codon in *Streptomyces* mRNA can only be efficiently translated by the leucyl-tRNA encoded by the *bldA* gene (Hackl and Bechthold, 2015), and loss of *bldA* was shown to reduce the expression of the *txt* biosynthetic genes and eliminate thaxtomin A production in *S. scabiei* (Bignell et al., 2014a). Four other genes (*bldC*, *bldD*, *bldG*, *bldH*) belonging to the *bld* (bald) gene family of global regulators are also known to modulate *txt* biosynthetic gene expression and the production of thaxtomin A (Bignell et al., 2014a).

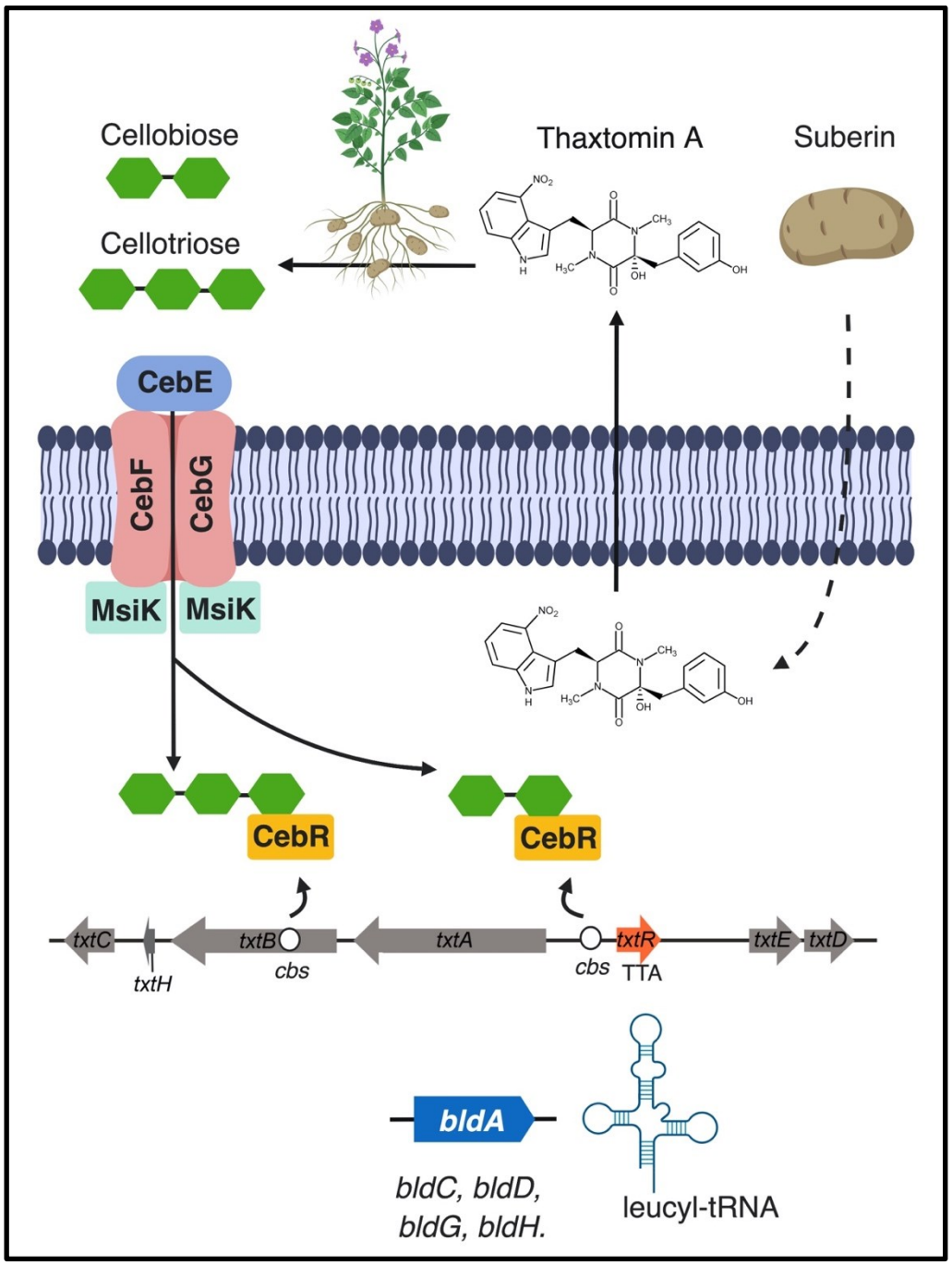


Figure 1.5 Model of thaxtomin A regulation in *S. scabiei*. Cello-oligosaccharides are predicted to be released from expanding cell walls and are transported into *S. scabiei* through the CebEFG-MsiK ABC transporter. Once inside the cell, the cello-oligosaccharides bind to CebR, and this results in the derepression of the *txt* genes by causing the release of CebR from the *cbs* (CebR-binding site, indicated by circles). Expression of the *txt* genes leads to the production of thaxtomin A, which then inhibits cellulose biosynthesis in expanding plant tissues and causes the release of more cellotriose,

thereby creating a positive feedback loop for thaxtomin A biosynthesis. The main component of the potato periderm, suberin, is also known to stimulate thaxtomin A production, presumably by affecting specialized metabolism in *S. scabiei*. In addition, the expression of *txtR* is regulated by the *bldA* tRNA, which is required for translation of the TTA codon within the *txtR* coding sequence, as well as by the transcriptional regulators *bldC*, *bldD*, *bldG* and *bldH*.

With the evidence suggesting that thaxtomin A targets cellulose biosynthesis in plants, it has been proposed that cello-oligosaccharides such as cellobiose and cellotriose may work as environmental cues for sensing the presence of expanding plant tissues that serve as infection sites for scab-causing *Streptomyces* species (Li et al., 2019). Both cellobiose and cellotriose are known inducers of *txt* biosynthetic gene expression and thaxtomin A production (Johnson et al., 2007). In addition, treatment with pure thaxtomin A causes the release of cellotriose from rapidly growing plant tissues and tobacco cell suspensions, suggesting that exposure to thaxtomin A creates a positive feedback loop (Figure 1.5) for the production of thaxtomin (Johnson et al., 2007).

More recently, the expression of *txtR* was shown to be controlled by CebR, which has been identified as a repressor of cello-oligosaccharides utilization in the non-pathogenic *Streptomyces reticuli* (Francis et al., 2015). CebR-binding sites (Figure 1.5) were identified upstream of *txtR* as well as within the *txtB* gene in *S. scabiei* and in other thaxtomin-producing *Streptomyces* pathogens. The transport of cellotriose and cellobiose into the cell (Figure 1.5) is facilitated by the CebEFG-MsiK ABC transporter (Jourdan et al., 2016). Once inside, the cello-oligosaccharides bind to CebR and induce the release of the repressor from its DNA targets (Figure 1.5), thus turning on the expression of the *txt* biosynthetic and regulatory genes (Francis et al., 2015). Deletion of *cebR* leads to overexpression of the *txt* genes and constitutive thaxtomin A production in the absence of cellobiose, and the mutant

exhibits a hypervirulence phenotype when compared to the wild-type strain (Francis et al., 2015). Together, these results suggest that CebR may serve as the "gatekeeper" of pathogenicity in scab-causing *Streptomyces* species.

Suberin, a major component of the potato periderm, is characterized as another inducer of thaxtomin production in *S. scabiei* (Komeil et al., 2013). It has been shown that *S. scabiei* grown on minimal medium containing cellobiose produces only a small amount of thaxtomin A, whereas supplementation with both cellobiose and suberin induces the expression of the *txt* biosynthetic genes and significantly enhances thaxtomin biosynthesis (Lerat et al., 2010). This suggests that suberin and cellobiose promote thaxtomin production in a synergistic way (Figure 1.5). Padilla-Reynaud and colleagues found that the production of glycosyl hydrolases such as cellulases in *S. scabiei* is enhanced by the addition of suberin to cellulose-containing medium (Padilla-Reynaud et al., 2015). Thus, it has been proposed that the pathogenicity of *S. scabiei* requires both suberin and cellulose in order to generate the cello-oligosaccharides that function as inducers of thaxtomin biosynthesis. In addition, the stimulation of thaxtomin production by suberin is thought to result from the induction of specialized metabolism in general through an unknown mechanism (Lerat et al., 2012).

1.6 Non-Ribosomal Peptide Synthetases (NRPSs)

1.6.1 General features of NRPSs

Unlike ribosomal peptide synthesis, which is restricted to the use of the 20 proteinogenic amino acids as building blocks, non-ribosomal peptide (NRP) synthesis can incorporate a variety of different amino acid substrates (eg β-amino acids, hydroxy amino

acids, methyl amino acids, halogenated amino acids) and generate peptides with tremendous structural and functional diversity (Finking and Marahiel, 2004; Marahiel et al., 1997; Sieber and Marahiel, 2005; Süssmuth and Mainz, 2017). NRPs are mainly produced by bacteria within the phyla Actinobacteria, Firmicutes, Proteobacteria, and Cyanobacteria; and by fungi belonging to the phylum Ascomycota (Süssmuth and Mainz, 2017). The complexity of NRP structures reflects the versatility of their biological activities, and they have been exploited in the development of important therapeutic agents, including the antibiotic daptomycin (Tedesco et al., 2003), the immunosuppressive agent cyclosporin A (Weber et al., 1994) and the anticancer drug bleomycin A2 (Du et al., 2000). Additionally, NRPs can have an important function in the producing organism; for instance, iron-chelation carried out by siderophores such as enterobactin (Gehring et al., 1998), and plant tissue colonization mediated by phytotoxins such as thaxtomin A (Li et al., 2019).

The synthesis of NRPs are orchestrated by large multienzyme machineries known as non-ribosomal peptide synthetases (NRPSs) that are arranged in modules, each of which incorporates one amino acid into the growing polypeptide chain. The modules can be subsequently dissected into independent enzymatic domains including three core domains: an adenylation (A-) domain, a thiolation (T-) domain and a condensation (C-) domain (Finking and Marahiel, 2004; Süssmuth and Mainz, 2017). The A-domain is responsible for selecting and activating the amino acid substrate that is then loaded onto the T-domain, also referred to as a peptidyl carrier protein (PCP-) domain. The activated amino acid is covalently bound to the 4′-phosphopantetheine (Ppant) prosthetic group of the T-domain, which facilitates the shuttling of building blocks between different catalytic domains. The

C-domain catalyzes the coupling of the amino acyl substrates bound to the T-domains from adjacent modules, generating a peptide bond between the amino acids. Canonical organization of the domains within an initiation module that incorporates the first amino acid is A-T at the minimum, and the subsequent condensation reaction is then carried out by the adjacent elongation module which harbours a C-A-T domain arrangement (Finking and Marahiel, 2004; Süssmuth and Mainz, 2017). Besides the core domains essential for producing the NRP backbone, many tailoring domains are able to catalyze reactions such as methylation, epimerization, formylation, heterocyclization, reduction and oxidation, thereby incorporating additional modifications to the NRP product. Finally, NRP synthesis is terminated by a thioesterase (Te-) domain that releases the mature oligopeptide from the last module on NRPS machinery. Fungal NRPSs employ an alternative strategy for product release, in which the Te-domain can be functionally replaced by a terminal C-domain for macrocyclization during the detachment process (Süssmuth and Mainz, 2017).

1.6.2 Adenylation domain

The synthesis of NRPs begins with the selection and activation of the amino acid substrate, which is catalyzed by the A-domain. This domain belongs to the ANL (**A**cyl-CoA synthetases, **N**RPS adenylation domains, and **L**uciferase enzymes) superfamily of adenylating enzymes (Gulick, 2009). Crystal structures of various A-domains have been solved, including the phenylalanine-activating A-domain of the gramicidin S synthetase A (PheA, also known as GrsA) from *Bacillus brevis*, and the 2,3-dihydroxybenzoate (DHB) activating A-domain (DhbE) from *B. subtilis* (Conti et al., 1997; May et al., 2002).

Although the amino acid sequences of A-domains share low identity, the overall crystal structures display a high level of similarity (Sieber and Marahiel, 2005). Typically, an A-domain is comprised of a N-terminal core domain (A_{core}, approximately 50 kDa) and a C-terminal subdomain (A_{sub}, approximately 10 kDa), which are linked by a hinge region of five residues (Conti et al., 1997). The reaction catalyzed by A-domains can be divided into two steps (Figure 1.6): (1) at the interface of the A_{core}- and A_{sub}-domain, an amino acid substrate is recruited at the expense of Mg^{2+} and ATP to form a high-energy aminoacyl adenylate species, which provides the activation energy for the second partial reaction, and (2) after pyrophosphate (PP_i) is released, the A_{sub}-domain undergoes a 140° rotation with respect to the A_{core}-domain, and this not only stops the adenylation reaction but also serves as a flexible hinge to facilitate the relocation of the downstream T-domain such that its Ppant (4′- phosphopantetheine) arm can approach the aminoacyl-AMP intermediate. Then, the A-domain catalyzes the second half reaction through a nucleophilic attack of a free thiol on the Ppant arm, resulting in the covalent attachment of the activated amino acid to the T-domain as thioester and the release of AMP. Once the substrate bound T-domain travels to the acceptor site of the C-domain, the A_{sub}-domain switches back to its adenylation state (Gulick, 2009; Sieber and Marahiel, 2005; Süssmuth and Mainz, 2017).

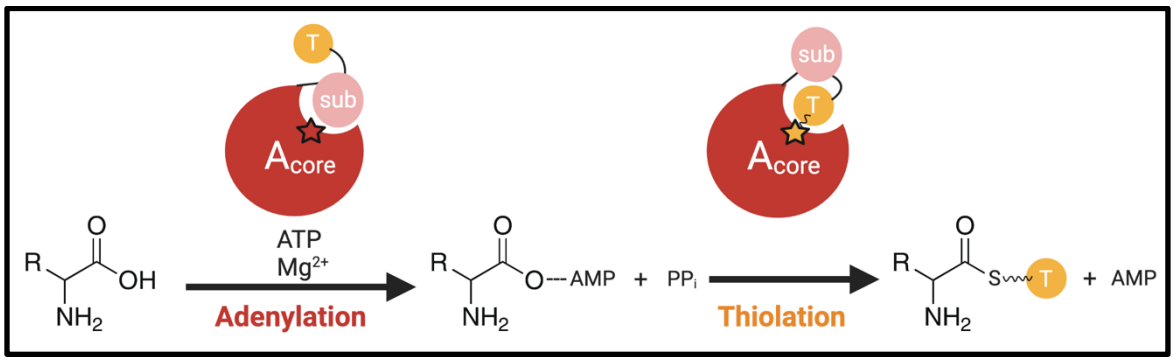


Figure 1.6 Reactions catalyzed by the NRPS adenylation (A-) domain. In the first half reaction (Adenylation), an amino acid substrate is activated by reacting with Mg^{2+} and ATP to form an aminoacyl-AMP intermediate and inorganic PP_i. In the second half reaction (Thiolation), the A_{sub}-domain undergoes a 140° rotation, which enables covalent attachment of the activated amino acid to the Ppant (4′-phosphopantetheine, represented by the squiggly line) arm of the thiolation (T-) domain with the release of AMP. The thiolation reaction is catalyzed by the A-domain using the same catalytic pocket as in the adenylation reaction. The active site of the A-domain for substrate binding is indicated by the red star at the interface between the A_{core}- and A_{sub}-domains. The Ppant attachment site of the T-domain is indicated by yellow star.

Ten sequence motifs (A1-10) are highly conserved in A-domains within the ANL superfamily, and these consensus motifs serve a variety of functions including structural (A1, A2, A5, A6 and A8), substrate binding (A3-A5, A7 and A8) and catalytic (A7, A9 and A10) roles (reviewed by Gulick, 2009; Labby et al., 2015). Additionally, an LPxP motif downstream of the A10 motif is conserved in some NRPSs at the linker region joining the A- and T-domains. This motif appears to have two different functions during the NRPS catalytic cycle. The LPxP motif interacts with a region on the A_{sub}-domain to stabilize the catalytic A10 motif for the adenylation reaction. This interaction also coordinates the movement of the T-domain with the rotation of the A_{sub}-domain when the enzyme adopts the thioester-forming conformation (Miller et al., 2014). The determination of crystal structures of members of this enzyme family have rationalized the roles of catalytic residues

during the adenylation reaction. In the structure of PheA (Figure 1.7), a strictly conserved aspartic acid residue (D413 of A7 motif) binds the hydroxyl groups on the ribose moiety of ATP. Additionally, the aromatic residue of tyrosine (Y323 of A5 motif) stacks against the adenine ring of ATP, and a highly conserved glutamic acid (E327 of A5 motif) binds to the Mg^{2+} ion (Conti et al., 1997; Gulick, 2009). Moreover, a strictly conserved lysine (K517 of A10 motif) and a highly conserved aspartic acid (D235 of A4 motif) stabilize the amino and carboxylate moieties, respectively, of the amino acid substrate, and the former also forms specific interactions with the AMP phosphate (Stachelhaus et al., 1999). Ten amino acids (Figure 1.7) located in between the A4-A5 motifs and within the A10 motif line the substrate-binding pocket and are the major determinants of the substrate specificity of A-domains (Eppelmann et al., 2002; Stachelhaus et al., 1999). Based on the sequence of these binding pocket constituents (referred to as the signature sequence), bioinformatics tools have been developed to enable prediction of the substrate specificity of an unknown A-domain (Challis et al., 2000; Conti et al., 1997; Stachelhaus et al., 1999). Furthermore, this selectivity-conferring code has been exploited to rationally reprogram the specificity of the NRPS A-domain using *in vitro* and *in vivo* approaches (Eppelmann et al., 2002; Thirlway et al., 2012).

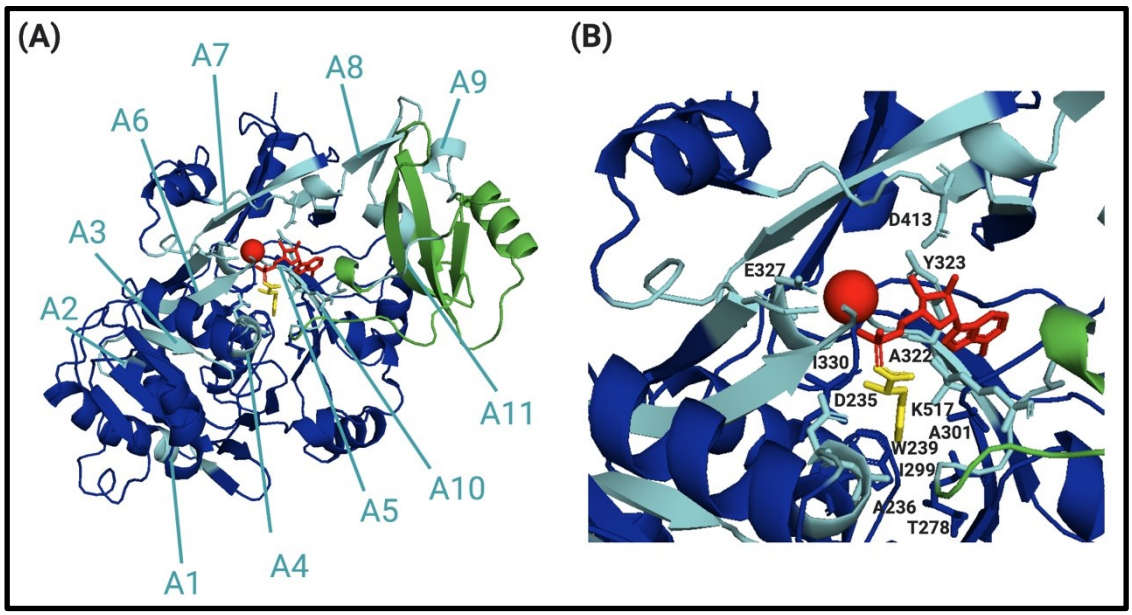


Figure 1.7 (A) Structure of the phenylalanine activating A-domain PheA (PDB ID: 1AMU, Conti et al., 1997). The N-terminal A_{core}-domain comprising residues 17-428 is indicated in dark blue, and the C-terminal A_{sub}-domain (residues 429–530) is indicated in green. The AMP (red stick), Mg^{2+} (red ball) and phenylalanine (yellow stick) are bound at the interface between the A_{core}- and A_{sub}-domains. The location of the eleven highly conserved motifs (A1-A11) are indicated in light blue. (B) Closer look of the structure showing the residues (light and dark blue sticks) lining the substrate binding pocket. D413 and E327 bind to the ribose hydroxyls of ATP and Mg^{2+} ion, respectively. The residues that function as the major determinants of the substrate specificity of PheA and which form the signature sequence of the domain are labeled.

1.6.3 MbtH-like proteins

Some NRPSs require accessory proteins for proper function, including a group of small proteins (approximately 60-70 amino acids) belonging to the MbtH-like protein (MLP) family. The name MLP stems from the MbtH protein, which is associated with the production of the virulence-conferring siderophore mycobactin in *Mycobacterium tuberculosis* (McMahon et al., 2012; Quadri et al., 1998). Homologs of MbtH are most prevalent in Actinobacteria, where the highest conservation (~3 per genome) is within the genus *Streptomyces* (Baltz, 2011). The genes encoding these proteins are normally found

within NRP BGCs (Baltz, 2011), which implies that they play a role in the production of NRPs. Indeed, several biochemical studies have demonstrated that MLPs function as chaperones to assist the proper folding of one or more of the NRPS enzymes encoded within the same BGC. In these studies, the absence of the MLP (referred to as the cognate MLP) was found to negatively impact the soluble expression of one or more NRPS components in *Escherichia coli* (Boll et al., 2011; Imker et al., 2010; Kaniusaite et al., 2020; McMahon et al., 2012; Zolova and Garneau-Tsodikova, 2012, 2014). In addition, some MLPs can modulate the adenylation activity of the associated NRPS. In these instances, the NRPS can be heterologously produced in soluble form in *E. coli* without its cognate MLP; however, the purified NRPS displays low or no adenylating activity *in vitro* unless the purified cognate MLP partner is added to the adenylation reaction, or the MLP is co-expressed with the NRPS (Al-Mestarihi et al., 2014; Boll et al., 2011; Davidsen et al., 2013; Felnagle et al., 2010; Heemstra et al., 2009; Miller et al., 2016; Schomer et al., 2018; Zhang et al., 2010). Some authors also noticed that optimal stimulation of adenylation activity by an MLP occurs when it is present in a 1:1 molar complex with the interacting A-domain (Boll et al., 2011; Davidsen et al., 2013). Together, these studies suggest MLPs can function as chaperones and/or activators for A-domains in some instances. However, the mechanism for how MLPs have these influences on the NRPSs remains elusive. Particularly, it has been reported that not all MLP-NRPS interactions are equal even within the same BGC. For instance, in the BGC producing enterobactin (ENT), mycobactin and capreomycin, at least one of the A-domains is MLP-independent despite the associated BGC containing an MLP-encoding gene (Felnagle et al., 2010; McMahon et al., 2012).

Genetic studies have revealed the importance of MLPs for the efficient production of NRPs *in vivo*, and they have demonstrated in some instances that MLPs from different biosynthetic pathways can functionally complement each other *in trans.* In these studies, the elimination of an MLP-encoding gene did not fully disrupt the production of the associated NRP, but rather production was only abolished when all of the MLP-encoding genes were removed from the genome (Lautru et al., 2007; Wolpert et al., 2007). In addition, complementation of the resulting mutant with the cognate MLP or with MLP homologues from other NRPS gene clusters (referred to as non-cognate MLPs) could restore the production of the desired NRP (Lautru et al., 2007; Wolpert et al., 2007). Other studies have demonstrated that non-cognate MLPs have the ability to promote the solubility and/or adenylation activity of NRPS enzymes at comparable or higher efficiency than the cognate MLP (Boll et al., 2011; Felnagle et al. 2010; Mori et al., 2018a; Schomer and Thomas, 2017; Zhang et al., 2010). Particularly, Schomer and Thomas (2017) utilized the *E. coli* ENT biosynthetic pathway as a model to investigate the ability of various non-cognate MLPs to compensate for the loss of the cognate MLP YbdZ in different *in vivo* and *in vitro* assays. Their results indicated that non-cognate MLPs vary in their ability to promote solubility and enzyme activity of the EntF NRPS. Furthermore, the authors noted that the level of *in vivo* complementation by non-cognate MLPs was not always correlated to the ability of the MLPs to promote solubility and activity of the NRPSs (Schomer and Thomas, 2017). These data suggest the MLP-NRPS interactions are sophisticated and require a better understanding of the specificity between the two protein partners.

Recent research has provided new insights into the roles of MLPs. For example, Schomer and Thomas (2017) reported a negative impact of a non-cognate MLP on

aminoacyl-*S*-PCP formation by EntF, which provides the first evidence of an MLP having an influence on the step in NRPS catalysis besides the adenylation reaction (Schomer and Thomas, 2017). Furthermore, Mori and colleagues showed that non-cognate MLPs can expand the substrate profile of the NRPS TioK by affecting the turnover rate of the adenylating enzyme (Mori et al., 2018a). In other studies, overexpression of cognate and non-cognate MLPs resulted in increased metabolite production in bacterial and fungal strains (Lee et al. 2016; Zwahlen et al., 2019). Interestingly, a fusion protein called LtxB, encoding an unusual cytochrome P450 monooxygenase that contains an N-terminal MLP domain, is involved in the biosynthesis of lyngbyatoxin in the cyanobacterium *Lyngbya majuscule* (Edwards and Gerwick, 2004; Huynh et al., 2010). It has been postulated that the MLP domain of LtxB may assist the tailoring enzyme in the oxygenation of the substrate during NRP assembly, however, further experimental evidence for this is required (Baltz, 2011; Süssmuth and Mainz, 2017).

1.6.4 The structures of MLPs and MLP-NRPS complexes

Structural analysis of MLPs, such as PA2412 from *Pseudomonas aeruginosa* and Rv2377c from *Mycobacterium tuberculosis*, has been conducted using X-ray crystallography and NMR spectroscopy. The structure of Rv2377c is composed of a three-stranded, anti-parallel β-sheet nestled against one C-terminal α-helix, while PA2412 has an extra two turn helix at the end of the C-terminus (Drake et al., 2007; Buchko et al., 2010). Multiple sequence alignments of various MLPs, including PA2412 and Rv2377c, has revealed a signature sequence (NxExQxSxWP-x5-PxGW-x13-L-x7-WTDxRP) that has

been used to predict functional MLP homologues in sequenced genomes (Baltz, 2011). Three tryptophan residues are universally conserved in all family members, and residues surrounding these tryptophan residues are also highly conserved. In the structure of PA2412, the strictly conserved tryptophan residues W25 and W35 are nearly parallel and form a small pocket with a distance of 7 Å and bordered by P32 and S23. The sequence surrounding W25 is the conserved motif SxWP, which is located at the center of the β2 sheet, and the sequence where W35 lies is the conserved motif PxGW that falls on the turn proceeding the β3 sheet. Between the two α-helices of PA2412, a third conserved motif around W55 is WXDXRP, in which R59 forms an ionic interaction with D57 and D68 from the second helix (Drake et al., 2007). The solution structures determined for PA2412 and Rv2377c indicate that this region is intrinsically disordered (Drake et al., 2007; Buchko et al., 2010). Disordered regions of proteins are often associated with functional diversity or with the binding to multiple protein partners (Haynes et al., 2006; Xie et al., 2007). This idea is supported by the observation that MLPs display functional redundancy and are able to exhibit cross-talk with different NRP biosynthetic pathways, but the importance of this highly conserved disordered region in MLP functional cross-talk warrants further investigation.

Several recent studies examining the structure of MLP-NRPS complexes (Herbst et al., 2013; Kreitler et al., 2019; Miller et al. 2016; Mori et al. 2018b; Tarry et al., 2017) have shed light on the interaction of the two protein partners at the molecular level. All of these structures suggest a similar interface between the MLP and the NRPS, and the residues at the interface are highly conserved. The first structure revealing the interaction interface between the two protein partners is that of the adenylating enzyme SlgN1 from

Streptomyces lydicus, which contains an MLP domain naturally tethered to the A_{core}-domain by a single turn helix linker (Herbst et al., 2013). The interface between the two domains involves residues in the α helix 11 and parts of the β strands 19 to 24 of the A_{core}-domain, and most residues within the MLP domain including the α helix 1, β strand 2 and the loop to β strand 3. Several highly conserved residues, including the strictly conserved tryptophan residues, play a critical role in the complex formation. Particularly, W25 and W35 of the MLP domain display a parallel orientation (a similar orientation has been observed in PA2412 and Rv2377c; Drake et al., 2007; Buchko et al., 2010), and this forms a small cavity coordinating the side chain of an alanine residue (A433) in the A_{core}-domain. Additionally, W35 forms a hydrogen bond with the main chain carbonyl oxygen of E442 in the A_{core}-domain. Furthermore, two key residues contributing to the interface stabilization are S23 and L24 of the MLP domain, which form hydrogen bonds with A433 and A428 of the A_{core}-domain, respectively. Mutational analysis of the S23Y point mutant in SlgN1 confirmed the importance of the serine residue at the binding interface for the adenylating activity of SlgN1 (Herbst et al., 2013).

The structures of the full length NRPS EntF bound to the cognate MLP YbdZ and the non-cognate MLP PA2412 have been reported (Miller et al., 2016). In these structures, both MLPs bind to the EntF A-domain at a site distal from the catalytic center, resembling the interface found in SlgN1. Additionally, the conformation of the EntF A-domain remains the same regardless of whether the MLP is bound to it or not (Miller et al., 2016). These findings obscure the role of MLPs in the adenylating activity of A-domains since it has been shown that the co-expression of EntF with YbdZ and the addition of MLPs (YbdZ or PA2412) to the individually purified EntF increases the affinity of the NRPS for its amino

acid substrate (Felnagle et al., 2010; Schomer and Thomas, 2017; Schomer et al., 2018). Even though the importance of MLPs during NRP biosynthesis has been previously established, the mechanism of MLPs as activators in the adenylation reaction is still poorly understood, and the role of MLPs in the entire NRPS assembly line remains ambiguous.

1.7 Objectives and Goals of Thesis Research

S. scabiei is an important causative agent of potato CS disease, and the key pathogenicity factor produced by this organism is the phytotoxic NRP thaxtomin A. The biosynthesis of thaxtomin A involves the megasynthetases TxtA and TxtB, both of which contain an A-domain that selects and activates the amino acid substrate that gets incorporated into the thaxtomin backbone. In addition, a gene (*txtH*) encoding a small protein belonging to the MLP family has been identified in the *txt* gene cluster, but its role in thaxtomin biosynthesis has not been elucidated. The biosynthesis of some NRPs requires the presence of an MLP for efficient production, where the MLP functions as a chaperone for ensuring the proper folding of the A-domain(s), and/or it stimulates or enhances the enzymatic activity of the A-domain(s). Remarkably, MLPs from different NRP pathways can functionally replace each other in promoting the production of NRPs, though the mechanism by which this occurs is still poorly understood. Whether MLPs from other NRP biosynthetic pathways have the ability to exhibit functional cross-talk with TxtH is currently unknown.

This thesis is organized into three research chapters, each of which examines different aspects of MLP function in the thaxtomin biosynthetic pathway in *S. scabiei*. In

Chapter 2, the role of TxtH in thaxtomin biosynthesis was investigated by constructing a *txtH* deletion mutant in *S. scabiei*. Two other endogenous MLP-encoding genes were also deleted along with *txtH* to see if either MLP can compensate for the loss of TxtH. The resulting mutant strains were analyzed for thaxtomin production, and the virulence phenotype of each was investigated using a potato tuber assay. Moreover, two MLPs from other *Streptomyces* species were overexpressed in a *S. scabiei* MLP triple mutant to assess their ability to exhibit functional cross-talk with TxtH. To gain insights into the function of TxtH, TxtA and TxtB A-domains were expressed as N-terminal 6×histidine (HIS_6)-tagged proteins in *E. coli* together or without HIS_6-tagged TxtH to determine whether TxtH is required for the soluble production for each A-domain. Finally, residues that contribute to the function of TxtH were identified by conducting site-directed mutagenesis of the TxtH protein.

In Chapter 3, the mechanism MLP functional redundancy was further investigated by examining the ability of various MLPs from different biosynthetic pathways to compensate for the loss of TxtH during the biosynthesis of thaxtomin A. An MLP phylogenetic tree was constructed using more than one hundred amino acid sequences of MLPs from the database, and several MLP candidates from diverse phylogenetic clades were selected for functional studies. Each MLP was expressed as a HIS_6-tagged protein together with the TxtA and TxtB A-domains in *E. coli* to assess whether any could promote the soluble expression of each A-domain. In addition, the MLPs were expressed in the *S. scabiei* MLP triple mutant to determine whether they can promote thaxtomin A biosynthesis in the absence of TxtH. Finally, *in silico* structural analysis of the protein

complex involving TxtH and the Txt A-domains was conducted in order to identify residues that may play a key role in the MLP-NRPS interaction.

In Chapter 4, the aim of the study was to investigate whether TxtH and other MLPs can influence the enzymatic activity of the Txt NRPS A-domains. *In vivo* chemical cross-linking was conducted in *E. coli* expressing HIS_6-tagged TxtH and TxtA A-domain to confirm that the proteins interact and to determine the stoichiometry of the protein complex. The HIS_6-tagged A-domain was co-expressed with both HIS_6-tagged and untagged txtH, and the A-domain was purified and tested for its ability to adenylate L-phenylalanine *in vitro* using a colorimetric assay. In addition, the AMT-domains of both TxtA and TxtB were co-expressed with TxtH as HIS_6-tagged proteins in *E. coli,* and an attempt to purify the TxtA AMT-domains was made in order to use the protein in *in vitro* enzyme assays.

1.8 References

Adams, M.J., Read, P.J., Lapwood, D.H., Cayley, G.R., and Hide, G.A. (1987) The effect of irrigation on powdery scab and other tuber diseases of potatoes. *Ann. Appl. Biol.* 110, 287–294. doi:10.1111/j.1744-7348.1987.tb03258.x.

Alkhalaf, L.M., Barry, S.M., Rea, D., Gallo, A., Griffiths, D., Lewandowski, J.R., et al. (2018) Binding of distinct substrate conformations enables hydroxylation of remote sites in thaxtomin D by cytochrome P450 TxtC. *J. Am. Chem. Soc.* 141, 216-222. doi: 10.1021/jacs.8b08864.

Al-Mestarihi, A.H., Villamizar, G., Fernández, J., Zolova, O.E., Lombó, F., and Garneau-Tsodikova, S. (2014) Adenylation and S-methylation of cysteine by the bifunctional

enzyme TioN in thiocoraline biosynthesis. *J. Am. Chem. Soc*. 136, 17350-17354. doi: 10.1021/ja510489j.

Arseneault, T., Goyer, C., and Filion, M. (2013) Phenazine production by Pseudomonas sp. LBUM223 contributes to the biological control of potato common scab. *Phytopathology*. 103, 995-1000. doi:10.1094/PHYTO-01-13-0022-R.

Arseneault, T., Goyer, C., and Filion, M. (2015) *Pseudomonas fluorescens* LBUM223 increases potato yield and reduces common scab symptoms in the field. *Phytopathology*. 105, 1311-1317. doi:10.1094/PHYTO-12-14-0358-R.

Arseneault, T., Roquigny, R., Novinscak, A., Goyer, C., and Filion, M. (2020) Phenazine-1-carboxylic acid-producing *Pseudomonas synxantha* LBUM223 alters the transcriptome of *Streptomyces scabies*, the causal agent of potato common scab. *Physiol. Mol. Plant Pathol*. 110, 101480. doi:10.1016/j.pmpp.2020.101480.

Baltz, R.H. (2011) Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J. Ind. Microbiol. Biotechnol*. 38, 1747. doi:10.1007/s10295-011-1022-8.

Barry, S.M., Kers, J.A., Johnson, E.G., Song, L.J., Aston, P.R., Patel, B., Krasnoff, S.B., Crane, B.R., Gibson, D.M., Loria, R. and Challis, G.L. (2012) Cytochrome P450-catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. *Nat Chem Biol*. 8, 814-816. doi:10.1038/nchembio.1048.

Bérdy, J. (2005) Bioactive Microbial Metabolites. *J Antibiot*. 58, 1-26. doi:10.1038/ja.2005.1.

Bernards, M.A., Lerat, S., Simao-Beaunoir, A.M., Padilla-Reynaud, R., and Beaulieu, C. (2015) Suberin Regulates the Production of Cellulolytic Enzymes in *Streptomyces*

scabiei, the Causal Agent of Potato Common Scab. *Microbes Environ.* 30, 245–253. doi:10.1264/jsme2.me15034.

Bignell, D.R.D., Francis, I.M., Fyans, J.K., and Loria, R. (2014a) Thaxtomin A Production and Virulence Are Controlled by Several *bld* Gene Global Regulators in *Streptomyces scabies*. *Mol. Plant Microbe Interact.* 27, 875–885. doi:10.1094/mpmi-02-14-0037-r.

Bignell, D.R.D., Fyans, J.K., and Cheng, Z. (2014b) Phytotoxins produced by plant pathogenic *Streptomyces* species. *J Appl Microbiol*. 116, 223-235. doi:10.1111/jam.12369.

Bignell, D.R., Huguet-Tapia, J.C., Joshi, M.V., Pettis, G.S., and Loria, R. (2010) What does it take to be a plant pathogen: genomic insights from *Streptomyces* species. *Antonie Van Leeuwenhoek*. 98, 179-194. doi: 10.1007/s10482-010-9429-1.

Bischoff, V., Cookson, S.J., Wu, S., and Scheible, W.R. (2009) Thaxtomin A affects CESA-complex density, expression of cell wall genes, cell wall composition, and causes ectopic lignification in *Arabidopsis thaliana* seedlings. *J Exp Bot*. 60, 955–965. doi:10.1093/jxb/ern344.

Boll, B., Taubitz, T., and Heide, L. (2011) Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. *J. Biol. Chem*. 286, 36281-36290. doi:10.1074/jbc.M111.288092.

Buchko, G.W., Kim, C.Y., Terwilliger, T.C., and Myler, P.J. (2010) Solution structure of Rv2377c-founding member of the MbtH-like protein family. *Tuberculosis*. 90, 245-251. doi:10.1016/j.tube.2010.04.002.

Challis, G.L., Ravel, J., and Townsend, C.A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* 7, 211–224. doi:10.1016/S1074-5521(00)00091-0.

Chater, K.F. (2006) *Streptomyces* inside-out: A new perspective on the bacteria that provide us with antibiotics. *Philos Trans R Soc Lond B Biol Sci.* 61, 761-768. doi:10.1098/rstb.2005.1758.

Chen, C.W., Huang, C.H., Lee, H.H., Tsai, H.H., and Kirby, R. (2002) Once the circle has been broken: Dynamics and evolution of *Streptomyces* chromosomes. *Trends Genet.* 18, 522-529. doi:10.1016/S0168-9525(02)02752-X.

Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO Journal*. doi:10.1093/emboj/16.14.4174.

Davidsen, J.M., Bartley, D.M., and Townsend, C.A. (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. doi: 10.1021/ja307710d.

Dees, M.W., Somervuo, P., Lysøe, E., Aittamaa, M., and Valkonen, J.P.T. (2012) Species' identification and microarray-based comparative genome analysis of *Streptomyces* species isolated from potato scab lesions in Norway. *Mol. Plant Pathol.* 13, 174-186. doi:10.1111/j.1364-3703.2011.00741.x.

Dees, M.W., and Wanner, L.A. (2012) In Search of Better Management of Potato Common Scab. *Potato Res*. 55, 249-268. doi:10.1007/s11540-012-9206-9.

Dharmaraj, S. (2010) Marine *Streptomyces* as a novel source of bioactive substances. *World J Microbiol Biotechnol*. 26, 2123-2139. doi: 10.1007/s11274-010-0415-6.

Dharmaraj, S., and Sumantha, A. (2009) Bioactive potential of *Streptomyces* associated with marine sponges. *World J. Microbiol. Biotechnol.* 25, 1971-1979. doi:10.1007/s11274-009-0096-1.

Drake, E.J., Cao, J., Qu, J., Shah, M.B., Straubinger, R.M., and Gulick, A.M. (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. doi:10.1074/jbc.M611833200.

Du, L., Sánchez, C., Chen, M., Edwards, D.J., and Shen, B. (2000) The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem. Biol.* 7, 623-642. doi:10.1016/S1074-5521(00)00011-9.

Duval, I., and Beaudoin, N. (2009) Transcriptional profiling in response to inhibition of cellulose synthesis by thaxtomin A and isoxaben in *1* suspension cells. *Plant Cell Rep.* 28, 811–830. doi:10.1007/s00299-009-0670-x.

Duval, I., Brochu, V., Simard, M., Beaulieu, C., and Beaudoin, N. (2005) Thaxtomin A induces programmed cell death in *Arabidopsis thaliana* suspension-cultured cells. *Planta.* 222, 820–831. doi:10.1007/s00425-005-0016-z.

Edwards, D.J., and Gerwick, W.H. (2004) Lyngbyatoxin biosynthesis: Sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. *J. Am. Chem. Soc*. 126, 11432-11433. doi:10.1021/ja047876g.

Egan, S.M. (2002) Growing Repertoire of AraC /XylS Activators. *J. Bacteriol.* 184, 5529–5532. doi:10.1128/JB.184.20.5529.

Elliot, M.A., Buttner, M.J., and Nodwell, J.R. (2007) Multicellular Development in *Streptomyces*. *Myxobacteria*. 419–438. doi:10.1128/9781555815677.ch24.

Eppelmann, K., Stachelhaus, T., and Marahiel, M.A. (2002) Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry*. 41, 9718-9726. doi:10.1021/bi0259406.

Errakhi, R., Dauphin, A., Meimoun, P., Lehner, A., Reboutier, D., Vatsa, P., et al. (2008) An early Ca^{2+} influx is a prerequisite to thaxtomin A-induced cell death in *Arabidopsis thaliana* cells. *J. Exp. Bot*. 59, 4259-4270. doi:10.1093/jxb/ern267.

Felnagle, E.A., Barkei, J.J., Park, H., Podevels, A.M., McMahon, M.D., Drott, D.W., et al. (2010) MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry.* 49, 8815–8817. doi:10.1021/bi1012854.

Fernández-Martínez, L.T., and Hoskisson, P.A. (2019) Expanding, integrating, sensing and responding: the role of primary metabolism in specialised metabolite production. *Curr. Opin. Microbiol*. 51, 16-21. doi:10.1016/j.mib.2019.03.006.

Filippova, S.N., and Vinogradova, K.A. (2017) Programmed cell death as one of the stages of *Streptomycete* differentiation. *Microbiology (Russian Federation)*. 86, 439-454. doi:10.1134/S0026261717040075.

Finking, R., and Marahiel, M.A. (2004) Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol*. 58, 453–488. doi:10.1146/annurev.micro.58.030603.123615.

Flärdh, K., and Buttner, M.J. (2009) *Streptomyces* morphogenetics: Dissecting differentiation in a filamentous bacterium. *Nat. Rev. Microbiol*. 7, 36-49. doi:10.1038/nrmicro1968.

Flärdh, K., Richards, D.M., Hempel, A.M., Howard, M., and Buttner, M.J. (2012) Regulation of apical growth and hyphal branching in *Streptomyces*. *Curr. Opin. Microbiol.* 15, 737-743. doi:10.1016/j.mib.2012.10.012.

Francis, I.M., Jourdan, S., Fanara, S., Loria, R., and Rigali, S. (2015) The cellobiose sensor CebR is the gatekeeper of *Streptomyces scabies* pathogenicity. *mBio*. 6, e02018-14. doi:10.1128/mBio.02018-14.

Fry, B.A., and Loria, R. (2002) Thaxtomin A: Evidence for a plant cell wall target. *Physiol. Mol. Plant Pathol*. 60, 1-8. doi:10.1006/pmpp.2001.0371.

Gallegos, M.T., Schleif, R., Bairoch, A., Hofmann, K., and Ramos, J.L. (1997) Arac/XylS family of transcriptional regulators. *Microbiol. Mol. Biol. Rev.* 61, 393–410. doi:9409145.

Gehring, A.M., Mori, I., and Walsh, C.T. (1998) Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry*. 37, 2648-2659. doi:10.1021/bi9726584.

Goyer, C., and Beaulieu, C. (1997) Host range of *Streptomycete* strains causing common scab. *Plant Disease*. 81, 901-904. doi:10.1094/PDIS.1997.81.8.901.

Goyer, C., Vachon, J., and Beaulieu, C. (1998) Pathogenicity of *Streptomyces scabies* mutants altered in thaxtomin a production. *Phytopathology*. 88, 442-445. doi:10.1094/PHYTO.1998.88.5.442.

Guan, D., Grau, B.L., Clark, C.A., Taylor, C.M., Loria, R., and Pettis, G.S. (2012) Evidence that thaxtomin C is a pathogenicity determinant of *Streptomyces ipomoeae*, the causative agent of *Streptomyces* soil rot disease of sweet potato. *Mol. Plant Microbe Interact.* 25, 393–401. doi:10.1094/mpmi-03-11-0073.

Grondin, G., Lerat, S., Forest, M., Beaulieu, C., Lauzier, A., and Lacelle, S. (2011) Potato Suberin Induces Differentiation and Secondary Metabolism in the Genus *Streptomyces*. *Microbes Environ.* 27, 36–42. doi:10.1264/jsme2.me11282.

Gulick, A.M. (2009) Conformational dynamics in the acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem. Biol.* 4, 811-827. doi:10.1021/cb900156h.

Hackl, S., and Bechthold, A. (2015) The Gene *bldA*, a Regulator of Morphological Differentiation and Antibiotic Production in *Streptomyces*. *Archiv der Pharmazie*. 348, 455-462. doi:10.1002/ardp.201500073.

Hamedi J., Poorinmohammad N. and Papiran R. (2017) Growth and life cycle of Actinobacteria. *Biology and biotechnology of Actinobacteria*. 29-50. Springer: DOI 10.1007/978-3-319-60339-1.

Harir, M., Bendif, H., Bellahcene, M., Fortas, Z. and Pogni, R. (2018) *Streptomyces* Secondary Metabolites. *Basic Biology and Applications of Actinobacteria.* 6, 99-122. doi:10.5772/intechopen.79890.

Haynes, C., Oldfield, C. J., Ji, F., Klitgord, N., Cusick, M.E., Radivojac, P., et al. (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput. Biol*. 2, e100. doi:10.1371/journal.pcbi.0020100.

Healy, F.G., Krasnoff, S.B., Wach, M., Gibson, D.M., and Loria, R. (2002) Involvement of a cytochrome P450 monooxygenase in thaxtomin A biosynthesis by *Streptomyces acidiscabies*. *J. Bacteriol*. 184, 2019-2029. doi:10.1128/JB.184.7.2019-2029.2002.

Healy, F.G., Wach, M., Krasnoff, S.B., Gibson, D.M., and Loria, R. (2000) The *txtAB* genes of the plant pathogen *Streptomyces acidiscabies* encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity. *Mol. Microbiol*. 38, 794-804. doi:10.1046/j.1365-2958.2000.02170.x.

Heemstra, J.R., Walsh, C.T., and Sattely, E.S. (2009) Enzymatic tailoring of ornithine in the biosynthesis of the rhizobium cyclic trihydroxamate siderophore vicibactin. *J. Am. Chem. Soc.* 131, 15317–15329. doi:10.1021/ja9056008.

Herbst, D.A., Boll, B., Zocher, G., Stehle, T., and 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. doi:10.1074/jbc.M112.420182.

Hill, J., and Lazarovits, G. (2005) A mail survey of growers to estimate potato common scab prevalence and economic loss in Canada. *Can. J. Plant Pathol*. 27, 46-52. doi:10.1080/07060660509507192.

Hiltunen, L.H., Weckman, A., Ylhäinen, A., Rita, H., Richter, E., and Valkonen, J.P.T. (2005) Responses of potato cultivars to the common scab pathogens, *Streptomyces scabies* and *S. turgidiscabies*. *Ann. Appl. Biol*. 146, 395-403. doi:10.1111/j.1744-7348.2005.040083.x.

Hopwood, D.A. (2006) Soil to genomics: The *Streptomyces* chromosome. *Annu. Rev. Genet*. 40, 1-23. doi:10.1146/annurev.genet.40.110405.090639.

Huguet-Tapia, J.C., Lefebure, T., Badger, J.H., Guan, D., Pettis, G.S., Stanhope, M.J., et al. (2016) Genome content and phylogenomics reveal both ancestral and lateral evolutionary pathways in plant-pathogenic *Streptomyces* species. *Appl. Environ. Microbiol.* 82, 2146-2155. doi:10.1128/AEM.03504-15.

Huynh, M.U., Elston, M.C., Hernandez, N.M., Ball, D.B., Kajiyama, S.I., Irie, K., et al. (2010) Enzymatic production of (-)-indolactam V by LtxB, a cytochrome P450 monooxygenase. *J. Nat. Prod.* 73, 71–74. doi:10.1021/np900481a.

Hwang, K.S., Kim, H.U., Charusanti, P., Palsson, B.T., and Lee, S.Y. (2014) Systems biology and biotechnology of *Streptomyces* species for the production of secondary metabolites. *Biotechnol. Adv*. 32, 255-268. doi:10.1016/j.biotechadv.2013.10.008.

Hwang, S., Lee, N., Jeong, Y., Lee, Y., Kim, W., Cho, S., et al. (2019) Primary transcriptome and translatome analysis determines transcriptional and translational regulatory elements encoded in the *Streptomyces clavuligerus* genome. *Nucleic Acids Res.* 47, 6114-6129. doi:10.1093/nar/gkz471.

Ibarra, J.A., Pérez-Rueda, E., Segovia, L., and Puente, J.L. (2008) The DNA-binding domain as a functional indicator: The case of the AraC/XylS family of transcription factors. *Genetica.* 133, 65–76. doi:10.1007/s10709-007-9185-y.

Imker, H.J., Krahn, D., Clerc, J., Kaiser, M., and Walsh, C.T. (2010) *N*-Acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF. *Chem. Biol.* 17, 1077–1083. doi:10.1016/j.chembiol.2010.08.007.

Jiang, G., Zuo, R., Zhang, Y., Powell, M.M., Zhang, P., Hylton, S.M., et al. (2018) One-Pot Biocombinatorial Synthesis of Herbicidal Thaxtomins. *ACS Catal.* 8, 10761–10768. doi:10.1021/acscatal.8b03317.

Johnson, E.G., Joshi, M.V., Gibson, D.M., and Loria, R. (2007) Cello-oligosaccharides released from host plants induce pathogenicity in scab-causing *Streptomyces* species. *Physiol. Mol. Plant Pathol*. 71, 18-25. doi:10.1016/j.pmpp.2007.09.003.

Johnson, E.G., Krasnoff, S.B., Bignell, D.R.D., Chung, W.C., Tao, T., Parry, R. J., et al. (2009) 4-Nitrotryptophan is a substrate for the non-ribosomal peptide synthetase TxtB in the thaxtomin A biosynthetic pathway. *Mol. Microbiol*. 73, 409-418. doi:10.1111/j.1365-2958.2009.06780.x.

Jones, S.E., and Elliot, M.A. (2017) *Streptomyces* Exploration: Competition, Volatile Communication and New Bacterial Behaviours. *Trends Microbiol*. 25, 522-531. doi:10.1016/j.tim.2017.02.001.

Jones, S.E., Ho, L., Rees, C.A., Hill, J.E., Nodwell, J.R., and Elliot, M.A. (2017) *Streptomyces* exploration is triggered by fungal interactions and volatile signals. *eLife*. 6, e21738. doi:10.7554/eLife.21738.

Jones, S.E., Pham, C.A., Zambri, M.P., McKillip, J., Carlson, E.E., and Elliot, M.A. (2019) *Streptomyces* volatile compounds influence exploration and microbial community dynamics by altering iron availability. *mBio*. 10, e00171-19. doi:10.1128/MBIO.00171-19.

Joshi, M.V., Bignell, D.R.D., Johnson, E.G., Sparks, J.P., Gibson, D.M., and Loria, R. (2007) The AraC/XylS regulator TxtR modulates thaxtomin biosynthesis and

virulence in *Streptomyces scabies*. *Mol Microbiol.* 66, 633–642. doi:10.1111/j.1365-2958.2007.05942.x.

Jourdan, S., Francis, I.M., Kim, M.J., Salazar, J.J.C., Planckaert, S., Frère, J.M., et al. (2016) The CebE/MsiK Transporter is a Doorway to the Cello-oligosaccharide-mediated Induction of *Streptomyces scabies* Pathogenicity. *Sci. Rep.* 6, 1–12. doi:10.1038/srep27144.

Kaltenpoth, M., Goettler, W., Dale, C., Stubblefield, J. W., Herzner, G., Roeser-Mueller, K., et al. (2006) Candidatus *Streptomyces philanthi*, an endosymbiotic *Streptomycete* in the antennae of *Philanthus* digger wasps. *Int J Syst Evol Microbiol*. 56, 1403-1411. doi:10.1099/ijs.0.64117-0.

Kaltenpoth, M., Schmitt, T., Polidori, C., Koedam, D., and Strohm, E. (2010) Symbiotic *Streptomycetes* in antennal glands of the South American digger wasp genus *Trachypus* (Hymenoptera, Crabronidae). *Physiol. Entomol*. 35, 196-200. doi:10.1111/j.1365-3032.2010.00729.x.

Kaniusaite, M., Tailhades, J., Kittilä, T., Fage, C.D., Goode, R.J.A., Schittenhelm, R. B., et al. (2020) Understanding the early stages of peptide formation during the biosynthesis of teicoplanin and related glycopeptide antibiotics. *FEBS J.* doi:10.1111/febs.15350. [Epub ahead of print].

Kers, J.A., Wach, M.J., Krasnoff, S.U., Widom, J., Cameron, K.D., Dukhalid, R.A., et al. (2004) Nitration of a peptide phytotoxin by bacterial nitric oxide synthase. *Nature.* 429, 79–82. doi:10.1038/nature02504.

King, R.R., and Calhoun, L.A. (2009) The thaxtomin phytotoxins: Sources, synthesis, biosynthesis, biotransformation and biological activity. *Phytochemistry*. 70, 833-841. doi:10.1016/j.phytochem.2009.04.013.

King, R.R., Lawrence, C.H., Calhoun, L.A., and Ristaino, J.B. (1994) Isolation and Characterization of Thaxtomin-Type Phytotoxins Associated with *Streptomyces ipomoeae*. *J Agric Food Chem*. 42, 1791-1794. doi:10.1021/jf00044a042.

King, R.R., Lawrence, C.H., and Clark, M.C. (1991) Correlation of phytotoxin production with pathogenicity of *Streptomyces scabies* isolates from scab infected potato tubers. *Am Potato J*. 68, 675-680. doi:10.1007/BF02853743.

King, R.R., Lawrence, C.H., Clark, M.C., and Calhoun, L.A. (1989) Isolation and characterization of phytotoxins associated with *Streptomyces scabies*. *J Chem Soc Chem Commun.* 13, 849-850. doi:10.1039/C39890000849.

Kinkel, L.L., Bowers, J.H., Shimizu, K., Neeno-Eckwall, E.C., and Schottel, J.L. (1998) Quantitative relationships among thaxtomin A production, potato scab severity, and fatty acid composition in *Streptomyces*. *Can J Microbiol*. 44, 768-776. doi:10.1139/w98-061.

Komeil, D., Simao-Beaunoir, A.M., and Beaulieu, C. (2013) Detection of potential suberinase-encoding genes in *Streptomyces scabiei* strains and other actinobacteria. *Can J Microbiol*. 59, 294-303. doi:10.1139/cjm-2012-0741.

Krastel, P., Zeeck, A., Gebhardt, K., Fiedler, H.P., and Rheinheimer, J. (2002) Endophenazines AD, new phenazine antibiotics from the athropod. Associated endosymbiont *Streptomyces anulatus*. *J Antibiot*. 55, 794-800. doi:10.7164/antibiotics.55.801.

Kreitler, D.F., Gemmell, E.M., Schaffer, J.E., Wencewicz, T.A., and Gulick, A.M. (2019) The structural basis of *N*-acyl-α-amino-β-lactone formation catalyzed by a nonribosomal peptide synthetase. *Nat Commun.* 10, 1–13. doi:10.1038/s41467-019-11383-7.

Krištůfek, V., Nováková, A., and Pižl, V. (2001) Coprophilous *Streptomycetes* and Fungi - Food Sources for Enchytraeid Worms (*Enchytraeidae*). *Folia Microbiol*. 46, 555-558. doi:10.1007/BF02818002.

Kroiss, J., Kaltenpoth, M., Schneider, B., Schwinger, M.G., Hertweck, C., Maddula, R.K., et al. (2010) Symbiotic *Streptomycetes* provide antibiotic combination prophylaxis for wasp offspring. *Nat Chem Biol*. 6, 261-263. doi:10.1038/nchembio.331.

Labby, K.J., Watsula, S.G., and Garneau-Tsodikova, S. (2015) Interrupted adenylation domains: Unique bifunctional enzymes involved in nonribosomal peptide biosynthesis. *Nat Prod Rep*. 32, 641-653. doi:10.1039/c4np00120f.

Lambert, D.H., and Loria, R. (1989) *Streptomyces scabies* sp. nov., nom. rev. *Int J Syst Evol Microbiol*. 39, 387-392. doi:10.1099/00207713-39-4-387.

Lapwood, D.H., and Hering, T.F. (1970) Soil moisture and the infection of young potato tubers by *Streptomyces scabies* (common scab). *Potato Res*. 13, 296-304. doi:10.1007/BF02358275.

Lautru, S., Oves-Costales, D., Pernodet, J.L., and Challis, G.L. (2007) MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. *Microbiology*. 153, 1405-1412. doi:10.1099/mic.0.2006/003145-0.

Lawrence, C.H. (1990) Induction of Common Scab Symptoms in Aseptically Cultured Potato Tubers by the Vivotoxin, Thaxtomin. *Phytopathology*. 80, 606-608. doi:10.1094/phyto-80-606.

Lee, K.S., Lee, B.M., Ryu, J.H., Kim, D.H., Kim, Y.H., and Lim, S.K. (2016) Increased vancomycin production by overexpression of MbtH-like protein in *Amycolatopsis orientalis* KFCC10990P. *Lett Appl Microbiol.* 63, 222-228. doi:10.1111/lam.12617.

Lee, N., Hwang, S., Lee, Y., Cho, S., Palsson, B., and Cho, B.K. (2019) Synthetic biology tools for novel secondary metabolite discovery in *Streptomyces*. *J Microbiol Biotechnol.* 29, 667-686. doi:10.4014/jmb.1904.04015.

Leiner, R.H., Fry, B.A., Carling, D.E., and Loria, R. (1996) Probable involvement of thaxtomin A in pathogenicity of *Streptomyces scabies* on seedlings. *Phytopathology*. doi:10.1094/Phyto-86-709.

Lerat, S., Forest, M., Lauzier, A., Grondin, G., Lacelle, S., and Beaulieu, C. (2012) Potato suberin induces differentiation and secondary metabolism in the genus *Streptomyces*. *Microbes Environ.* 27, 36–42. doi:10.1264/jsme2.ME11282

Lerat, S., Simao-Beaunoir, A.M., and Beaulieu, C. (2009) Genetic and physiological determinants of *Streptomyces scabies* pathogenicity. *Mol Plant Pathol.* 10, 579–585. doi:10.1111/j.1364-3703.2009.00561.x.

Lerat, S., Simao-Beaunoir, A.M., Wu, R., Beaudoin, N., and Beaulieu, C. (2010) Involvement of the plant polymer suberin and the disaccharide cellobiose in triggering thaxtomin A biosynthesis, a phytotoxin produced by the pathogenic agent *Streptomyces scabies*. *Phytopathology*. 100, 91-96. doi: 10.1094/PHYTO-100-1-0091.

Li, Y., Liu, J., Díaz-Cruz, G., Cheng, Z., and Bignell, D.R.D. (2019) Virulence mechanisms of plant-pathogenic *Streptomyces* species: An updated review. *Microbiology.* 165, 1025-1040. doi:10.1099/mic.0.000818.

Lin, Z., Antemano, R.R., Hughen, R.W., Tianero, M.D.B., Peraud, O., Haygood, M.G., et al. (2010) Pulicatins A-E, neuroactive thiazoline metabolites from cone snail-associated bacteria. *J Nat Prod*. 73, 1922-1926. doi:10.1021/np100588c.

Lindholm, P., Kortemaa, H., Kokkola, M., Haahtela, K., Salkinoja-Salonen, M., and Valkonen, J. P.T. (1997) *Streptomyces* spp. isolated from potato scab lesions under nordic conditions in Finland. *Plant Dis*. 81, 1317-1322. doi:10.1094/PDIS.1997.81.11.1317.

Loria, R., Bukhalid, R.A., Creath, R.A., Leiner, R.H., Olivier, M., and Steffens, J.C. (1995) Differential production of thaxtomins by pathogenic *Streptomyces* species *in vitro*. *Phytopathology*. 85, 537-541. doi:10.1094/Phyto-85-537.

Loria, R., Bukhalid, R.A., Fry, B.A., and King, R.R. (1997) Plant pathogenicity in the genus *Streptomyces*. *Plant Dis*. 81, 836-846. doi:10.1094/PDIS.1997.81.8.836.

Loria, R., Coombs, J., Yoshida, M., Kers, J., and Bukhalid, R. (2003) A paucity of bacterial root diseases: *Streptomyces* succeeds where others fail. *Physiol Mol Plant Pathol.* 62, 65-72. doi:10.1016/S0885-5765(03)00041-9.

Loria, R., Kers, J., and Joshi, M. (2006) Evolution of Plant Pathogenicity in *Streptomyces*. *Annu Rev Phytopathol.* 44, 469–487. doi: 10.1146/annurev.phyto.44.032905.091147.

Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem Rev*. 97, 2651-2674. doi:10.1021/cr960029e.

May, J.J., Kessler, N., Marahiel, M.A., and Stubbs, M.T. (2002) Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci USA.* 99, 12120-12125. doi:10.1073/pnas.182156699.

McCormick, J.R., and Flärdh, K. (2012) Signals and regulators that govern *Streptomyces* development. *FEMS Microbiol Rev*. 36, 206-231. doi:10.1111/j.1574-6976.2011.00317.x.

McMahon, M.D., Rush, J.S., and Thomas, M.G. (2012) Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *J Bacteriol.* 194, 2809–2818. doi:10.1128/JB.00088-12.

Miguélez, E.M. (2000) *Streptomycetes*: A new model to study cell death. *Int Microbiol.* 3, 153-158. doi:10.2436/im.v3i3.9269.

Miller, B.R., Drake, E.J., Shi, C., Aldrich, C.C., and Gulick, A.M. (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. doi:10.1074/jbc.M116.746297.

Miller, B.R., Sundlov, J.A., Drake, E.J., Makin, T.A., and Gulick, A.M. (2014) Analysis of the linker region joining the adenylation and carrier protein domains of the modular nonribosomal peptide synthetases. *Proteins*. 82, 2691-2702. doi:10.1002/prot.24635.

Mori, S., Green, K.D., Choi, R., Buchko, G.W., Fried, M.G., and Garneau-Tsodikova, S. (2018a) Using MbtH-Like Proteins to Alter the Substrate Profile of a Nonribosomal Peptide Adenylation Enzyme. *ChemBioChem.* 19, 2186–2194. doi:10.1002/cbic.201800240.

Mori, S., Pang, A.H., Lundy, T.A., Garzan, A., Tsodikov, O.V., and Garneau-Tsodikova, S. (2018b) Structural basis for backbone *N*-methylation by an interrupted adenylation domain. *Nat Chem Biol.* 14, 428–430. doi:10.1038/s41589-018-0014-7.

O'Brien, J., and Wright, G.D. (2011) An ecological perspective of microbial secondary metabolism. *Curr Opin Biotechnol*. 22, 552-558. doi: 10.1016/j.copbio.2011.03.010.

Olanrewaju, O.S., and Babalola, O.O. (2019) *Streptomyces*: implications and interactions in plant growth promotion. *Appl Microbiol Biotechnol*. 103, 1179-1188. doi: 10.1007/s00253-018-09577-y.

O'Neill, L. (2019) Interspecies interactions and 'exploratory' behavior of plant pathogenic *Streptomyces* species. B.Sc. Honours Thesis. Memorial University of Newfoundland.

Park, S.B., Lee, I.A., Suh, J.W., Kim, J.G., and Lee, C.H. (2011) Screening and identification of antimicrobial compounds from *Streptomyces bottropensis* suppressing rice bacterial blight. *J Microbiol Biotechnol.* 21, 1236-1242. doi: 10.4014/jmb.1106.06047

Pathom-aree, W., Stach, J.E.M., Ward, A.C., Horikoshi, K., Bull, A.T., and Goodfellow, M. (2006) Diversity of actinomycetes isolated from Challenger Deep sediment

(10,898 m) from the Mariana Trench. *Extremophiles*. 10, 181-189. doi:10.1007/s00792-005-0482-z.

Quadri, L.E.N., Sello, J., Keating, T.A., Weinreb, P.H., and Walsh, C.T. (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. doi:10.1016/S1074-5521(98)90291-5.

Riedlinger, J., Schrey, S.D., Tarkka, M.T., Hampp, R., Kapur, M., and Fiedler, H.P. (2006) Auxofuran, a novel metabolite that stimulates the growth of fly agaric, is produced by the mycorrhiza helper bacterium *Streptomyces* strain AcH 505. *Appl Environ Microbiol*. 72, 3550-3557. doi:10.1128/AEM.72.5.3550-3557.2006.

Sadeghi, A., Karimi, E., Dahaji, P.A., Javid, M.G., Dalvand, Y., and Askari, H. (2012) Plant growth promoting activity of an auxin and siderophore producing isolate of *Streptomyces* under saline soil conditions. *World J. Microbiol. Biotechnol*. 28, 1503-1509. doi: 10.1007/s11274-011-0952-7.

Sarwar, A., Latif, Z., Zhang, S., Hao, J., and Bechthold, A. (2019) A potential biocontrol agent *Streptomyces violaceusniger* AC12AB for managing potato common scab. *Front Microbiol.* 10, 202. doi:10.3389/fmicb.2019.00202.

Scheible W.R., Fry B., Kochevenko A., Schindelasch D., Zimmerli L., Somerville S., et al. (2003) An *Arabidopsis* Mutant Resistant to Thaxtomin A, a Cellulose Synthesis Inhibitor from *Streptomyces* Species. *Plant Cell.* 15, 1781–1794. doi:10.1105/tpc.013342.

Schomer, R.A., Park, H., Barkei, J.J., and Thomas, M.G. (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. doi:10.1021/acs.biochem.8b00552.

Schomer, R.A., and Thomas, M.G. (2017) Characterization of the Functional Variance in MbtH-like Protein Interactions with a Nonribosomal Peptide Synthetase. *Biochemistry.* 56, 5380–5390. doi:10.1021/acs.biochem.7b00517.

Seipke, R.F., Kaltenpoth, M., and Hutchings, M.I. (2012) *Streptomyces* as symbionts: An emerging and widespread theme? *FEMS Microbiol Rev.* 36, 862–876. doi:10.1111/j.1574-6976.2011.00313.x.

Sieber, S.A., and Marahiel, M.A. (2005) Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics. *Chem Rev.* 105, 715–738. doi:10.1021/cr0301191.

Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol*. 6, 493-505. doi:10.1016/S1074-5521(99)80082-9.

Süssmuth, R.D., and Mainz, A. (2017) Nonribosomal Peptide Synthesis – Principles and Prospects. *Angew Chem.* 56, 3770–3821. doi:10.1002/anie.201609079.

Tarry, M.J., Haque, A.S., Bui, K.H., and Schmeing, T.M. (2017) X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure.* 25, 783-793. doi:10.1016/j.str.2017.03.014.

Tedesco, K.L., and Rybak, M.J. (2003) Daptomycin. *Pharmacotherapy.* 24, 41-57. doi: 10.1592/phco.24.1.41.34802.

Tegg, R.S., Melian, L., Wilson, C.R., and Shabala, S. (2005) Plant cell growth and ion flux responses to the *Streptomycete* phytotoxin thaxtomin A: Calcium and hydrogen flux patterns revealed by the non-invasive MIFE technique. *Plant Cell Physiol.* 46, 638–648. doi:10.1093/pcp/pci069.

Tegg, R.S., Shabala, S., Cuin, T.A., and Wilson, C.R. (2016) Mechanisms of thaxtomin A-induced root toxicity revealed by a thaxtomin A sensitive *Arabidopsis* mutant (ucu2-2/gi-2). *Plant Cell Rep*. 35, 347-356. doi:10.1007/s00299-015-1888-4.

Thirlway, J., Lewis, R., Nunns, L., Al Nakeeb, M., Styles, M., Struck, A.W., et al. (2012) Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity. *Angew Chem.* 51, 7181-7184. doi:10.1002/anie.201202043.

van Keulen, G., and Dyson, P.J. (2014) Production of specialized metabolites by *Streptomyces coelicolor* A3(2). *Adv Appl Microbiol.* 89, 217-266. doi:10.1016/B978-0-12-800259-9.00006-8.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F., et al. (2007) Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev*. 71, 495-548. doi:10.1128/MMBR.00005-07.

Wan, M., Li, G., Zhang, J., Jiang, D., and Huang, H.C. (2008) Effect of volatile substances of *Streptomyces platensis* F-1 on control of plant fungal diseases. *Biol Control*. 46, 552-559. doi:10.1016/j.biocontrol.2008.05.015.

Wang, X.J., Yan, Y.J., Zhang, B., An, J., Wang, J.J., Tian, J., et al. (2010) Genome sequence of the milbemycin-producing bacterium *Streptomyces bingchenggensis*. *J Bacteriol.* 192, 4526-4527. doi:10.1128/JB.00596-10.

Wanner, L.A. (2009) A patchwork of *Streptomyces* species isolated from potato common scab lesions in North America. *Am Potato J Res*. 86, 247-264. doi:10.1007/s12230-009-9078-y.

Wanner, L.A., and Kirk, W.W. (2015) *Streptomyces* – from Basic Microbiology to Role as a Plant Pathogen. *Am Potato J Res*. 92, 236-242. doi:10.1007/s12230-015-9449-5.

Weber, G., Schörgendorfer, K., Schneider-Scherzer, E., and Leitner, E. (1994) The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Curr Genet*. 26, 120-125. doi:10.1007/BF00313798.

Wolpert, M., Gust, B., Kammerer, B., and Heide, L. (2007) Effects of deletions of mbtH-like genes on clorobiocin biosythesis in *Streptomyces coelicolor*. *Microbiology*. 153, 1413–1423. doi:10.1099/mic.0.2006/002998-0.

Xie, H., Vucetic, S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Obradovic, Z., et al. (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J Proteome Res*. 6, 1917-1932. doi:10.1021/pr060394e.

Yagüe, P., Lopez-Garcia, M.T., Rioseras, B., Sanchez, J., and Manteca, A. (2012) New insights on the development of *Streptomyces* and their relationships with secondary metabolite production. *Trends Microbiol.* 8, 65–73. doi: 10651/33649.

Yim, G., Wang, H.H., and Davies, J. (2007) Antibiotics as signalling molecules. *Philos Trans R Soc Lond B Biol Sci.* 362, 1195-1200. doi:10.1098/rstb.2007.2044.

Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V., and Luzhetskyy, A. (2014) Insights into naturally minimised *Streptomyces albus* J1074 genome. *BMC Genomics*. 15, 97. doi:10.1186/1471-2164-15-97.

Zhang, W., Heemstra, J.R., Walsh, C.T., and Imker, H.J. (2010) Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins. *Biochemistry.* 49, 9946–9947. doi:10.1021/bi101539b.

Zhang, Y., Bignell, D.R.D., Zuo, R., Fan, Q., Huguet-Tapia, J.C., Ding, Y., et al. (2016) Promiscuous pathogenicity islands and phylogeny of pathogenic *Streptomyces* spp. *Mol. Plant Microbe Interact*. 29, 640-650. doi:10.1094/MPMI-04-16-0068-R.

Zhou, Z., Gu, J., Li, Y.Q., and Wang, Y. (2012) Genome plasticity and systems evolution in *Streptomyces*. *BMC bioinformatics*. 13, s8. doi:10.1186/1471-2105-13-S10-S8.

Zolova, O.E., and Garneau-Tsodikova, S. (2012) Importance of the MbtH-like protein TioT for production and activation of the thiocoraline adenylation domain of TioK. *MedChemComm.* 3, 950–955. doi:10.1039/c2md20131c.

Zolova, O.E., and Garneau-Tsodikova, S. (2014) KtzJ-dependent serine activation and O-methylation by KtzH for kutznerides biosynthesis. *J Antibiot*. 67, 59-64. doi:10.1038/ja.2013.98.

Zwahlen, R.D., Pohl, C., Bovenberg, R.A.L., and Driessen, A.J.M. (2019) Bacterial MbtH-like Proteins Stimulate Nonribosomal Peptide Synthetase-Derived Secondary Metabolism in Filamentous Fungi. *ACS Synth Biol.* 8, 1776–1787. doi:10.1021/acssynbio.9b00106.

Co-Authorship Statement

Chapter 2 is a version of a manuscript published in Molecular Plant Pathology [Li, Y., Liu, J., Adekunle, D., Bown, L., Tahlan, K., and Bignell, D.R.D. (2019) TxtH is a key component of the thaxtomin biosynthetic machinery in the potato common scab pathogen *Streptomyces scabies*. *Mol Plant Pathol*. 20,1379-93. doi: 10.1111/mpp.12843.]. The initial study concept was designed by D.R.D. Bignell and K. Tahlan, and the experimental methodology was designed by Y. Li, D.R.D. Bignell and K. Tahlan. Y. Li conducted all of the described work except for the following. J. Liu constructed the *S. scabiei* Δ*txtH* mutant, D. Adekunle constructed two overexpression plasmids (pRLDB50-1a/*cdaX* and pRLDB50-1a/*SCLAV_p1293*) and L. Bown conducted the LC-HRESIMS analysis. Y. Li conducted the data analysis and interpretation. The manuscript was drafted and prepared by Y. Li, D.R.D. Bignell and K. Tahlan, with some editorial input provided by the other co-authors.

Chapter 3 is a version of a manuscript published in Frontiers in Microbiology [Li, Y., Tahlan, K., and Bignell, D.R.D. (2020) Functional cross-talk of MbtH-like proteins during thaxtomin biosynthesis in the potato common scab pathogen *Streptomyces scabiei. Front Microbiol*. 11, 585456. doi: 10.3389/fmicb.2020.585456.]. The study concept and experimental methodology were designed by Y. Li, D.R.D Bignell and K. Tahlan. All of the experimental procedures, data analyses and interpretation were carried out by Y. Li. The manuscript was drafted and prepared by Y. Li with editorial input by D.R.D. Bignell and K. Tahlan.

Chapter 4 is a manuscript in preparation for future submission. The study concept and experimental methodology were designed by Y. Li, D.R.D. Bignell and K. Tahlan. Y.

Li conducted all of the experimental work and data analyses, and the manuscript was drafted and prepared by Y. Li with editorial input by D.R.D. Bignell and K. Tahlan.

CHAPTER 2

TxtH is a key component of the thaxtomin biosynthetic machinery in the potato common scab pathogen *Streptomyces scabiei*

Yuting Li, Jingyu Liu, Damilola Adekunle, Luke Bown, Kapil Tahlan and Dawn R.D. Bignell

 Molecular Plant Pathology, 20 (10), 1379-1393. doi: 10.1111/mpp.12843

2.1 Abstract

Streptomyces scabiei causes potato common scab disease, which reduces the quality and market value of affected tubers. The predominant pathogenicity determinant produced by *S. scabiei* is the thaxtomin A phytotoxin, which is essential for common scab disease development. Production of thaxtomin A involves the nonribosomal peptide synthetases (NRPSs) TxtA and TxtB, both of which contain an adenylation (A-) domain for selecting and activating the appropriate amino acid during thaxtomin biosynthesis. The genome of *S. scabiei* 87.22 contains three small MbtH-like protein (MLP)-coding genes, one of which (*txtH*) is present in the thaxtomin biosynthesis gene cluster. MLP family members are typically required for the proper folding of NRPS A-domains and/or stimulating their activities. This study investigated the importance of TxtH during thaxtomin biosynthesis in *S. scabiei*. Biochemical studies showed that TxtH is required for promoting the soluble

expression of both the TxtA and TxtB A-domains in *Escherichia coli*, and amino acid residues essential for this activity were identified. Deletion of *txtH* in *S. scabiei* significantly reduced thaxtomin A production, and deletion of one of the two additional MLP homologs in *S. scabiei* completely abolished production. Engineered expression of all three *S. scabiei* MLPs could restore thaxtomin A production in a triple MLP-deficient strain, while engineered expression of MLPs from other *Streptomyces* spp. could not. Furthermore, the constructed MLP mutants were reduced in virulence compared to wild-type *S. scabiei.* The results of our study confirm that TxtH plays a key role in thaxtomin A biosynthesis and plant pathogenicity in *S. scabiei.*

2.2 Introduction

Over 580 species of *Streptomyces* have been identified to date (Garrity *et al.*, 2007), of which only a very small number have the ability to infect living plant tissue and cause plant diseases (Bignell *et al.*, 2010a). One of the best studied plant-pathogenic species is *Streptomyces scabiei* (syn. *S. scabies*), which causes common scab disease of potato (Bignell *et al.*, 2014; Loria *et al.*, 2006). The main symptom associated with this disease is the formation of superficial, raised or deep-pitted lesions on the tuber surface, and these lesions reduce the market value of affected potatoes, leading to economic losses for potato growers (Dees and Wanner, 2012). As *S. scabiei* is neither tissue nor host specific, it can cause scab disease symptoms on other economically important root crops such as radish, carrot, beet and turnip (Dees and Wanner, 2012). Also, the seedlings of model plants such

as *Arabidopsis thaliana* and *Nicotiana tabacum* can be infected by *S. scabiei*, resulting in root stunting, swelling, necrosis and seedling death (Loria *et al.*, 2006).

The primary pathogenicity determinant produced by *S. scabiei* consists of a family of specialized phytotoxic metabolites called the thaxtomins, which are cyclic dipeptides (King and Calhoun, 2009). Eleven different thaxtomins have been described, of which thaxtomin A is the predominant member produced by *S. scabiei* and other scab-causing pathogens such as *Streptomyces turgidiscabies, Streptomyces acidiscabies, Streptomyces europaeiscabiei* and *Streptomyces stelliscabiei* (King *et al.*, 1989; King and Calhoun, 2009). A positive correlation was established between the pathogenicity of *S. scabiei* strains and their ability to produce thaxtomin A (King *et al.*, 1991), and disruption of thaxtomin biosynthesis in *S. acidiscabies* abolished the ability of the pathogen to cause necrotic lesions on potato tubers (Healy *et al.*, 2000). Thaxtomin A targets the plant cell wall by functioning as a cellulose synthesis inhibitor (Scheible *et al.*, 2003), and its production is induced by cellobiose and cellotriose, which are the smallest subunits of cellulose (Johnson *et al.*, 2007). In *A. thaliana*, thaxtomin A has been shown to affect the expression of genes involved in cell wall synthesis, and it also reduces the number of cellulose synthase complexes in the plant cell plasma membrane (Bischoff *et al.*, 2009). In addition, thaxtomin A elicits an early defense response in *Arabidopsis* by inducing the influx of Ca^{2+} and the efflux of H^{+} ions (Bischoff *et al.*, 2009; Errakhi *et al.*, 2008; Tegg *et al.*, 2005).

The biosynthetic gene cluster responsible for the synthesis of thaxtomin A and related analogs is highly conserved in scab-causing *Streptomyces* spp. and consists of seven genes - *txtA*, *txtB*, *txtC*, *txtD*, *txtE*, *txtH* and *txtR* (Figure 2.1). *txtD* encodes a nitric oxide synthase that generates nitric oxide (NO) from L-arginine, and *txtE* encodes a novel

cytochrome P450 monooxygenase that nitrates L-tryptophan using the NO to produce the intermediate 4-nitro-L-tryptophan (Barry *et al.*, 2012; Johnson *et al.*, 2009). Two nonribosomal peptide synthetases (NRPSs) encoded by *txtA* and *txtB* have been proposed to synthesize thaxtomin D using L-phenylalanine and 4-nitro-L-tryptophan as substrates, respectively (Healy *et al.,* 2000; Johnson *et al.*, 2009; Loria *et al.*, 2008). NRPSs are a family of large proteins that produce nonribosomal peptide molecules with diverse structures and activities. NRPSs consist of multiple enzymatic domains, of which the adenylation domain (A-domain) is responsible for selecting and activating the amino acid substrate for incorporation into the peptide product (Süssmuth and Mainz, 2017). The *txtC* gene encodes a cytochrome P450 monooxygenase that introduces two hydroxyl groups onto the thaxtomin D backbone to generate the final thaxtomin A product (Figure 2.1) (Healy *et al.*, 2002), and *txtR* encodes a cluster-situated regulator that activates the expression of the thaxtomin biosynthetic genes (Joshi *et al.*, 2007). Additionally, a small gene called *txtH* is located between *txtB* and *txtC* and encodes a protein belonging to the MbtH-like protein (MLP) family (Bignell *et al.*, 2010a).

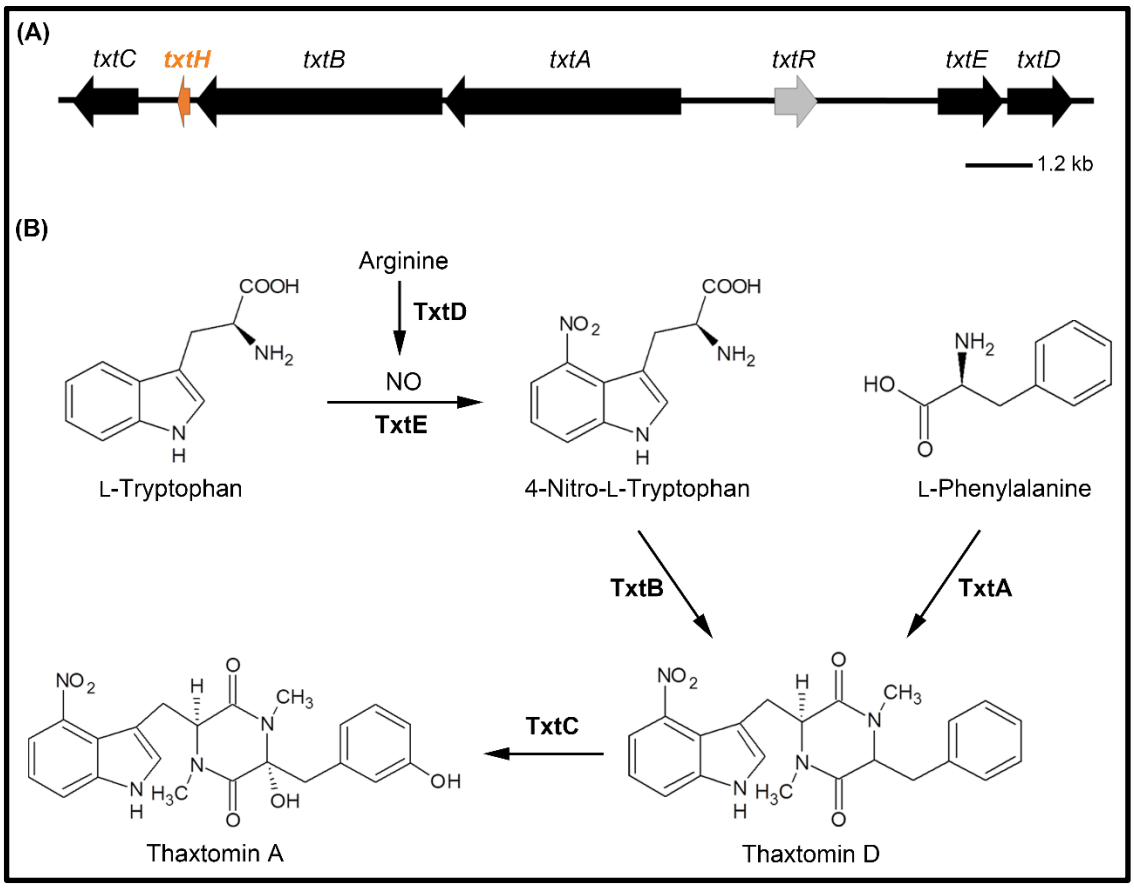


Figure 2.1 (A) Organization of the thaxtomin biosynthetic gene cluster in *Streptomyces scabiei* 87.22. The block arrows represent the genes within the cluster, and the direction of each arrow indicates the direction of transcription. Biosynthetic genes *txtA*, *txtB*, *txtC*, *txtD* and *txtE* are represented in black, the regulatory gene *txtR* is grey, and the MbtH-like protein (MLP)-encoding *txtH* gene is orange. (B) The proposed biosynthetic pathway of thaxtomin A in *S. scabiei* 87.22.

MLPs are small (~70 amino acids) proteins that are usually encoded within NRPS gene clusters (Baltz, 2011). Recent studies have shown that they play essential roles in promoting the proper folding and activity of the NRPS A-domains, though for reasons currently unknown, not all NRPS A-domains require an MLP for proper function (Felnagle *et al.,* 2010; Zhang *et al.,* 2010; Boll *et al.,* 2011; McMahon *et al.,* 2012; Schomer and Thomas, 2017). It has also been revealed that MLPs from different pathways can, in some

instances, functionally complement each other with varying efficiencies (Lautru *et al.*, 2007; Wolpert *et al.,* 2007; Zhang *et al*., 2010; Boll *et al*., 2011; Schomer and Thomas, 2017; Mori *et al.,* 2018).

In the current study, we used multiple approaches to investigate the requirement of TxtH and other MLPs during the biosynthesis of thaxtomin A in *S. scabiei*.

2.3 Materials and Methods

2.3.1 Bacterial strains, cultivation and maintenance

Bacterial strains used in this study are listed in Table 2.1. *E. coli* strains were cultivated at 37°C unless otherwise stated. Liquid cultures were grown with shaking (200 – 250 rpm) in Luria-Bertani (LB) Lennox broth (Fisher Scientific, Ottawa, ON, Canada), low salt LB broth (1% w/v tryptone; 0.5% w/v yeast extract; 0.25% w/v NaCl), super optimal broth (SOB) or super optimal broth with catabolite repression (SOC) medium (New England Biolabs, Whitby, ON, Canada), while solid cultures were grown on LB Lennox (or low salt LB) medium containing 1.5% w/v agar. When necessary, the growth medium was supplemented with 50 µg/mL apramycin (Sigma Aldrich, Oakville, ON, Canada), 50 µg/mL kanamycin or hygromycin B (Millipore Sigma, Canada), or with 25 µg/mL chloramphenicol (Acros Organics, Geel, Belgium) (final concentration). *E. coli* strains were maintained at 4°C for short-term storage or at -80°C as glycerol stocks for long-term storage.

Table 2.1 Bacterial strains used in this study.

Strain	Description	Resistance†	Reference or source
***Escherichia coli* strains**			
DH5α	General cloning host	n/a	Gibco-BRL
NEB5α	DH5α derivative, high efficiency competent cells	n/a	New England Biolabs
BL21(DE3)*ybdZ:aac(3)IV*	BL21(DE3) derivative, *ybdZ* replaced with an apramycin resistance cassette (*aac(3)IV*)	$Apra^R$	Herbst *et al.*, 2013
ET12567/pUZ8002	*dam*$^-$, *dcm*$^-$, *hsdS*$^-$; nonmethylating conjugation host	Kan^R, Cml^R	MacNeil *et al.*, 1992
***Streptomyces scabiei* strains**			
87.22	Wild-type strain	n/a	Loria *et al.*, 1995
87.22/Δmlp_{lipo}_int	Strain 87.22 containing plasmid pIJ12738/Δmlp_{lipo} inserted into the chromosome	$Apra^R$	This study
Δmlp_{lipo}	mlp_{lipo} deletion mutant derivative of strain 87.22	n/a	This study
Δ*txtH*	*txtH* deletion mutant derivative of strain 87.22	$Apra^R$	This study
Δmlp_{lipo}/Δ*txtH*	*txtH* deletion mutant derivative of strain Δmlp_{lipo}	$Apra^R$	This study
Δ*txtH*/Δmlp_{scab}	mlp_{scab} deletion mutant derivative of strain Δ*txtH*	$Apra^R$, Hyg^R	This study
Δmlp_{lipo}/Δ*txtH*/Δmlp_{scab}	mlp_{scab} deletion mutant derivative of strain Δmlp_{lipo}/Δ*txtH*	$Apra^R$, Hyg^R	This study
Other *Streptomyces* strains			
Streptomyces coelicolor A3(2) M145	Source of genomic DNA for amplifying	n/a	Kieser et al., 2000

	the *cdaX* coding sequences		
Streptomyces clavuligerus ATCC27064	Source of genomic DNA for amplifying the *SCLAV_p1293* coding sequence	n/a	ATCC

† $Apra^R$, Kan^R, Cml^R and Hyg^R = apramycin, kanamycin, chloramphenicol and hygromycin resistance, respectively.
n/a = not applicable.

S. scabiei strains were cultured at 28°C unless otherwise indicated. Liquid cultures were typically grown with shaking (200 rpm) in trypticase soy broth (TSB; BD Biosciences, Mississauga, ON, Canada) medium with stainless steel springs. Plate cultures were routinely grown on potato mash agar (PMA; Fyans *et al.*, 2015), nutrient agar (BD Biosciences), soy flour mannitol (SFM) agar (Kieser *et al.*, 2000) and modified yeast extract-malt extract-starch agar (mYMS). The mYMS is the same as YMS (Ikeda et al. 1987) except that it contains Bacto Malt Extract Broth (BD Biosciences) in place of malt extract. When necessary, the growth medium was supplemented with 50 µg/mL apramycin (Sigma Aldrich), 60 µg/mL nalidixic acid (Fisher Scientific), or 25 µg/mL thiostrepton (Sigma Aldrich, Canada) (final concentration). Seed cultures for RNA extraction were prepared by inoculating 100 µL of a *S. scabiei* spore stock into 5 mL of TSB in a 50 mL spring flask followed by incubation for 48 hours until dense mycelial growth was obtained. The seed cultures (50 µL) were then spread onto the surface of cellophane discs (75 mm diameter) on oat bran agar (Johnson *et al.*, 2007) containing 0.35% w/v cellobiose (OBAC), after which the plates were incubated for 42 hours. Cultures for analysis of thaxtomin A production were prepared by inoculating 50 µL of TSB seed cultures into 5 mL of oat bran broth containing 0.35% w/v cellobiose (OBBC; Johnson *et al.*, 2009) in 6-well tissue

culture plates (Fisher Scientific) and then incubating at 25°C and 125 rpm for 7 days. Strains used for potato tuber slice bioassays were cultured at 28°C for 14 days on yeast extract-malt extract-starch (YMS) agar (Ikeda *et al.*, 1987) that had been modified by replacing the malt extract with Bacto Malt Extract Broth (BD Biosciences).

2.3.2 Plasmids, primers and DNA manipulation

Plasmids and cosmids used in this study are listed in Table 2.2. Standard molecular biology procedures were implemented for all DNA manipulations performed in this study (Sambrook and Russell, 2001). Restriction enzymes were purchased from New England Biolabs unless otherwise stated. PCR was routinely performed using Phusion DNA polymerase (New England Biolabs) according to the manufacturer's instructions, except that 5% v/v DMSO was included in the reactions. All oligonucleotide primers used for cloning, PCR, site-directed mutagenesis and sequencing were purchased from Integrated DNA Technologies (Coralville, IA, USA) and are listed in Supplementary Table 2.1. DNA sequencing was performed by The Centre for Applied Genomics (Toronto, Canada). *Streptomyces* genomic DNA was isolated from mycelia harvested from 1-2 days old TSB cultures using the QIAamp® DNA mini kit as per the manufacturer's protocol (QIAgen Inc, Toronto, ON, Canada).

Table 2.2 Plasmids and cosmids used in this study.

Plasmid or cosmid	**Description**	**Resistance**[†]	**Reference or source**
pGEM-T	General cloning vector	Amp^R	Promega Corporation
pGEM-T EASY	General cloning vector	Amp^R	Promega Corporation
pGEM-T EASY/Δ*mlp*$_{lipo}$	pGEM-T EASY derivative containing a 3725 bp insert with a deletion of the *mlp*$_{lipo}$ gene	Amp^R	This study
pET28b	N- or C- terminal 6×histidine fusion tag protein expression vector with T7 promoter and *lac* operator	Kan^R	Novagen
pET28b/HIS$_6$-*txtH*	pET28b derivative containing a DNA fragment for expression of the HIS$_6$-TxtH protein	Kan^R	This study
pET28b/HIS$_6$-*mlp*$_{lipo}$	pET28b derivative containing a DNA fragment for expression of the HIS$_6$-MLP$_{lipo}$ protein	Kan^R	This study
pET28b/HIS$_6$-*mlp*$_{scab}$	pET28b derivative containing a DNA fragment for expression of the HIS$_6$-MLP$_{scab}$ protein	Kan^R	This study
pET28b/*txtH*	pET28b derivative containing a DNA fragment for expression of the untagged TxtH protein	Kan^R	This study
pACYCDuet-1	N- terminal 6×histidine fusion tag expression vector with T7 promoter and *lac* operator	Cml^R	Novagen
pACYCDuet-1/HIS$_6$-*txtA*^{A}	pACYCDuet-1 derivative containing a DNA fragment for expression of the HIS$_6$-TxtA^A protein	Cml^R	This study
pACYCDuet-1/HIS$_6$-*txtB*^{A}	pACYCDuet-1 derivative containing a DNA fragment for expression of the HIS$_6$-TxtB^A protein	Cml^R	This study
pIJ12738	Conjugative plasmid, non-replicative in *Streptomyces*, containing MCS and *I-Sce*I site	$Apra^R$	Fernández-Martínez and Bibb, 2014
pIJ12738/Δ*mlp*$_{lipo}$	pIJ12738 derivative containing two flanking regions of *mlp*$_{lipo}$	$Apra^R$	This study

pIJ12742	Conjugative plasmid containing the temperature-sensitive replication origin and the codon optimised *I-Sce*I gene under the control of the strong constitutive *ermE*p* promoter	Thio^R	Fernández-Martínez and Bibb, 2014
pIJ773	Template for PCR amplification of the *aac(3)IV-oriT* cassette used for PCR targeting	Apra^R	Gust *et al.*, 2003a
pIJ10700	Template for PCR amplification of the *hyg-oriT* cassette used for PCR targeting	Hyg^R	Gust *et al.*, 2003b
Cosmid 1989	SuperCos1 derivative containing the *S. scabiei* thaxtomin A biosynthetic gene cluster	Amp^R, Kan^R	Zhang *et al.*, 2016
Cosmid 57	SuperCos1 derivative containing the *S. scabiei* mlp_{scab} gene	Amp^R, Kan^R	This study
Cosmid 1989/Δ*txtH*	Cosmid 1989 derivative containing the *aac(3)IV-oriT* cassette in place of the *txtH* gene	Amp^R, Kan^R, Apra^R	This study
Cosmid 57/Δmlp_{scab}	Cosmid 57 derivative containing the *hyg-oriT* cassette in place of the mlp_{scab} gene	Amp^R, Kan^R, Hyg^R	This study
pRLDB50-1a	Overexpression plasmid containing the strong constitutive *ermE*p* promoter	Apra^R, Thio^R	Bignell *et al.*, 2010
pRLDB50-1a/*txtH*	pRLDB50-1a derivative containing the *S. scabiei txtH* gene	Apra^R, Thio^R	This study
pRLDB50-1a/mlp_{lipo}	pRLDB50-1a derivative containing the *S. scabiei* mlp_{lipo} gene	Apra^R, Thio^R	This study
pRLDB50-1a/mlp_{scab}	pRLDB50-1a derivative containing the *S. scabiei* mlp_{scab} gene	Apra^R, Thio^R	This study
pRLDB50-1a/*SCLAV_p1293*	pRLDB50-1a derivative containing the *S. clavuligerus SCLAV_p1293* gene	Apra^R, Thio^R	This study
pRLDB50-1a/*cdaX*	pRLDB50-1a derivative containing the *S. coelicolor cdaX* gene	Apra^R, Thio^R	This study

† Amp^R, Apra^R, Kan^R, Cml^R and Thio^R = ampicillin, apramycin, kanamycin, chloramphenicol and thiostrepton resistance, respectively.

2.3.3 Construction of protein expression plasmids

Plasmids were constructed for overexpression of TxtH in *E. coli* with and without an N-terminal 6×histidine (HIS_6) tag as well as for overexpression of N-terminal HIS_6-tagged MLP_{lipo} and MLP_{scab}. The *txtH* gene was PCR-amplified using Cosmid 1989 as template and using primers PL150 and PL36 for construction of the untagged TxtH expression plasmid, and PL35 and PL36 for construction of the HIS_6-TxtH expression plasmid. The resulting PCR products were directly cloned into the expression vector pET28b via the *Nde*I/*EcoR*I and *Nco*I/*Eco*RI restriction sites to give pET28b/HIS_6-*txtH* and pET28b/*txtH*, respectively. mlp_{lipo} was PCR-amplified from genomic DNA using primers PL163 and PL164, and mlp_{scab} was PCR-amplified from Cosmid 57 using primers PL165 and PL166. The PCR products were directly cloned into the *Nde*I/*EcoR*I restriction sites of pET28b to give pET28b/HIS_6-mlp_{lipo} and pET28b/HIS_6-mlp_{scab}.

Plasmids were also constructed for overexpression of the TxtA and TxtB A-domains (referred to herein as $TxtA^A$ and $TxtB^A$) as N-terminal HIS_6-tagged proteins. The DNA sequences encoding $TxtA^A$ and $TxtB^A$ were PCR amplified using the primer pairs PL37/PL38 and PL40/PL41, respectively, and using Cosmid 1989 as template. The products were cloned into the pGEM-T vector as per the manufacturer's instructions, after which the inserts were released by digestion with *Eco*RI and *Hin*dIII and were cloned into similarly digested pACYCDuet-1 to give pACYCDuet-1/HIS_6-$txtA^A$ and pACYCDuet-1/HIS_6-$txtB^A$. The cloned inserts in all constructed expression vectors were verified by DNA sequencing.

2.3.4 Site-directed mutagenesis of TxtH

Site-directed mutagenesis of TxtH was performed using the QuikChange II Site-directed Mutagenesis Kit (Agilent Technologies Canada, Inc., Mississauga, ON, Canada) as per the manufacturer's instructions. Mutagenic primers for the desired mutation was designed online with QuikChange® Primer Design Program (https://www.genomics.agilent.com/primerDesignProgram.jsp). The desired mutations were verified by DNA sequencing.

2.3.5 Co-expression of HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$ with MLPs

The BL21(DE3) *ybdZ:aac(3)IV* bacterial strain was used for co-expression of HIS_6-$TxtA^A$ or HIS_6-$TxtB^A$ with tagged or untagged MLP proteins. Strains containing either pACYCDuet-1/HIS_6-$txtA^A$ or pACYCDuet-1/HIS_6-$txtB^A$ with and without pET28b/*txtH*, pET28b/HIS_6-*txtH* (wild-type or point mutants), pET28b/HIS_6-mlp_{lipo} or pET28b/HIS_6-mlp_{scab} were grown overnight in 3 mL of LB medium supplemented with 1% glucose, apramycin and chloramphenicol. Kanamycin was additionally included for strains containing the MLP expression plasmids. The overnight cultures were subcultured (1% v/v) into 50 mL of fresh LB containing appropriate antibiotics, and the cultures were incubated at 37°C and 200 rpm until the OD_{600} was 0.4 - 0.6. Then, the cells were induced with 1 mM isopropyl β-D thiogalactopyranoside (IPTG) and were further incubated at 16°C and 200 rpm for 48 hours. Cells from 1 mL of culture were harvested by centrifugation and were resuspended in 200 µL of 50 mM Tris-HCl (pH 8.0) containing 1× cOmplete EDTA-free

protease inhibitor (Roche Diagnostics, Laval, QC, Canada). The cells were then lysed by sonication for 25 seconds (10 seconds pulses alternating with 10 seconds pauses, 40% Amp) and the cell debris was removed by centrifugation (1 minute at 16,000 rpm). The supernatants containing soluble proteins were collected and the protein concentration was quantified using a Bradford protein assay kit (Fisher Scientific).

2.3.6 Western blot analysis

Soluble protein extracts (10 μg) were subjected to standard sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) before being transferred to Amersham™ Hybond™ ECL membrane (GE Healthcare Canada, Inc., Mississauga, ON, Canada) as per the manufacturer's instructions. Membranes were blocked overnight in TBS-T buffer (50 mM Tris-HCl pH 7.6, 150 mM NaCl and 0.05% v/v Tween 20) containing 5% w/v skim milk, and were then incubated with 6×HIS Epitope Tag Antibody (mouse IgG2b) (Fisher Scientific) at a 1:2000 dilution. The membranes were washed several times with TBS-T buffer and were then incubated with the secondary antibody (goat anti-mouse IgG2b, HRP conjugate) (Fisher Scientific) at a 1:2000 dilution. The membranes were processed using ECL™ Western blotting detection reagent (GE Healthcare) and were visualized an ImageQuant LAS4000 Biomolecular Imager (GE Healthcare).

2.3.7 Construction of an MLP-deficient strain of *S. scabiei*

A marker-less deletion mutant of the mlp_{lipo} gene was generated using the meganuclease *I-Sce*I system (Fernández-Martínez and Bibb, 2014). A 1766 bp region

upstream of *mlp*$_{lipo}$ (5′ *mlp*$_{lipo}$) was amplified using *S. scabiei* 87.22 genomic DNA as template and using primers PL3 and PL4 to generate a DNA fragment with terminal *Xba*I and *Bam*HI sites. A 1959 bp region downstream of the gene (3′ *mlp*$_{lipo}$) was separately amplified using the same template and primers PL5 and PL6 to generate a DNA fragment with terminal *Bam*HI and *Eco*RI sites. These two flanking fragments were each cloned into pGEM-T EASY (Table 2.2) generating pGEM-T EASY/5′ *mlp*$_{lipo}$ and pGEM-T EASY/3′ *mlp*$_{lipo}$, the inserts of which were confirmed by DNA sequencing. The 3′ *mlp*$_{lipo}$ insert was then released following digestion with *Eco*RI and *Bam*HI and was cloned into similarly digested pGEM-T EASY/5′ *mlp*$_{lipo}$ to generate pGEM-T EASY/Δ*mlp*$_{lipo}$, which contained a 3725 bp insert with a deletion of the *mlp*$_{lipo}$ gene. Next, the 3725 bp insert was released by digestion with *Xba*I and *Eco*RI and was cloned into similarly digested pIJ12738 to give pIJ12738/Δ*mlp*$_{lipo}$, which was then introduced into *S. scabiei* 87.22 by intergeneric conjugation with *E. coli* as described before (Kieser *et al.*, 2000). Apramycin-resistant exconjugants (assigned as 87.22/Δ*mlp*$_{lipo}$_int) were selected and verified by PCR using primer PL62 and PL63. Then, the delivery vector pIJ12742 containing the codon optimized *I-Sce*I gene under the control of the *ermE*p* promoter was introduced into verified *S. scabiei* 87.22/Δ*mlp*$_{lipo}$_int by conjugation with *E. coli*. The exconjugants were cultured on PMA at 37°C in order to promote the loss of pIJ12742, which was confirmed by screening for sensitivity to thiostrepton. Spores of thiostrepton-sensitive exconjugants were then serially diluted in sterile water and were plated onto PMA plates to obtain single colonies, which were then screened for sensitivity to apramycin. Successful deletion of *mlp*$_{lipo}$ was confirmed by PCR (Supplementary Figure 2.1).

The Redirect PCR targeting system (Gust *et al.*, 2003a, b) was used to construct the Δ*txtH*, Δmlp_{lipo}/Δ*txtH,* Δ*txtH*/Δmlp_{scab} and Δmlp_{lipo}/Δ*txtH*/Δmlp_{scab} mutant strains. The *txtH* gene on Cosmid 1989 was replaced with an extended apramycin resistance cassette [*aac(3)IV-oriT*] that was PCR-amplified using pIJ773 as template and using primers DRB627 and DRB628. The mlp_{scab} gene on Cosmid 57 was replaced with an extended hygromycin resistance cassette [*hyg-oriT*] that was PCR-amplified using pIJ10700 as template and using primers PL153 and PL154. The Δ*txtH* and Δmlp_{scab} mutant cosmids were verified by PCR (Supplementary Figure 2.2; Supplementary Figure 2.4) and were then introduced into *S. scabiei* by intergeneric conjugation with *E. coli*. The resulting mutant strains were analyzed by PCR to confirm replacement of the target genes (Supplementary Figure 2.3 and 2.4).

2.3.8 Construction of MLP overexpression plasmids

The *txtH,* mlp_{lipo} and mlp_{scab} genes from *S. scabiei,* together with the *SCLAV_p1293* and *cdaX* MLP-encoding genes from *S. clavuligerus* and *S. coelicolor,* respectively, were PCR-amplified using Cosmid 1989 (for *txtH*) or genomic DNA (for mlp_{lipo}, mlp_{scab}, *SCLAV_p1293* and *cdaX*) as template and using gene-specific primers with *Bam*HI and *Xba*I restriction sites added (Supplementary Table 2.1). The resulting products were digested with *Bam*HI and *Xba*I and were ligated into similarly digested pRLDB50-1a (Bignell *et al.* 2010b) to generate pRLDB50-1a/*txtH*, pRLDB50-1a/mlp_{lipo}, pRLDB50-1a/mlp_{scab}, pRLDB50-1a/ *SCLAV_p1293* and pRLDB50-1a/*cdaX* (Table 2.2). The expression plasmids along with the control plasmid (pRLDB50-1a) were then introduced

into *S. scabiei* 87.22 and the $\Delta mlp_{lipo}/\Delta txtH/\Delta mlp_{scab}$ mutant by intergeneric conjugation with *E. coli*.

2.3.9 Quantification of thaxtomin A production

Thaxtomin A was extracted from *S. scabiei* OBBC cultures as described by Fyans et al. (2016). The extracts were analyzed using an Agilent 1260 Infinity Quaternary LC system (Agilent Technologies Canada Inc.) with a Poroshell 120 EC-C18 column (4.6 × 50 mm, 2.7 µm particle size; Agilent Technologies Canada, Inc.) held at a constant temperature of 40°C. An isocratic mobile phase consisting of 30% acetonitrile and 70% water at a constant flow rate of 1.0 mL/min was used for metabolite separation, and metabolites were monitored using a detection wavelength of 380 nm. Quantification of thaxtomin A in the culture extracts was by reverse phase HPLC using a standard curve that was constructed from known amounts of a pure thaxtomin A standard (Sigma Aldrich). The thaxtomin A production levels were normalized using dry cell weights (DCWs) as described before (Fyans *et al.*, 2016) and were reported as ng thaxtomin A/mg DCW. Statistical analysis of the results was conducted in Minitab 18 using one-way ANOVAs with *a posteriori* multiple comparisons of least squared means performed using the Tukey test. P values ≤ 0.05 were considered statistically significant in all analyses.

2.3.10 LC-HRESIMS analysis of *S. scabiei* culture extracts

Liquid chromatography-high resolution electrospray ionization mass spectrometry (LC-HRESIMS) analysis of *S. scabiei* culture extracts was performed at the Memorial

University Centre for Chemical Analysis, Research and Training using an Agilent 1260 Infinity HPLC system interfaced to an Agilent 6230 orthogonal time-of-flight mass analyzer. Separation was achieved using a ZORBAX SB-C18 analytical column (4.6 × 150 mm, 5 µm particle size) held at a constant temperature of 40°C and an isocratic mobile phase consisting of 30% acetonitrile and 70% water at a constant flow rate of 1.0 mL/min. Metabolites were monitored by absorbance at 380 nm and by electrospray ionization MS in negative ion mode.

2.3.11 Potato tuber slice bioassay

The virulence phenotype of *S. scabiei* strains was assessed using a potato tuber slice bioassay as described before (Loria *et al.*, 1995). *S. scabiei* strains were cultured on mYMS agar for 14 days until well sporulated. Agar plugs were then prepared from the plates and were inverted onto the tuber slices. The tuber slices were incubated at room temperature (~22-25°C) in the dark in a moist chamber and were photographed after 10 days. The assay was performed twice in total.

2.3.12 Total RNA isolation

S. scabiei mycelia (100-200 mg) from 42 hours OBAC plates were placed into sterile 1.7 mL microcentrifuge tubes and were flash frozen in a dry ice/ethanol bath and then stored at -80°C. Total RNA was isolated using an innuPREP RNA Mini Kit 2.0 and a SpeedMill PLUS tissue homogenizer (Analytik Jena AG, Jena, Germany) as per the manufacturer's instructions. The resulting RNA samples were treated with DNase I (New

England Biolabs) as directed by the manufacturer to remove trace amounts of genomic DNA, after which the DNase-treated RNA samples were quantified using a NanoDrop™ 1000 Spectrophotometer (Fisher Scientific). The integrity of the RNA was confirmed by agarose gel electrophoresis using a 1.2 % w/v RNase-free agarose gel in 1× TBE (Tris-Borate-EDTA) buffer. The RNA samples were stored at -80°C.

2.3.13 Reverse transcription PCR

Reverse transcription (RT) was performed using SuperScript IV reverse transcriptase (Fisher Scientific) with 2 µg of DNase-treated total RNA and random hexamer primers as per the manufacturer's instructions. A negative control reaction lacking the reverse transcriptase enzyme was included to verify the absence of genomic DNA in the RNA samples. RNA was removed from the synthesized cDNA by adding 1 uL of RNAse H and incubating at 37°C for 20 minutes. PCR was performed using 2 µL of the cDNA template. Amplification was conducted using Taq DNA polymerase (New England Biolabs) with 1× Standard Taq Reaction Buffer, 250 µM dNTPs, 0.5 µM of gene-specific primers (Supplementary Table 2.1) and 5% v/v DMSO. The PCR reactions were initiated by denaturing at 95°C for 2 minutes followed by 22 (*txtA, txtB, txtC, txtH*), 25 (*gyrA*) or 27 (mlp_{lipo}, mlp_{scab}) cycles of 95°C for 15 seconds, 60°C for 30 seconds and 68°C for 15 seconds. After the amplification, 10 µL of each PCR product was analyzed on a 1% agarose gel by electrophoresis.

2.3.14 Bioinformatics analysis

Identification of the adenylation domain within the TxtA and TxtB amino acid sequences was performed using the Pfam database (http://pfam.xfam.org/) (Finn et al. 2013). TxtH homologs were identified using the NCBI Protein Basic Local Alignment Search Tool (BLASTP) (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Amino acid sequence alignment of TxtH and other MLPs was performed using ClustalW within the Geneious version 6.1.2 software (Biomatters Inc., Newark, NJ, USA). The accession numbers for the protein sequences used in the alignment are listed in Supplementary Table 2.1. The MLP phylogenetic tree was constructed by maximum likelihood with the MEGA 7 software (Kumar *et al.*, 2016) using the Whelan and Goldman plus gamma (WAG+G) substitution model (Whelan and Goldman, 2001). Bootstrap analyses were performed with 1000 replicates.

2.4 Results and Discussion

2.4.1 Bioinformatics analysis of *S. scabiei* TxtH

The TxtH amino acid sequence was aligned with that of other MLP homologues from the database, including some that have been previously characterized (Figure 2.2A; Supplementary Table 2.2). Pairwise comparisons revealed that TxtH shares the greatest degree of amino acid identity (100%) with the corresponding homologues from the thaxtomin biosynthetic gene clusters in the potato scab pathogens *S. acidiscabies* and *S. europaeiscabiei* (Supplementary Table 2.3). In contrast, the TxtH homologue from the thaxtomin biosynthetic gene cluster of another scab pathogen, *S. turgidiscabies,* shares only

80% amino acid identity with the *S. scabiei* TxtH. This is consistent with a previous phylogenetic analysis which suggested that the thaxtomin biosynthetic gene clusters from *S. scabiei* and *S. acidiscabies* are more closely related to each other than to the *S. turgidiscabies* gene cluster (Huguet-Tapia *et al.*, 2016). As expected, the TxtH homologues from the pathogenic *Streptomyces* spp. formed a well-supported clade in the constructed phylogenetic tree (Figure 2.2B). Interestingly, an MLP (ACM01_RS10820) from the non-pathogenic species *Streptomyces viridochromogenes* also clustered together with the TxtH homologues from the pathogenic species and showed a very high degree of amino acid identity (97%) with TxtH from *S. scabiei, S. europaeiscabiei* and *S. acidiscabies* (Figure 2.2B and Supplementary Table 2.3). An analysis of the *S. viridochromogenes* genome sequence (accession number PRJNA238534) revealed that the MLP is encoded in the vicinity of four other genes that show strong similarity to thaxtomin biosynthetic genes, though three of the genes appear to be pseudogenes. Two other MLPs encoded in the *S. scabiei* 87.22 genome, SCAB_3331 (herein referred to as MLP_{lipo}) and SCAB_85461 (herein referred to as MLP_{scab}), both share only 52.3% amino acid identity with TxtH (Supplementary Table 2.3) and cluster together in a separate clade as compared to the one containing TxtH (Figure 2.2B). The MLP_{lipo}-encoding gene is located within an NRPS gene cluster that is responsible for the biosynthesis of a putative lipopeptide metabolite (Yaxley 2009), whereas the MLP_{scab}-encoding gene is localized within the NRPS gene cluster that synthesizes the siderophore scabichelin (Kodani *et al.*, 2013). An orphan MLP (MXAN_3118) from *Myxococcus xanthus* DK 1622 showed the least amino acid identity (33.8%) with TxtH in the pairwise comparison (Supplementary Table 2.3). A recent study showed that MXAN_3118, which is not located within or near any NRPS biosynthetic gene

clusters, can interact *in vivo* and *in vitro* with several different NRPSs in *M. xanthus* (Esquilin-Lebron *et al.*, 2018).

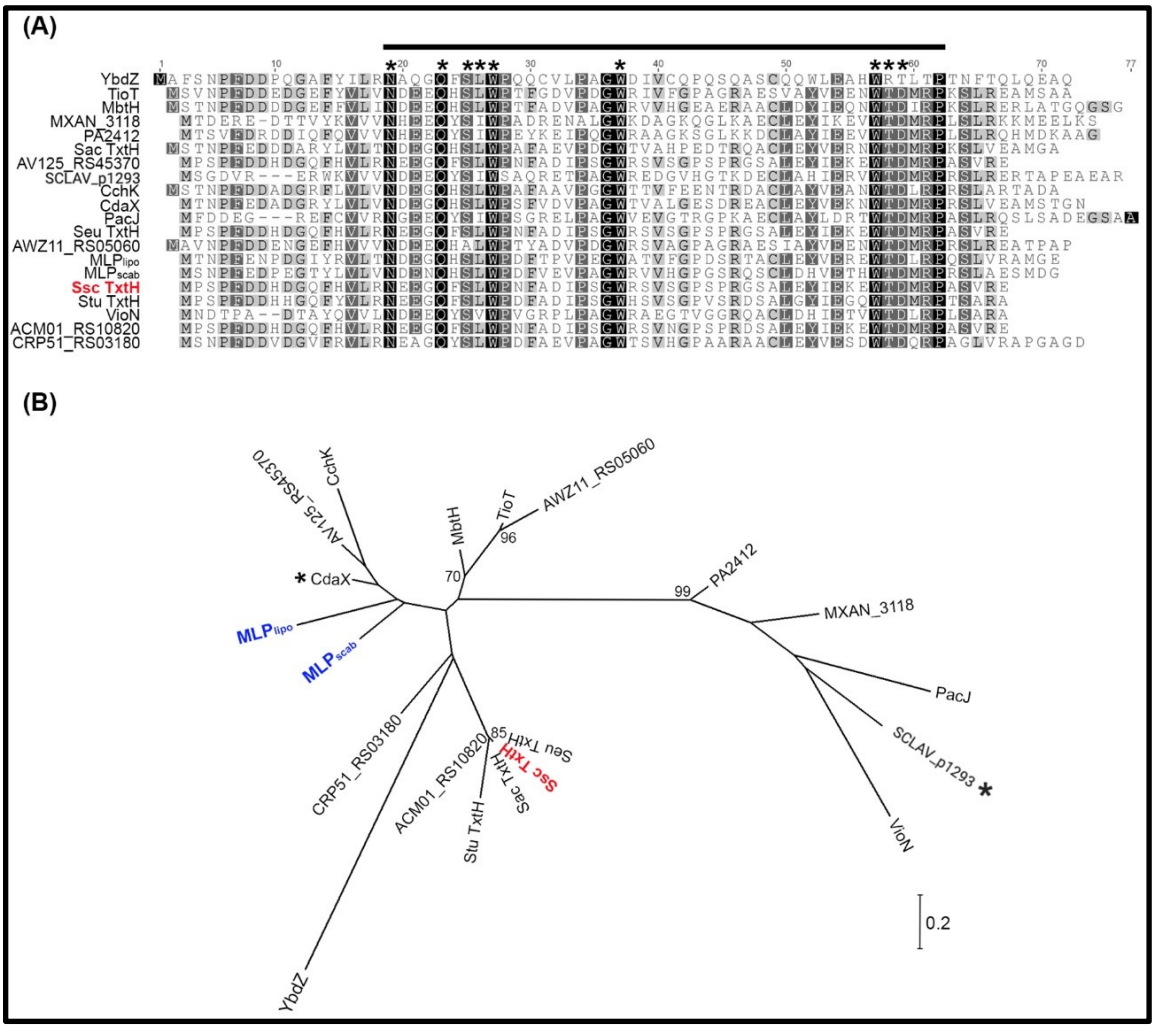


Figure 2.2 (A) Amino acid alignment of TxtH from *Streptomyces scabiei* with other MbtH-like protein (MLP) homologues. Highly conserved amino acids are highlighted as follows: black, 100% identity; dark grey, 80–99% identity; grey, 60–79% identity; light grey, <60% identity. The MLP signature sequence is indicated by the black line above the alignment, and the conserved residues subjected to mutation in the *S. scabiei* TxtH are indicated with the asterisks. (B) Phylogenetic analysis of the MLP homologues. The tree was constructed using the maximum likelihood algorithm, and bootstrap values >50% for 1000 repetitions are shown. The scale bar indicates the number of amino acid substitutions per site. The *S. scabiei* TxtH is highlighted in red, while the other MLPs encoded in the *S. scabiei* genome are shown in blue. The non-cognate MLPs used for the complementation experiments are

indicated with the black asterisks. Ssc, *Streptomyces scabiei*; Sac, *Streptomyces acidiscabies*; Stu, *Streptomyces turgidiscabies*; Seu, *Streptomyces europaeiscabiei*.

Baltz (2011) previously proposed a signature sequence [NxExQxSxWP$(x)_5$PxGW$(x)_{12}$L$(x)_6$WTDxRP] consisting of multiple amino acid residues that are invariant in most MLPs, all of which are also conserved in TxtH and in other MLP homologues analyzed here (Figure 2.2A). Structural analysis of the PA2412 MLP from *Pseudomonas aeruginosa* PAO1 revealed that several of these residues, including the three highly conserved tryptophan residues, lie on one face of the protein, which is thought interact with conserved components of the cognate NRPS (Drake *et al.*, 2007). The structure of SlgN1, a 3-methylaspartate-adenylating enzyme with an MLP domain at its N-terminus, revealed that two of the conserved tryptophan residues (W25 and W35) from the MLP domain are located at the interface between the MLP and the A-domain and are important for this interaction (Herbst *et al.*, 2013). Analysis of mutants defective in equivalent residues (W22A/W32A) in another MLP, PacJ, showed that they contribute to PacJ's ability to form a complex with the cognate PacL NRPS to stimulate the adenylation activity of the synthetase (Zhang *et al.*, 2010). Based on these studies, we predict that the conserved amino acid residues in TxtH play an important role in its interaction with the thaxtomin NRPS in *S. scabiei.*

2.4.2 TxtH is required for promoting the solubility of the TxtA and TxtB A-domains

Previously it was shown that YbdZ, an MLP encoded in the enterobactin biosynthetic gene cluster of *E. coli*, can interact with adenylating enzymes from different

NRPS biosynthesis pathways (Felnagle *et al.*, 2010). Therefore, the TxtA and TxtB A-domains were expressed as N-terminal HIS_6-tagged proteins in an *E. coli ybdZ* mutant [BL21(DE3)*ybdZ:aac(3)IV*] to avoid any potential interference caused by YbdZ during co-expression studies using TxtH (Table 2.1). Each A-domain was expressed in the presence or absence of TxtH, which itself either contained or lacked an N-terminal HIS_6-tag, to rule out any influence that the tag might have on the function of TxtH. The ability of TxtH to promote the solubility of each A-domain was then determined by Western blot analysis of isolated soluble protein fractions using an anti-HIS antibody.

As shown in Figure 2.3A, only trace levels of soluble HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$ protein were detected in *E. coli* when expressed in the absence of TxtH, whereas both proteins were readily detectable in soluble form when co-expressed with the MLP. The solubility promoting activity of TxtH was observed regardless of whether or not the protein contained an N-terminal HIS_6 tag (Figure 2.3A), indicating that the tag did not interfere with the activity of the protein. Our results therefore suggest that TxtH likely functions as a chaperone that is essential for the proper folding of both A-domains in the thaxtomin NRPSs, a role that is consistent with that proposed for other MLPs (Zhang *et al.*, 2010; Imker *et al.*, 2010; Zolova and Garneau-Tsodikova, 2012).

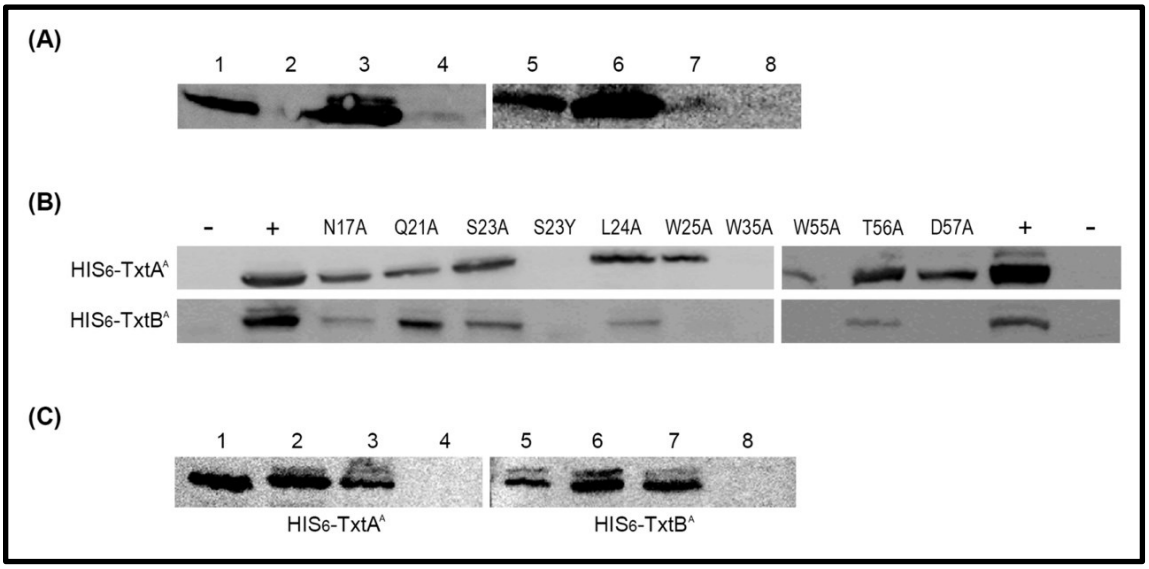


Figure 2.3 (A) Western blot analysis of soluble HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$ expressed in the presence and absence of His-tagged and untagged TxtH. Lanes: 1, HIS_6-$TxtB^A$ co-expressed with HIS_6-TxtH; 2, HIS_6-$TxtB^A$ expressed without HIS_6-TxtH; 3, HIS_6-$TxtA^A$ co-expressed with HIS_6-TxtH; 4, HIS_6-$TxtA^A$ expressed without HIS_6-TxtH; 5, HIS_6-$TxtB^A$ co-expressed with TxtH; 6, HIS_6-$TxtA^A$ co-expressed with TxtH; 7, HIS_6-$TxtB^A$ expressed without TxtH; 8, HIS_6-$TxtA^A$ expressed without TxtH. (B) Western blot analysis of soluble HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$ that was co-expressed with wild-type and mutant His_6-TxtH proteins. The lanes corresponding to the different HIS_6-TxtH point mutants are indicated, and lanes containing A-domain produced in the presence (+) or absence (–) of wild-type HIS_6-TxtH are also shown. (C) Western blot analysis of soluble HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$ expressed in the absence of an MbtH-like protein (MLP) (lanes 4 and 8) or co-expressed with HIS_6-TxtH (lanes 3 and 7), HIS_6-MLP_{lipo} (lanes 2 and 6) or HIS_6-MLP_{scab} (lanes 1 and 5).

To further explore the role of the highly conserved amino acid residues in the MLP signature sequence of TxtH, we constructed several HIS_6-TxtH point mutants (N17A, Q21A, S23A, S23Y, L24A, W25A, W35A, W55A, T56A and D57A) and then co-expressed each mutant protein with HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$. As shown in Figure 2.3B, the solubility of both the HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$ proteins was reduced or abolished when co-expressed with all of the TxtH point mutants. Of particular note is the S23Y mutation, which resulted in complete loss of soluble protein for both A-domains. Herbst

and colleagues showed that the same mutation in the MLP domain of the SlgN1 hybrid adenylase resulted in a 5-fold reduction in adenylation activity of the enzyme, most likely due to impairment of the interaction between the MLP and adenylation domains by the bulky tyrosyl residue (Herbst *et al.*, 2013). In contrast, the S23A mutation in TxtH caused a drastic reduction of soluble HIS_6-$TxtB^A$ protein but did not lead to a complete loss of soluble protein, and it only slightly reduced the solubility of the HIS_6-$TxtA^A$ protein (Figure 2.3B). This is possibly due to the fact that an alanine side chain is less bulky than a tyrosine side chain and may therefore cause less steric interference during the interaction of the MLP with the A-domains. All three highly conserved tryptophan residues in TxtH (W25, W35, W55) (Figure 2.2A) were found to be essential for promoting the solubility of HIS_6-$TxtB^A$, whereas only W35 and W55 are essential for promoting HIS_6-$TxtA^A$ solubility (Figure 2.3B). The W55 residue is part of the highly conserved WTDxRP motif, which in the *P. aeruginosa* PA2412 occurs between two alpha helices and was proposed to play a role in the proper orientation of the C-terminal helix (Drake *et al.*, 2007), whereas in MbtH from *M. tuberculosis* the motif lies within a disordered region (Buchko *et al.*, 2010). Our results show that in addition to W55, two other residues within this motif (T56, D57) are critical for the ability of TxtH to promote the solubility of HIS_6-$TxtB^A$, whereas neither residue is essential for obtaining soluble HIS_6-$TxtA^A$, though HIS_6-$TxtA^A$ solubility was clearly affected in the presence of these point mutants. Other TxtH residues that were found to be are essential for promoting the solubility of HIS_6-$TxtB^A$ are N17 and L24. Overall, our results show that all of the highly conserved amino acid residues found in the MLP signature sequence are important for the solubility-promoting activity of TxtH. We anticipate that structural studies examining the interaction of TxtH with each A-domain

will provide further insights into the specific function of these residues during such interactions.

2.4.3 Loss of MLPs abolishes thaxtomin A production in *S. scabiei*

To examine the *in vivo* role of *txtH* in the thaxtomin A biosynthetic pathway, we deleted *txtH* from the *S. scabiei* chromosome (Supplementary Figure 2.3). Four mutant isolates were examined for thaxtomin A production, and all were found to produce significantly less thaxtomin A as compared to the wild-type strain (Figure 2.4A and B). Production in the Δ*txtH*1 mutant isolate was partially restored when *txtH* was expressed from an integrative plasmid using the strong, constitutive *ermE*p* promoter (Figure 2.4C). Notably, two other metabolites with retention times of 3.82 and 4.64 minutes were found to accumulate at very low levels in the Δ*txtH* mutant isolates but not in wild-type *S. scabiei* (Figure 2.4B, peaks ▼ and ▽). LC-HRESIMS analysis of the Δ*txtH*1 mutant culture extract in negative ion mode revealed a pseudomolecular $[M-H]^-$ ion at *m/z* 421.1524 for peak ▼ and a pseudomolecular $[M-H]^-$ ion at *m/z* 405.1577 for peak ▽, which is consistent with the accumulation of thaxtomin B and D, respectively (King and Calhoun, 2009). Thaxtomin D was previously reported to accumulate in a Δ*txtC* mutant of *S. acidiscabies* (Healy *et al.*, 2002), suggesting that there may be some polar effects on the expression of *txtC* caused by the deletion of *txtH,* even though the orientation of the inserted apramycin resistance cassette was the same as the original *txtH* gene (Supplementary Figure 2.2, 2.3). Indeed, semi-quantitative RT-PCR analysis showed that the *txtC* transcription level was reduced in the Δ*txtH*1 mutant compared to the wild-type strain, though expression of *txtC*

could still be detected in the mutant, especially at higher PCR cycle numbers (Figure 2.5; data not shown).

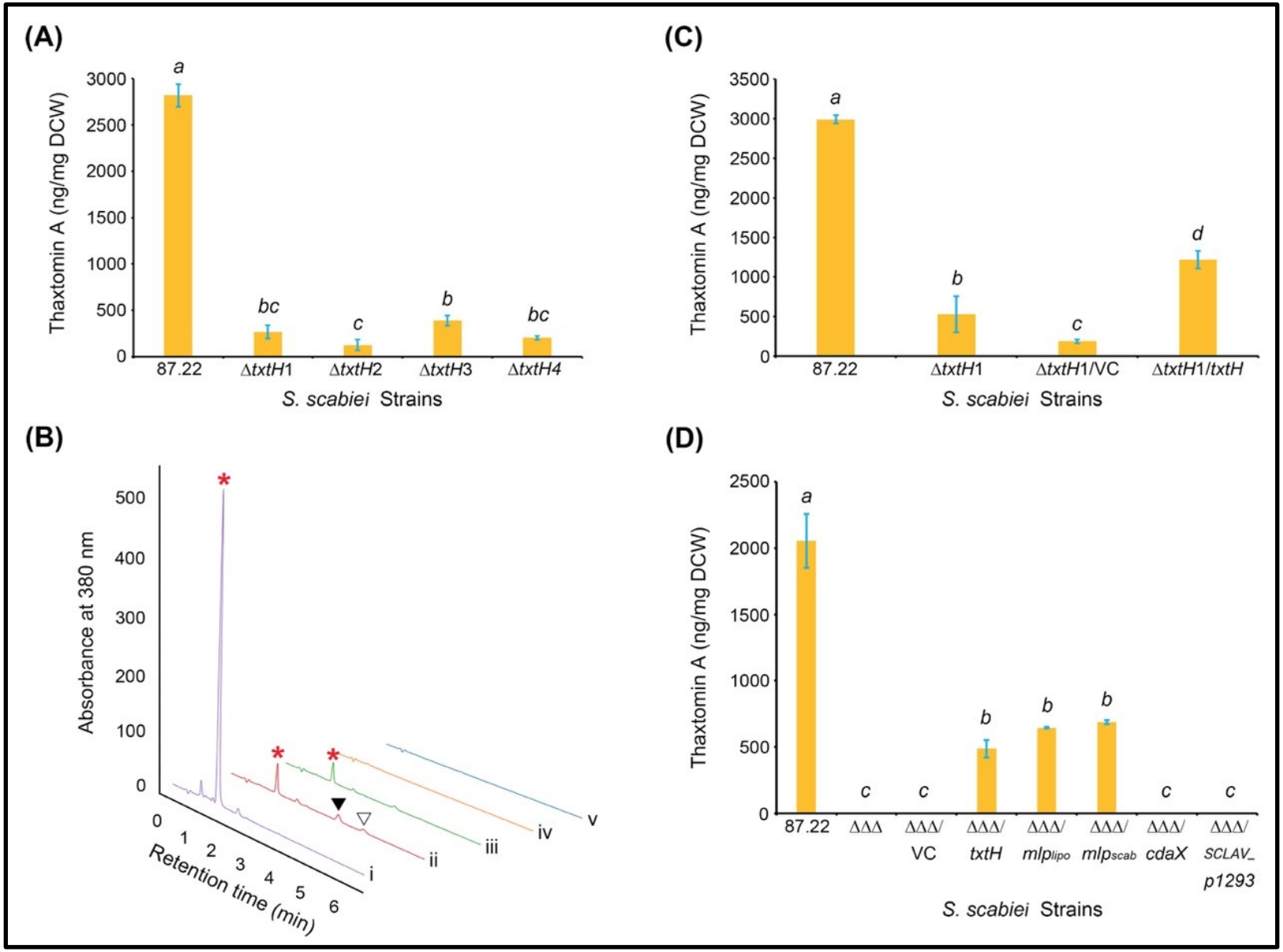

(A)
Thaxtomin A (ng/mg DCW)
3000
2500
2000
1500
1000
500
0
a
bc
c
b
bc
87.22
ΔtxtH1
ΔtxtH2
ΔtxtH3
ΔtxtH4
S. scabiei Strains
(C)
Thaxtomin A (ng/mg DCW)
3500
3000
2500
2000
1500
1000
500
0
a
b
c
d
87.22
ΔtxtH1
ΔtxtH1/VC
ΔtxtH1/txtH
S. scabiei Strains
(B)
Absorbance at 380 nm
500
400
300
200
100
0
0
1
2
3
4
5
6
Retention time (min)
i
ii
iii
iv
v
(D)
Thaxtomin A (ng/mg DCW)
2500
2000
1500
1000
500
0
a
c
c
b
b
b
c
c
87.22
ΔΔΔ
ΔΔΔ/ VC
ΔΔΔ/ txtH
ΔΔΔ/ mlplipo
ΔΔΔ/ mlpscab
ΔΔΔ/ cdaX
ΔΔΔ/ SCLAV_ p1293
S. scabiei Strains

Figure 2.4 Production of thaxtomin A by *Streptomyces scabiei* strains. Shown are the mean thaxtomin A production levels (ng thaxtomin A/mg dry cell weight) from triplicate cultures of each strain, with error bars representing the standard deviation from the mean. Means with different letters (*a*, *b*, *c*, *d*) were determined to be significantly different ($P \leqslant 0.05$). (A) Thaxtomin A production levels in *S. scabiei* 87.22 and in the Δ*txtH* mutant isolates 1–4. (B) HPLC chromatograms of culture extracts from *S. scabiei* 87.22 (i), Δ*txtH*1 (ii), $\Delta mlp_{lipo}/\Delta txtH$ (iii), $\Delta txtH/\Delta mlp_{scab}$ (iv) and $\Delta mlp_{lipo}/\Delta txtH/\Delta mlp_{scab}$ (v). The peak corresponding to thaxtomin A in each chromatogram is indicated with the red asterisks, and the peaks corresponding to the thaxtomin B and thaxtomin D intermediates are indicated with ▼ and ∇, respectively. (C) Thaxtomin A production levels in the Δ*txtH*1 mutant isolate following complementation with the *txtH* gene. (D) Thaxtomin A production levels in the $\Delta mlp_{lipo}/\Delta txtH/\Delta mlp_{scab}$ (ΔΔΔ) mutant following complementation with the *txtH*, mlp_{lipo}, mlp_{scab}, *cdaX* and *SCLAV_p1293* genes. VC, vector control.

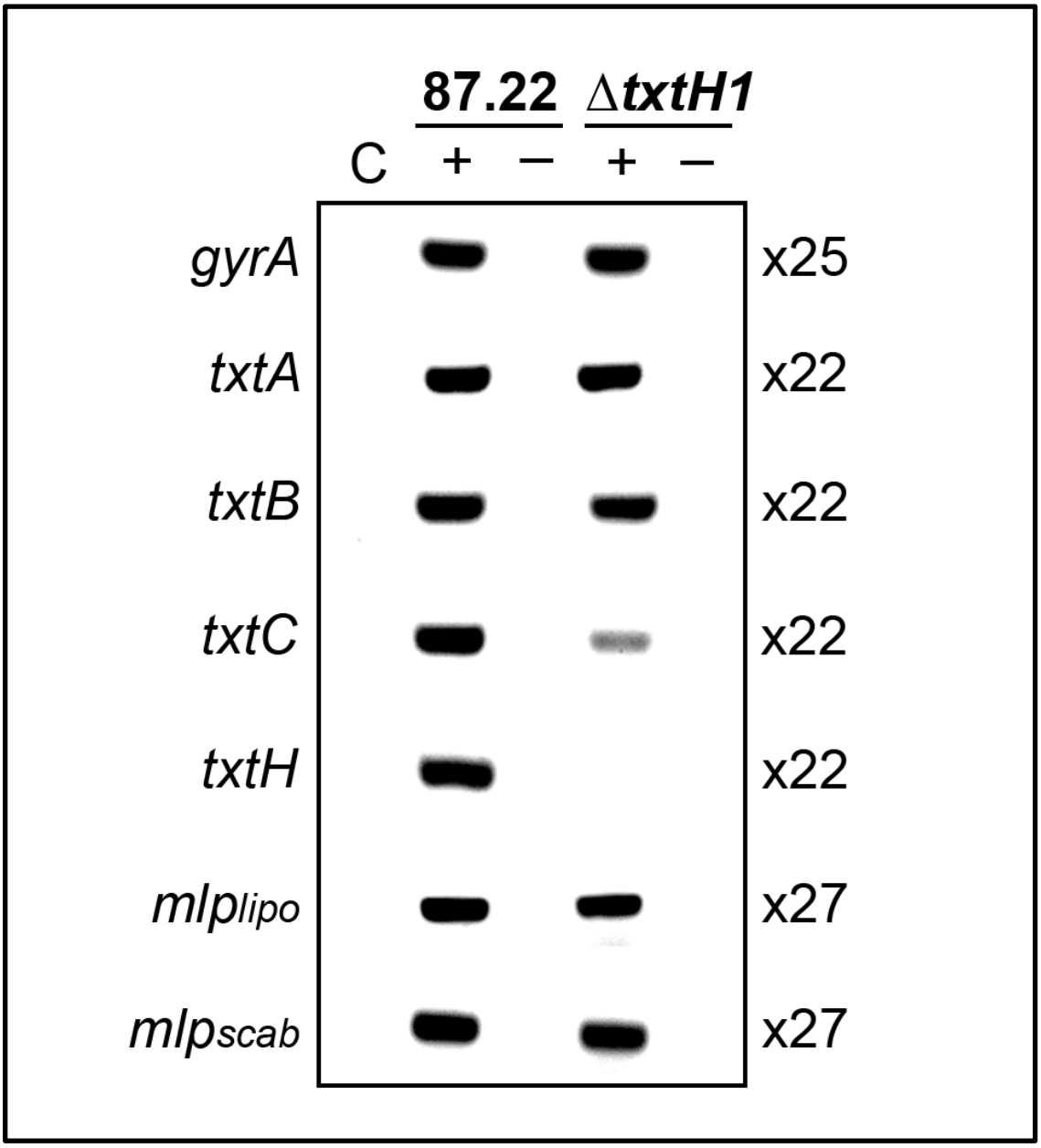


Figure 2.5 RT-PCR analysis of gene expression in *Streptomyces scabiei* 87.22 and the Δ*txtH*1 mutant. Reverse transcription reactions containing (+) or lacking (–) reverse transcriptase enzyme were used as a template for the PCR, while control (C) reactions contained water in place of the cDNA template. The number of cycles used for each set of gene-specific primers is indicated. The *gyrA* gene encoding the DNA gyrase subunit A was included as a loading control.

It has been reported that MLPs from different pathways can functionally complement each other (Lautru *et al.*, 2007; Wolpert *et al.,* 2007; Zhang *et al*., 2010; Boll *et al*., 2011; Schomer and Thomas, 2017; Mori *et al.,* 2018). In organisms where multiple MLPs are encoded in a single genome, the deletion of a single MLP often does not abolish the production of the cognate metabolite, but instead metabolite production is eliminated only when all copies of MLP-encoding genes are removed from the host genome (Lautru *et al.*, 2007; Wolpert *et al.,* 2007). As the *S. scabiei* genome harbours two additional MLP-encoding genes, mlp_{lipo} and mlp_{scab}, it is possible that either or both MLPs might be able to partially compensate for the loss of *txtH* in the Δ*txtH* mutant. When we deleted mlp_{lipo} from the wild-type *S. scabiei* chromosome (Supplementary Figure 2.1), thaxtomin A production was similar in the Δmlp_{lipo} mutant as compared to the wild-type strain (data not shown). Deletion of both *txtH* and mlp_{lipo} resulted in thaxtomin A production levels that are similar or slightly reduced as compared to the Δ*txtH* single mutant, whereas deletion of *txtH* and mlp_{scab} abolished thaxtomin A production completely, and similar results were observed when all three MLP genes were deleted (Figure 2.4B). Both the mlp_{lipo} and mlp_{scab} genes were shown to be expressed in wild-type *S. scabiei* and in the Δ*txtH*1 mutant under thaxtomin-inducing conditions (Figure 2.5), suggesting that the lack of thaxtomin A production in the $\Delta txtH/\Delta mlp_{scab}$ mutant was not due to a lack of transcription of the mlp_{lipo} gene. Interestingly, both MLP_{lipo} and MLP_{scab} were able to promote the soluble expression of the TxtA and TxtB A-domains in *E. coli,* though the solubility-promoting activity of MLP_{scab} was less efficient for the HIS_6-$TxtB^A$ protein (Figure 2.3C). This suggests that

despite the inability of the Δ*txtH*/Δ*mlp*$_{scab}$ mutant to produce detectable levels of thaxtomin A, both MLP$_{lipo}$ and MLP$_{scab}$ have the ability to functionally replace TxtH in its interaction with the thaxtomin NRPS A-domains. Further investigations will be required to determine the reason for the lack of detectable thaxtomin A production in the Δ*txtH*/Δ*mlp*$_{scab}$ mutant.

2.4.4 Engineered expression of MLPs in wild-type *S. scabiei* and in the MLP triple mutant

To further explore the ability of MLPs from different biosynthetic pathways to promote thaxtomin A production in the absence of *txtH,* we constructed several plasmids that overexpress different MLP-encoding genes using the *ermE*p* promoter and then introduced them into the Δ*mlp*$_{lipo}$/Δ*txtH*/Δ*mlp*$_{scab}$ mutant. As shown in Figure 2.4D, overexpression of *mlp*$_{lipo}$ and *mlp*$_{scab}$ from *S. scabiei* restored thaxtomin A production in the triple mutant to levels similar to that observed when *txtH* was overexpressed, confirming that both MLPs can functionally replace *txtH* in the thaxtomin biosynthetic pathway. We note that overexpression of *txtH, mlp*$_{lipo}$ and *mlp*$_{scab}$ also led to accumulation of the thaxtomin B and D biosynthetic intermediates (Supplementary Figure 2.5), confirming that there are some polar effects of the Δ*txtH* mutation on expression of the downstream *txtC* gene. In contrast, overexpression of the MLP-encoding genes *cdaX* from *S. coelicolor* and *SCLAV_p1293* from *S. clavuligerus* did not restore thaxtomin metabolite production in the *S. scabiei* triple MLP mutant (Figure 2.4D), suggesting that neither MLP can exhibit functional cross-talk with TxtH. Both CdaX and SCLAV_p1293 localize in different phylogenetic clades from TxtH (Figure 2.2B), though CdaX is predicted to be

closely related to MLP_{lipo} and MLP_{scab}, both of which can exhibit cross-talk with TxtH (Figure 2.4D). Interestingly, a recent study by Schomer and Thomas (2017) also showed that while some non-cognate MLPs are able to functionally replace the YbdZ MLP in the *E. coli* enterobactin biosynthetic pathway, others cannot, and no apparent correlation between MLP functionality and sequence similarity could be identified (Schomer and Thomas, 2017).

Previously, it was reported that the overexpression of cognate and non-cognate MLPs *in vivo* increases vancomycin production in the high producing strain *Amycolatopsis orientalis* KFCC10990P (Lee et al., 2016). We investigated whether overexpression of *txtH,* mlp_{lipo}, mlp_{scab}, *cdaX* and *SCLAV_p1293* in *S. scabiei* 87.22 enhances thaxtomin A production in this strain; however, none of the overexpression strains produced significantly higher levels of thaxtomin A compared to the control strain (data not shown). Other studies have shown that an A-domain requires a 1:1 molar ratio with its MLP partner for the maximum enzyme activity, and increasing the amount of MLPs beyond this optimal ratio did not stimulate the adenylating activity beyond a point (Boll *et al.,* 2011; Davidsen *et al.*, 2013). Our results suggest that a similar situation may exist with TxtH and its cognate NRPS, though further investigations into this are needed.

2.4.5 Plant pathogenic phenotype of the *S. scabiei* MLP mutants

We conducted a potato tuber slice assay in order to compare the virulence phenotype of the different *S. scabiei* MLP mutant strains. As expected, *S. scabiei* 87.22 readily colonized the surface of the potato tuber tissue and caused significant necrosis of

the tissue after 10 days post-inoculation (Figure 2.6). The Δ*txtH* and Δmlp_{lipo}/Δ*txtH* mutants also colonized the tissue and induced tissue necrosis, though both strains were less efficient at doing so than the wild-type strain. In contrast, there was very little visible growth of the Δmlp_{scab}/Δ*txtH* and Δmlp_{lipo}/Δ*txtH*/Δmlp_{scab} mutant strains on the tuber tissue, and both strains caused very little necrosis of the tissue (Figure 2.6). Given that a positive correlation has been noted between the production of thaxtomin A and the virulence of scab-causing *Streptomyces* spp. (King *et al.*, 1991; Healy *et al.*, 2000), the observed virulence phenotype of the different MLP mutant strains is consistent with the corresponding thaxtomin A production profiles observed in liquid culture (Figure 2.4B). It remains to be determined whether production of the putative lipopeptide metabolite and the scabichelin siderophore are also affected in the MLP mutant strains and whether these metabolites also contribute to the pathogenicity of *S. scabiei*. As siderophore production is known to contribute to the virulence phenotype of plant pathogenic bacteria (Franza *et al.*, 2005; Taguchi *et al.*, 2010), it will be interesting to further investigate the role of scabichelin in *S. scabiei* plant pathogenicity.

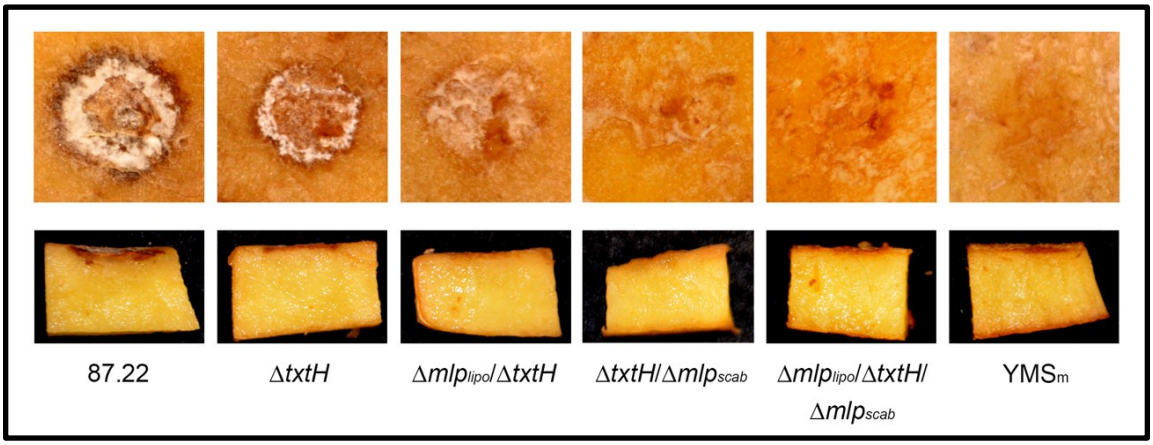


Figure 2.6 Potato tuber slice assay for assessing the virulence phenotype of *Streptomyces scabiei* strains. Tuber slices were inoculated with wild-type and mutant *S. scabiei* strains and were incubated for 10 days. Uninoculated medium (YMS_m) was included as a negative control. The bioassay was performed twice in total and representative results are shown.

2.5 Conclusion

This study demonstrated the importance of TxtH in the biosynthesis of thaxtomin A in *S. scabiei*. Particularly, TxtH is required for promoting the soluble expression of both A-domains from the thaxtomin NRPS in *E. coli*, suggesting that it performs a chaperone-like role to enable the proper folding of the NRPS in *S. scabiei*. Amino acid residues that contribute to the solubility-promoting activity of TxtH have been revealed in this study, and future structural investigations will provide important insights into the role of these residues in mediating interactions between TxtH and the thaxtomin NRPSs. We also showed that MLP_{lipo} from the putative lipopeptide biosynthetic pathway and MLP_{scab} from the scabichelin biosynthetic pathway can functionally replace TxtH in the thaxtomin biosynthetic pathway, whereas two MLPs from other *Streptomyces* spp. cannot. Further investigations are required to better understand the mechanisms behind MLP cross-talk and why certain MLPs from different pathways can functionally complement each other while

others are unable to do so. Finally, our study confirmed that TxtH is important for the plant pathogenic phenotype of *S. scabiei*.

2.6 Acknowledgement

We thank Prof. Lutz Heide (University of Tubingen) for the gift of the BL21(DE3)/Δ*ybdZ* strain, and Brandon Piercey for assistance with figure preparation and statistical analyses. This work was supported by a Natural Sciences Engineering Research Council of Canada Grant No. 386696-2010 to D.R.D.B. and by a Natural Sciences Engineering Research Council of Canada Grant No. 386417-2010 to K.T. Y.L. was supported in part by the President's Doctoral Student Investment Fund at Memorial University, and J.L was supported by a China Scholarship Council Grant No. 201603250060.

2.7 References

Baltz, R.H. (2011) Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J Ind Microbiol Biotechnol*. 38, 1747-1760. doi: 10.1007/s10295-011-1022-8.

Barry, S.M., Kers, J.A., Johnson, E.G., Song, L.J., Aston, P.R., Patel, B., Krasnoff, S.B., Crane, B.R., Gibson, D.M., Loria, R. and Challis, G.L. (2012) Cytochrome P450-catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. *Nat Chem Biol*. 8, 814-816. doi: 10.1038/nchembio.1048.

Bignell, D.R.D., Fyans, J.K. and Cheng, Z. (2014) Phytotoxins produced by plant pathogenic *Streptomyces* species. *J Appl Microbiol*. 116, 223-235. doi: 10.1111/jam.12369.

Bignell, D.R.D., Huguet-Tapia, J.C., Joshi, M.V., Pettis, G.S. and Loria, R. (2010a) What does it take to be a plant pathogen: genomic insights from *Streptomyces* species. *Antonie Van Leeuwenhoek*. 98, 179-194. doi: 10.1007/s10482-010-9429-1.

Bignell, D.R.D., Seipke, R.F., Huguet-Tapia, J.C., Chambers, A.H., Parry, R.J. and Loria, R. (2010b) *Streptomyces scabies* 87-22 contains a coronafacic acid-like biosynthetic cluster that contributes to plant-microbe interactions. *Mol Plant-Microbe Interact*. 23, 161-175. doi:10.1094/MPMI-23-2-0161.

Bischoff, V., Cookson, S.J., Wu, S. and Scheible, W.R. (2009) Thaxtomin A affects CESA-complex density, expression of cell wall genes, cell wall composition, and causes ectopic lignification in *Arabidopsis thaliana* seedlings. *J Exp Bot*. 60, 955-965. doi: 10.1093/jxb/ern344.

Boll, B., Taubitz, T. and Heide, L. (2011) The role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. *J Biol Chem*. 286, 36281-36290. doi: 10.1074/jbc.M111.288092.

Buchko, G.W., Kim, C.Y., Terwilliger, T.C. and Myler, P.J. (2010) Solution structure of Rv2377c-founding member of the MbtH-like protein family. *Tuberculosis*. 90, 245-251. doi: 10.1016/j.tube.2010.04.002.

Davidsen, J.M., Bartley, D.M., and Townsend, C.A. (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. doi: 10.1021/ja307710d.

Dees, M.W. and Wanner, L.A. (2012) In search of better management of potato common scab. *Potato Res*. 55, 249-268. doi: 10.1007/s11540-012-9206-9.

Drake, E.J., Cao, J., Qu, J., Shah, M.B., Straubinger, R.M. and Gulick, A.M. (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. doi: 10.1074/jbc.M611833200.

Esquilín-Lebrón, K.J., Boynton, T.O., Shimkets, L.J. and Thomas, M.G. (2018) An orphan MbtH-like protein interacts with multiple nonribosomal peptide synthetases in *Myxococcus xanthus* DK1622. *J Bacteriol*. 200, e00346-18. doi: 10.1128/JB.00346-18.

Errakhi, R., Dauphin, A., Meimoun, P., Lehner, A., Reboutier, D., Vatsa, P., Briand, J., Madiona, K., Rona, J.P., Barakate, M., Wendehenne, D., Beaulieu, C. and Bouteau, F. (2008) An early Ca^{2+} influx is a prerequisite to thaxtomin A-induced cell death in *Arabidopsis thaliana* cells. *J Exp Bot*. 59, 4259-4270. doi: 10.1093/jxb/ern267.

Felnagle, E.A., Barkei, J.J., Park, H., Podevels, A.M., McMahon, M.D., Drott, D.W. and Thomas, M.G. (2010) MbtH-Like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry*. 49, 8815-8817. doi: 10.1021/bi1012854.

Fernandez-Martinez, L.T. and Bibb, M.J. (2014) Use of the meganuclease *I-Sce*I of *Saccharomyces cerevisiae* to select for gene deletions in actinomycetes. *Sci Rep*. 4, 7100. doi: 10.1038/srep07100.

Finn R.D., Coggill P., Eberhardt R.Y., Eddy S.R., Mistry J., Mitchell A.L., Potter S.C., Punta M., Qureshi M., Saangrador-Vegas A., Salazar G.A., Tate J., Bateman A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 44: D279-D285. doi: 10.1093/nar/gkv1344.

Franza, T., Mahe, B. and Expert, D. (2005) *Erwinia chrysanthemi* requires a second iron transport route dependent of the siderophore achromobactin for extracellular growth and plant infection. *Mol Microbiol*. 55, 261-275. doi: 10.1111/j.1365-2958.2004.04383.x.

Fyans, J.K., Altowairish, M.S., Li, Y.T. and Bignell, D.R.D. (2015) Characterization of the coronatine-like phytotoxins produced by the common scab pathogen *Streptomyces scabies*. *Mol Plant-Microbe Interact*. 28, 443-454. doi: 10.1094/MPMI-09-14-0255-R.

Fyans, J.K., Bown, L. and Bignell, D.R.D. (2016) Isolation and characterization of plant-pathogenic *Streptomyces* species associated with common scab-infected potato tubers in Newfoundland. *Phytopathology*. 106, 123-131. doi: 10.1094/PHYTO-05-15-0125-R.

Garrity, G.M., Lilburn, T.G., Cole, J.R., Harrison, S.H., Euzéby, J. and Tindall, B.J. (2007) Part 10: The Bacteria: Phylum "*Actinobacteria*": Class *Actinobacteria*, 399–541. *The Taxonomic Outline of the Bacteria and Archaea*. Release 7.7. Michigan State University Board of Trustees, East Lansing, MI. doi: 10.1601/nm.5712.

Gust, B., Challis, G.L., Fowler, K., Kieser, T. and Chater, K.F. (2003a) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis

of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci USA*. 100, 1541-1546. doi: 10.1073pnas.0337542100.

Gust, B., O'Rourke, S., Bird, N., Kieser, T. and Chater, K.F. (2003b) Recombineering in *Streptomyces coelicolor*. The John Innes Foundation: Norwich, UK.

Healy, F.G., Krasnoff, S.B., Wach, M., Gibson, D.M. and Loria, R. (2002) Involvement of a cytochrome P450 monooxygenase in thaxtomin a biosynthesis by *Streptomyces acidiscabies*. J *Bacteriol*. 184, 2019-2029. doi: 10.1128/JB.184.7.2019–2029.2002.

Healy, F.G., Wach, M., Krasnoff, S.B., Gibson, D.M. and Loria, R. (2000) The *txtAB* genes of the plant pathogen *Streptomyces acidiscabies* encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity. *Mol Microbiol*. 38, 794-804. doi: 10.1046/j.1365-2958.2000.02170.x.

Herbst, D.A., Boll, B., Zocher, G., Stehle, T. and 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. doi: 10.1074/jbc.M112.420182.

Ikeda, H., Kotaki, H. and Omura, S. (1987) Genetic studies of avermectin biosynthesis in *Streptomyces avermitilis*. *J Bacteriol*. 169, 5615-5621. doi: 10.1128/jb.169.12.5615-5621.1987.

Imker, H.J., Krahn, D., Clerc, J., Kaiser, M. and Walsh, C.T. (2010) *N*-acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF. *Chem Biol*. 17, 1077-1083. doi: 10.1016/j.chembiol.2010.08.007.

Kieser, T., Bibb, M., Buttner, M., Chater, K. and Hopwood, D. (2000) Practical *Streptomyces* genetics. Norwich, UK: The John Innes Foundation.

Kumar S., Stecher G., Tamura K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 33:1870-4. doi: 10.1093/molbev/msw054.

Johnson, E.G., Joshi, M.V., Gibson, D.M. and Loria, R. (2007) Cello-oligosaccharides released from host plants induce pathogenicity in scab-causing *Streptomyces* species. *Physiol Mol Plant Pathol*. 71, 18-25. doi: 10.1016/j.pmpp.2007.09.003.

Johnson, E.G., Krasnoff, S.B., Bignell, D.R.D., Chung, W.C., Tao, T., Parry, R.J., Loria, R. and Gibson, D.M. (2009) 4-Nitrotryptophan is a substrate for the non-ribosomal peptide synthetase TxtB in the thaxtomin A biosynthetic pathway. *Mol Microbiol*. 73, 409-418. doi: 10.1111/j.1365-2958.2009.06780.x.

Joshi, M., Rong, X., Moll, S., Kers, J., Franco, C. and Loria, R. (2007) *Streptomyces turgidiscabies* secretes a novel virulence protein, Nec1, which facilitates infection. *Mol Plant-Microbe Interact*. 20, 599-608. doi: 10.1094/MPMI-20-6-0599.

King, R.R. and Calhoun, L.A. (2009) The thaxtomin phytotoxins: sources, synthesis, biosynthesis, biotransformation and biological activity. *Phytochemistry*. 70, 833-841. doi: 10.1016/j.phytochem.2009.04.013.

King, R.R., Lawrence, C.H. and Clark, M.C. (1991) Correlation of phytotoxin production with pathogenicity of *Streptomyces scabies* isolates from scab infected potato-tubers. *Am Potato J*. 68, 675-680. doi: 10.1007/BF02853743.

King, R.R., Lawrence, C.H., Clark, M.C. and Calhoun, L.A. (1989) Isolation and characterization of phytotoxins associated with *Streptomyces scabies*. *J Chem Soc Chem Commun*. 13, 849-850. doi: 10.1039/c39890000849.

Kodani, S., Bicz, J., Song, L.J., Deeth, R.J., Ohnishi-Kameyama, M., Yoshida, M., Ochi, K. and Challis, G.L. (2013) Structure and biosynthesis of scabichelin, a novel tris-hydroxamate siderophore produced by the plant pathogen *Streptomyces scabies* 87.22. *Org Biomol Chem*. 11, 4686-4694. doi: 10.1039/c3ob40536b.

Lautru, S., Oves-Costales, D., Pernodet, J.L. and Challis, G.L. (2007) MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. *Microbiology.* 153, 1405-1412. doi: 10.1099/mic.0.2006/003145-0.

Lee, K.S., Lee, B.M., Ryu, J.H., Kim, D.H., Kim, Y.H. and Lim, S.K. (2016) Increased vancomycin production by overexpression of MbtH-like protein in *Amycolatopsis orientalis* KFCC10990P. *Lett Appl Microbiol*. 63, 222-228. doi: 10.1111/lam.12617.

Loria, R., Bignell, D.R., Moll, S., Huguet-Tapia, J.C., Joshi, M.V., Johnson, E.G., Seipke, R.F. and Gibson, D.M. (2008) Thaxtomin biosynthesis: the path to plant pathogenicity in the genus *Streptomyces*. *Antonie Van Leeuwenhoek*. 94, 3-10. doi: 10.1007/s10482-008-9240-4.

Loria, R., Bukhalid, R.A., Creath, R.A., Leiner, R.H., Olivier, M. and Steffens, J.C. (1995) Differential production of thaxtomins by pathogenic *Streptomyces* species *in vitro*. *Phytopathology*. 85, 537-541. doi: 10.1094/Phyto-85-537.

Loria, R., Kers, J. and Joshi, M. (2006) Evolution of plant pathogenicity in *Streptomyces*. *Annu Rev Phytopathol*. 44, 469-487. doi: 10.1146/annurev.phyto.44.032905.091147.

McMahon, M.D., Rush, J.S. and Thomas, M.G. (2012) Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *J Bacteriol*. 194, 2809-2818. doi: 10.1128/JB.00088-12.

Mori, S., Green, K.D., Choi, R., Buchko, G.W., Fried, M.G. and Garneau-Tsodikova, S. (2018) Using MbtH-like proteins to alter the substrate profile of a nonribosomal peptide adenylation enzyme. *Chembiochem*. 19, 1-10. doi: 10.1002/cbic.201800240.

Sambrook, J. and Russell, D.W. (2001) Molecular cloning: a laboratory manual (3-volume set). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Scheible, W.R., Fry, B., Kochevenko, A., Schindelasch, D., Zimmerli, L., Somerville, S., Loria, R. and Somerville, C.R. (2003) An Arabidopsis mutant resistant to thaxtomin A, a cellulose synthesis inhibitor from *Streptomyces* species. *Plant Cell*. 15, 1781-1794. doi: 10.1105/tpc.013342.

Schomer, R.A. and Thomas, M.G. (2017) Characterization of the functional variance in MbtH-like protein interactions with a nonribosomal peptide synthetase. *Biochemistry*. 56, 5380-5390. doi: 10.1021/acs.biochem.7b00517.

Süssmuth, R.D. and Mainz, A. (2017) Nonribosomal peptide synthesis – Principles and prospects. *Angew Chem Int Ed.* 56, 3370-3821. doi: 10.1002/anie.201609079.

Taguchi, F., Suzuki, T., Inagaki, Y., Toyoda, K., Shiraishi, T. and Ichinose, Y. (2010) The siderophore pyoverdine of *Pseudomonas syringae* pv. tabaci 6605 is an intrinsic virulence factor in host tobacco infection. *J Bacteriol*. 192, 117-126. doi: 10.1128/jb.00689-09.

Tegg, R.S., Melian, L., Wilson, C.R. and Shabala, S. (2005) Plant cell growth and ion flux responses to the streptomycete phytotoxin thaxtomin A: Calcium and hydrogen flux patterns revealed by the non-invasive MIFE technique. *Plant Cell Physiol*. 46, 638-648. doi: 10.1093/pcp/pci069.

Whelan, S. and Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 18, 691-699. doi: 10.1093/oxfordjournals.molbev.a003851.

Wolpert, M., Gust, B., Kammerer, B. and Heide, L. (2007) Effects of deletions of *mbtH*-like genes on clorobiocin biosynthesis in *Streptomyces coelicolor*. *Microbiology*. 153, 1413-1423. doi: 10.1099/mic.0.2006/002998-0.

Yaxley, A.M. (2009) Study of the complete genome sequence of *Streptomyces scabies* (or *scabiei*) 87.22. (Doctoral dissertation, University of Warwick).

Zhang, W.J., Heemstra, J.R., Walsh, C.T. and Imker, H.J. (2010) Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins. *Biochemistry.* 49, 9946-9947. doi: 10.1021/bi101539b.

Zolova, O.E. and Garneau-Tsodikova, S. (2012) Importance of the MbtH-like protein TioT for production and activation of the thiocoraline adenylation domain of TioK. *MedChemComm*. 3, 950-955. doi: 10.1039/c2md20131c.

2.8 Supplementary Information

Supplementary Table 2.1 Oligonucleotide primers used in this study.

Primer	Sequence (5′ - 3′) †	Use
PL3	GCGC**TCTAGA**GCTGAACACCAACGGCAA	Forward primer for construction of pGEM-T EASY/5′ mlp_{lipo}
PL4	GCGC**GGATCC**GTTCTCGAAGGGGTTGGTCAT	Reverse primer for construction of pGEM-T EASY/5′ mlp_{lipo}
PL5	GCGC**GGATCC**GTCCGGGCCATGGGCGAGTGA	Forward primer for construction of pGEM-T EASY/3′ mlp_{lipo}
PL6	GCGC**GAATTC**CCATGCCACTGAGGGACT	Reverse primer for construction of pGEM-T EASY/3′ mlp_{lipo}
PL35	GCGC**CATATG**CCCTCACCCTTCGACGAC	Forward primer for construction of the overexpression plasmid pET28b/HIS$_6$-*txtH* and verification of Δ*txtH* deletion in *S. scabiei*
PL36	GCGC**GAATTC**TCATTCACGGACGGACGCCG	Reverse primer for construction of the overexpression plasmid pET28b/HIS$_6$-*txtH* and verification of Δ*txtH* deletion in *S. scabiei*
PL37	GCGC**GAATTC**GATGTCGCACCTGACCGGTGAA	Forward primer for construction of the overexpression plasmid pACYCDuet-1/ HIS$_6$-$txtA^A$
PL38	GCGC**AAGCTT**CCAGTAGCTTTCGCAGTCAC	Reverse primer for construction of the overexpression plasmid pACYCDuet-1/ HIS$_6$-$txtA^A$
PL40	GCGC**GAATTC**GATGTCCATGCTGCCGCCGGG	Forward primer for construction of the overexpression plasmid pACYCDuet-1/ HIS$_6$-$txtB^A$
PL41	GCGC**AAGCTT**ACGGATGCTGTCGACCGTG	Reverse primer for construction of the overexpression plasmid pACYCDuet-1/ HIS$_6$-$txtB^A$
PL62	GACCCATCGACCCACCGA	Forward primer for verification of *S. scabiei* strain 87.22/Δmlp_{lipo}_int
PL63	AGGTTGTGGCCACGGAAC	Reverse primer for verification of *S. scabiei* strain 87.22/Δmlp_{lipo}_int
PL67	GCGC**GGATCC**CCACCGAAAGCACCGTCAAT	Forward primer for construction of the overexpression plasmid pRLDB50-1a/*txtH*

PL68	GCGC**TCTAGA**GG CCCCGCTCGATGT TATTG	Reverse primer for construction of the overexpression plasmid pRLDB50-1a/*txtH*
PL69	GCGC**GGATCC**AT CGACCCGTCGACC CATCG	Forward primer for construction of the overexpression plasmid pRLDB50-1a/mlp_{lipo}
PL70	GCGC**TCTAGA**TCC CTTTCGTCCGCTG CGTC	Reverse primer for construction of the overexpression plasmid pRLDB50-1a/ mlp_{lipo}
PL71	GCGC**GGATCC**TG GAGGGGCTCGTCG CACGG	Forward primer for construction of the overexpression plasmid pRLDB50-1a/mlp_{scab} and verification of Δmlp_{scab} deletion in *S. scabiei*
PL72	GCGC**TCTAGA**CCC CTTGCGTGTGCCC GTGT	Reverse primer for construction of the overexpression plasmid pRLDB50-1a/mlp_{scab} and verification of Δmlp_{scab} deletion in *S. scabiei*
PL73	GCTACTGGGAGTT CGTCACG	Forward primer for verification of Δmlp_{lipo} deletion in *S. scabiei*
PL74	GCAGTTCCGCCGC CACAT	Reverse primer for verification of Δmlp_{lipo} deletion in *S. scabiei*
PL75	GTATCTCCTGCTG CTGTCCG	Forward primer for verification of Δ*txtH* deletion in *S. scabiei*
PL76	TCCAGCACCGCCC AAGCGCT	Reverse primer for verification of Δ*txtH* deletion in *S. scabiei*
PL82	GCGC**GGATCC**TG CTGAGTACCGAGA GCCTG	Forward primer for construction of the overexpression plasmid pRLDB50-1a/*SCLAV_p1293*
PL83	GCGC**TCTAGA**GTC CGTCCGCTCCGGG GAAA	Reverse primer for construction of the overexpression plasmid pRLDB50-1a/*SCLAV_p1293*
PL84	GCGC**GGATCC**AG GGTCCGCGACCCG CGCAG	Forward primer for construction of the overexpression plasmid pRLDB50-1a/*cdaX*
PL85	GCGC**TCTAGA**GTC GTGGTCCGGTCAG TTGC	Reverse primer for construction of the overexpression plasmid pRLDB50-1a/*cdaX*
PL104	GCGAAATTCGGCC AGAGATAGAACT GGCCTTCCTCG	Forward primer for construction of the site-directed mutant plasmid pET28b/HIS_6-*txtH*(S23Y)
PL105	CGAGGAAGGCCA GTTCTATCTCTGG CCGAATTTCGC	Reverse primer for construction of the site-directed mutant plasmid pET28b/HIS_6-*txtH*(S23Y)
PL106	TCGGCGAAATTCG GCGCGAGTGAGA ACTGGCC	Forward primer for construction of the site-directed mutant plasmid pET28b/HIS_6-*txtH*(W25A)

PL107	GGCCAGTTCTCACTCGCGCCGAATTTCGCCGA	Reverse primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(W25A)
PL108	GCTCACGGAACGCGCCCCGGAGGGGATG	Forward primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(W35A)
PL109	CATCCCCTCCGGGGCGCGTTCCGTGAGC	Reverse primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(W35A)
PL110	GGCGAAATTCGGCCAGGCTGAGAACTGGCCTTCC	Forward primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(L24A)
PL111	GGAAGGCCAGTTCTCAGCCTGGCCGAATTTCGCC	Reverse primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(L24A)
PL112	GAACTGGCCTTCCTCGGCGCGGAGCACATGGAAC	Forward primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(N17A)
PL113	GTTCCATGTGCTCCGCGCCGAGGAAGGCCAGTTC	Reverse primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(N17A)
PL114	GGCCAGAGTGAGAACGCGCCTTCCTCGTTGCG	Forward primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(Q21A)
PL115	CGCAACGAGGAAGGCGCGTTCTCACTCTGGCC	Reverse primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(Q21A)
PL116	AATTCGGCCAGAGTGCGAACTGGCCTTCCTC	Forward primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(S23A)
PL117	GAGGAAGGCCAGTTCGCACTCTGGCCGAATT	Reverse primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(S23A)
PL133	AGACAGTCCGTCTCCGTCGT	Forward primer for verification of Cosmid 1989/Δ*txtH* with forward primer Apra For
PL134	CGCATGTCCGTCGCTTCCTTCTCGATGTACTCAAGG	Forward primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(W55A)
PL135	CCTTGAGTACATCGAGAAGGAAGCGACGGACATGCG	Reverse primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(W55A)
PL136	GGGCGCATGTCCGCCCATTCCTTCTCG	Forward primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(T56A)

PL137	CGAGAAGGAATGGGCGGACATGCGCCC	Reverse primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(T56A)
PL138	GCCGGGCGCATGGCCGTCCATTCCT	Forward primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(D57A)
PL139	AGGAATGGACGGCCATGCGCCCGGC	Reverse primer for construction of the site-directed mutant plasmid pET28b/HIS$_6$-*txtH*(D57A)
PL150	GCGC**CCATGG**GCGTGCCCTCACCCTTCGAC	Forward primer for construction of the overexpression plasmid pET28b/*txtH* with reverse primer PL36
PL153	CGACGCGAGCGACGCGAAGTGAGAGAGAGGAACGACATGATTCCGGGGATCCGTCGACC	Forward primer for amplification of [*hyg-oriT*] cassette for construction of Cosmid 57/Δmlp_{scab}
PL154	CGACCGGCCCCCCTTGCGTGTGCCCGTGTGCCCGGGTCATGTAGGCTGGAGCTGCTTC	Reverse primer for amplification of [*hyg-oriT*] cassette for construction of Cosmid 57/Δmlp_{scab}
PL155	AGCAACCCGTTCGAGGACCC	Forward primer for verification of Δmlp_{scab} deletion in *S. scabiei*
PL156	TCCATCGACTCGGCCAGGCT	Reverse primer for verification of Δmlp_{scab} deletion in *S. scabiei*
PL157	CAACGAGGAAGGCCAGTTCT	Forward primer, *txtH* RT-PCR analysis
PL158	ATGTACTCAAGGGCGCTTCC	Reverse primer, *txtH* RT-PCR analysis
PL159	GACCAACCCCTTCGAGAACC	Forward primer, mlp_{lipo} RT-PCR analysis
PL160	GTTCGACGTACTCCAGGCAG	Reverse primer, mlp_{lipo} RT-PCR analysis
PL161	TGGTCAACGACGAGAACCAG	Forward primer, mlp_{scab} RT-PCR analysis
PL162	CATGTCGGTCCAGTGGGTC	Reverse primer, mlp_{scab} RT-PCR analysis
PL163	GCGC**CATATG**ACCAACCCCTTCGAGAAC	Forward primer for construction of the overexpression plasmid pET28b/HIS$_6$-mlp_{lipo}
PL164	GCGC**GAATTC**TCACTCGCCCATGGCCCGGA	Reverse primer for construction of the overexpression plasmid pET28b/HIS$_6$- mlp_{lipo}

PL165	GCGC**CATATG**AGCAACCCGTTCGAGGAC	Forward primer for construction of the overexpression plasmid pET28b/HIS$_6$- *mlp$_{scab}$*
PL166	GCGC**GAATTC**TCAGCCGTCCATCGACTCGG	Reverse primer for construction of the overexpression plasmid pET28b/HIS$_6$- *mlp$_{scab}$*
DRB13	GAGCGACTGTCCTTCATGG	Forward primer, *txtA* RT-PCR analysis
DRB14	CGTCGTCCAGTACCACGAG	Reverse primer, *txtA* RT-PCR analysis
DRB23	GGACATCCAGACGCAGTACA	Forward primer, *gyrA* RT-PCR analysis
DRB24	CTCGGTGTTGAGCTTCTCCT	Reverse primer, *gyrA* RT-PCR analysis
DRB48	CGGCTACTTCCCGATGGAT	Forward primer, *txtB* RT-PCR analysis
DRB49	CTCGATGTCACTCCTGGTCA	Reverse primer, *txtB* RT-PCR analysis
DRB54	CTCACCTTCCACGAGACCAT	Forward primer, *txtC* RT-PCR analysis
DRB55	GCTGCAGTGCATAACTCACC	Reverse primer, *txtC* RT-PCR analysis
DRB627	TGCCGGGCCCTCTTTGCCGACTAGGAGAAATTCACCGTGATTCCGGGGATCCGTCGACC	Forward primer for amplification of [*aac(3)IV-oriT*] cassette for construction of Cosmid 1989/Δ*txtH*
DRB628	GGCGACCCGTGGCCCCGCTCGATGTTATTGGCCGGGTCATGTAGGCTGGAGCTGCTTC	Reverse primer for amplification of [*aac(3)IV-oriT*] cassette for construction of Cosmid 1989/Δ*txtH*
Apra For	TCGATGGGCAGGTACTTCTC	Reverse primer for verification of Cosmid 1989/Δ*txtH* with forward primer PL133

† Non-homologous extensions are underlined, while engineered restriction sites are indicated in bold.

Supplementary Table 2.2 Accession numbers of MLP protein sequences used for constructing the amino acid alignment and phylogenetic tree.

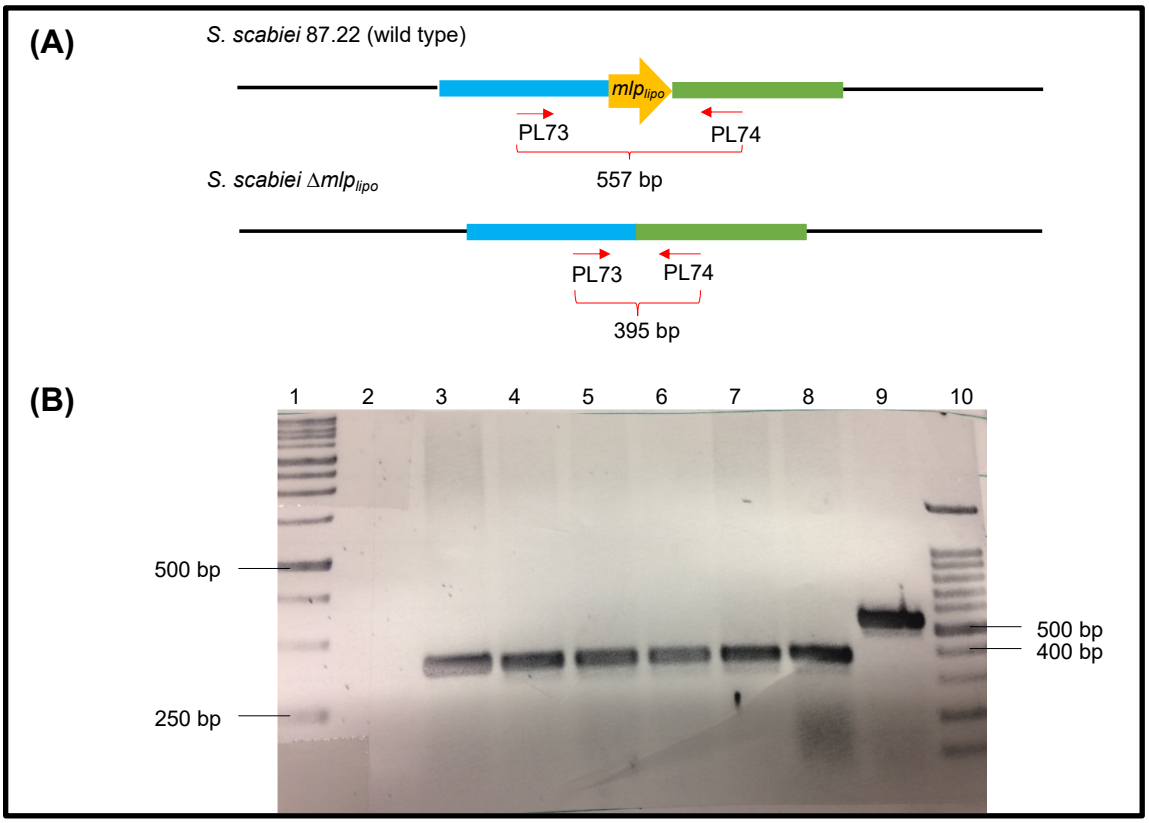
Proteins	**Accession number**
Streptomyces scabiei 87.22 TxtH	CBG70277.1
Streptomyces scabiei 87.22 MLP_{scab} (SCAB_85461)	WP_013005929.1
Streptomyces scabiei 87.22 MLP_{lipo} (SCAB_3331)	WP_012998279.1
Streptomyces coelicolor A3(2) CdaX	AAD18046.1
Streptomyces clavuligerus ATCC 27064 SCLAV_p1293	WP_003958107.1
Streptomyces turgidiscabies T45 TxtH	GAQ77365.1
Streptomyces acidiscabies a10 TxtH	GAQ51743.1
Streptomyces acidiscabies a10 AV125_RS45370	WP_010357635.1
Streptomyces europaeiscabiei 89-04 TxtH	WP_010350602.1
Streptomyces europaeiscabiei 89-04 AWZ11_RS05060	WP_046706407.1
Streptomyces viridochromogenes NRRL 3414 ACM01_RS10820	WP_048580916.1
Streptomyces zhaozhouensis CGMCC 4.7095 CRP51_RS03180	WP_097229363.1
Micromonospora sp. ML1 TioT	CAJ34376.1
Mycobacterium tuberculosis MbtH	CNH28865.1
Streptomyces vinaceus ATCC 11861 VioN	AAP92504
Pseudomonas aeruginosa PAO1 PA2412	AAG05800.1
Streptomyces coeruleorubidus NRRL 18370 PacJ	ADN26246.1
Streptomyces coelicolor A3(2) CchK	NP_624806.1
Escherichia coli ATCC 8739 YbdZ	WP_000885798.1
Myxococcus xanthus DK 1622 MXAN_3118	ABF91873.1

Supplementary Table 2.3 Pairwise comparison of amino acid identity (lower tier) and similarity (upper tier) for the MLPs included in this study.

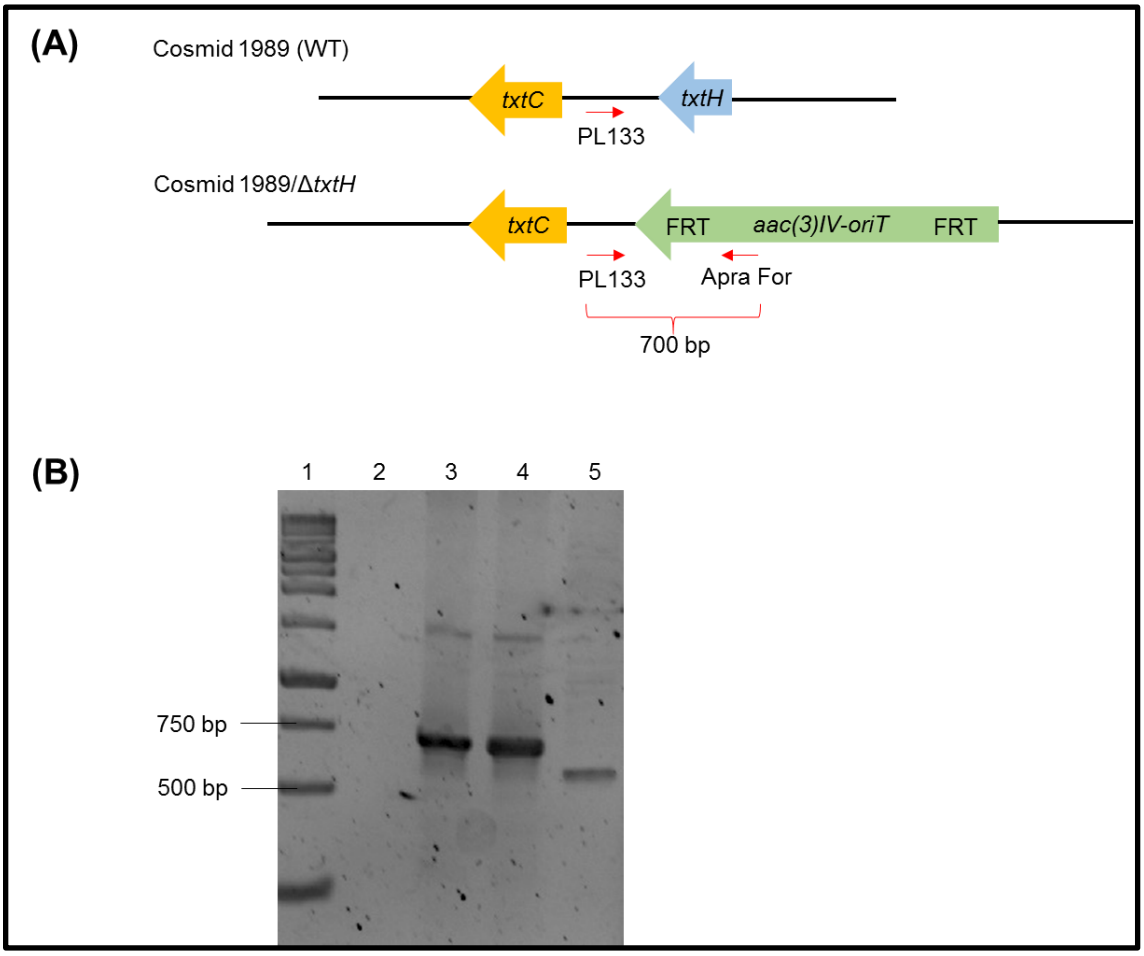
	Ssc TxtH†	MLP_{lipo}	MLP_{scab}	CdaX	SCLAV_p1293	*Stu* TxtH	*Sac* TxtH	AV125_RS45370	*Seu* TxtH	AWZ11_RS05060	ACM01_RS10820	CRP51_RS03180	TioT	MbtH	VioN	PA2412	PacJ	CchK	YbdZ	MXAN_3118
Ssc TxtH†		**75**	**73**	**73**	**61**	**85**	**100**	**68**	**100**	**80**	**98**	**79**	**76**	**77**	**62**	**63**	**68**	**70**	**60**	**59**
MLP_{lipo}	**52**		76	73	61	70	74	76	74	67	74	77	68	72	62	61	58	75	58	55
MLP_{scab}	**52**	61		82	62	69	72	77	72	71	73	74	75	80	71	66	61	72	57	58
CdaX	**57**	65	67		64	67	73	86	73	72	75	70	72	82	65	65	61	81	56	61
SCLAV_p1293	**35**	32	45	38		56	61	63	61	66	63	55	66	71	71	74	67	62	45	74
Stu TxtH	**80**	51	52	49	34		85	65	85	74	87	73	67	69	62	57	60	70	63	50
Sac TxtH	**100**	52	52	57	35	80		68	100	80	98	79	76	77	62	63	68	70	60	59
AV125_RS45370	**49**	67	61	75	39	45	49		68	67	69	69	68	77	61	63	61	75	56	59
Seu TxtH	**100**	52	52	57	35	80	100	49		80	98	79	76	77	62	63	68	70	60	59
AWZ11_RS05060	**55**	46	52	53	40	51	55	51	55		79	68	92	84	64	72	67	71	52	64
ACM01_RS10820	**97**	52	52	57	37	80	97	49	97	55		80	76	79	62	63	68	72	60	59
CRP51_RS03180	**60**	62	55	58	31	57	60	59	60	51	60		63	69	63	58	58	67	60	53
TioT	**57**	55	57	60	37	54	57	60	57	80	57	54		83	60	68	70	65	53	64
MbtH	**59**	54	58	63	38	54	59	64	59	62	59	58	69		66	64	69	75	59	67
VioN	**37**	42	46	40	51	37	37	39	37	35	37	35	37	45		73	67	60	50	65
PA2412	**48**	38	42	41	50	45	48	42	48	43	48	37	46	45	52		65	61	47	79

PacJ	**40**	33	38	35	51	37	40	35	40	40	40	39	41	41	49	49		58	69	52
CchK	**51**	61	57	67	39	51	51	69	51	56	52	58	61	63	37	42	35		58	57
YbdZ	**39**	35	39	30	20	43	39	29	39	31	39	41	37	34	25	23	29	38		46
MXAN_31 18	**34**	38	38	39	49	32	34	42	34	34	34	32	36	35	48	58	48	32	17	

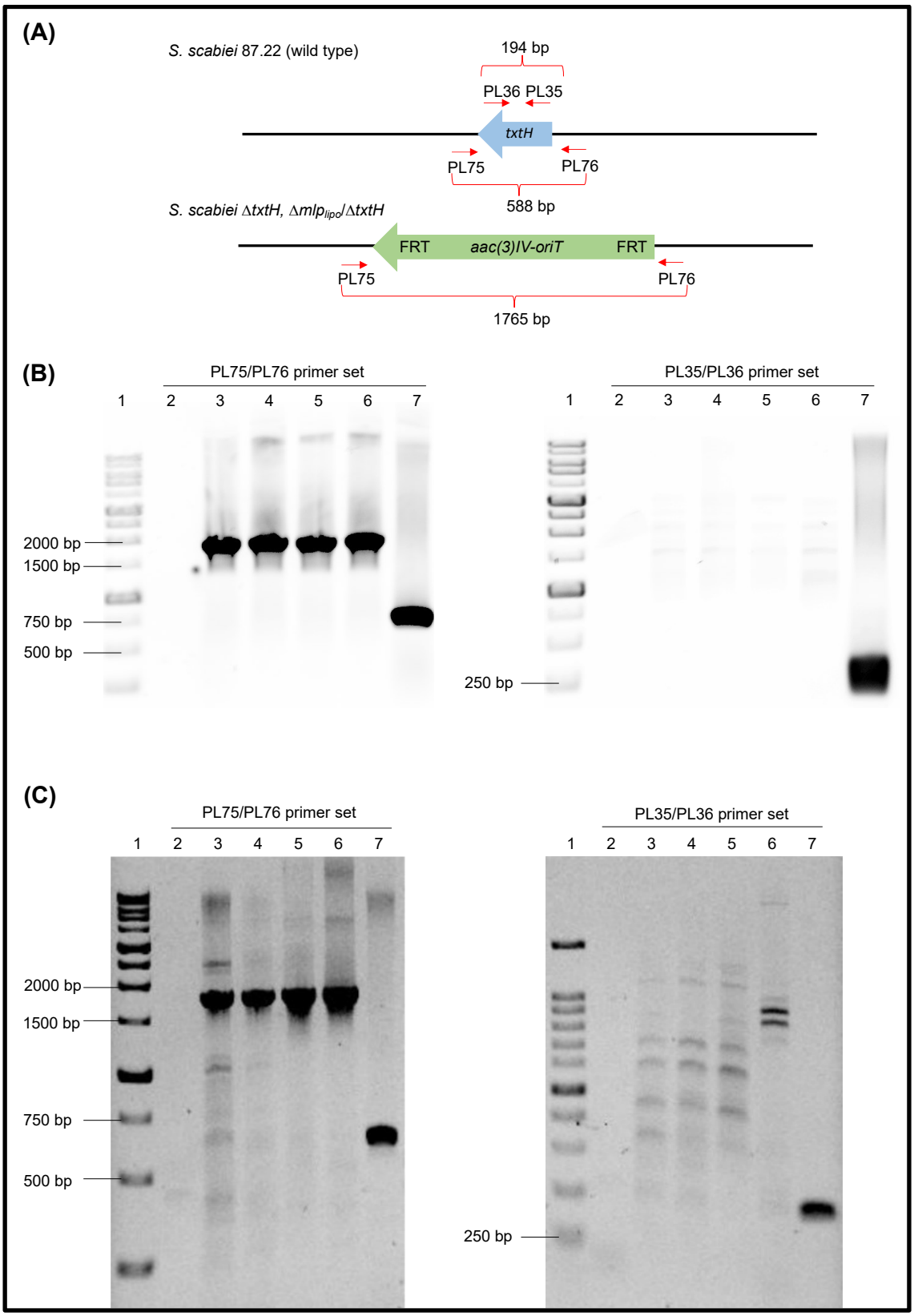
† Comparisons with TxtH from *S. scabiei* 87.22 are indicated in bold.
Ssc= S. scabiei, *Stu= S. turgidiscabies*, *Sac= S. acidiscabies* and *Seu= S. europaeiscabiei*



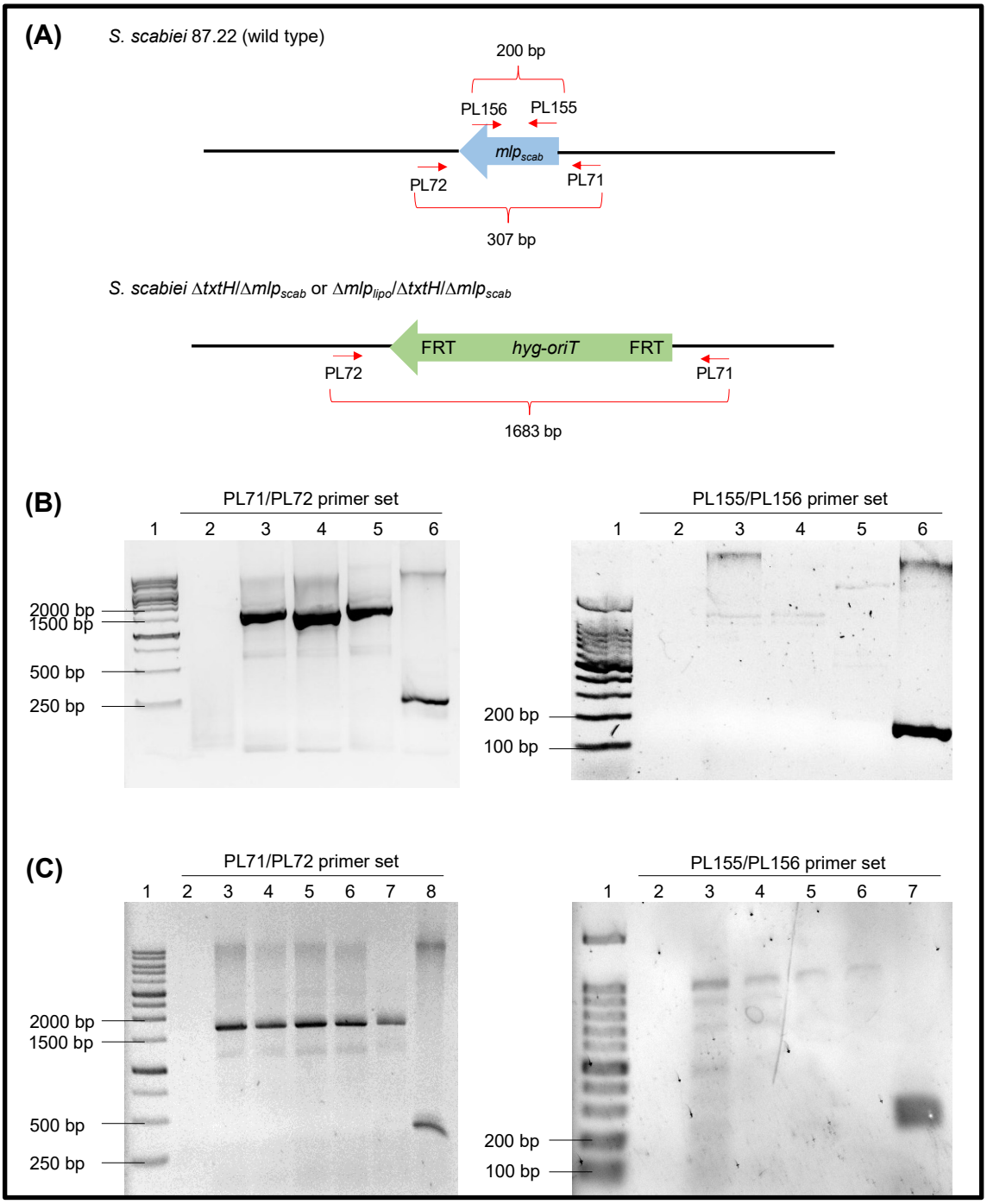
Supplementary Figure 2.1 PCR verification of the *S. scabiei* Δ*mlp*$_{lipo}$ deletion mutant. (A) Schematic diagram showing the annealing sites of the primers (indicated by the red arrows) used for the PCR verification. The expected product sizes for *S. scabiei* 87.22 (wild type) and the Δ*mlp*$_{lipo}$ mutant are indicated. The blue and green shaded areas represent the upstream and downstream regions used to construct the Δ*mlp*$_{lipo}$ deletion plasmid. (B) Agarose gel electrophoresis of the PCR products generated using genomic DNA from *S. scabiei* 87.22 (lane 9) and from the Δ*mlp*$_{lipo}$ mutant isolates 1-6 (lanes 3-8). A negative control reaction was conducted using water in place of template DNA (lane 2). The size (kb) of each product was estimated by comparison with the 1kb ladder (lane 1) and with the 100 bp ladder (lane 10).



Supplementary Figure 2.2 PCR verification of the orientation of the extended apramycin resistance cassette in the Δ*txtH* mutant cosmid. (A) Strategy used to verify the orientation of the *aac(3)IV-oriT* cassette in Cosmid 1989/Δ*txtH*. The primers used for PCR amplification are indicated by the red arrows, and the expected product size is also shown. Cosmid 1989 lacks the binding site for the Apra For primer and thus should not generate a product. FRT, Flip recombinase recognition sites. (B) Agarose gel electrophoresis of the PCR products generated using Cosmid 1989/Δ*txtH* (lanes 3 and 4) and Cosmid 1989 (lane 5) as template. A negative control reaction (lane 2) was conducted using water in place of template DNA. The size (kb) of the products was estimated by comparison with the 1kb ladder (lane 1).

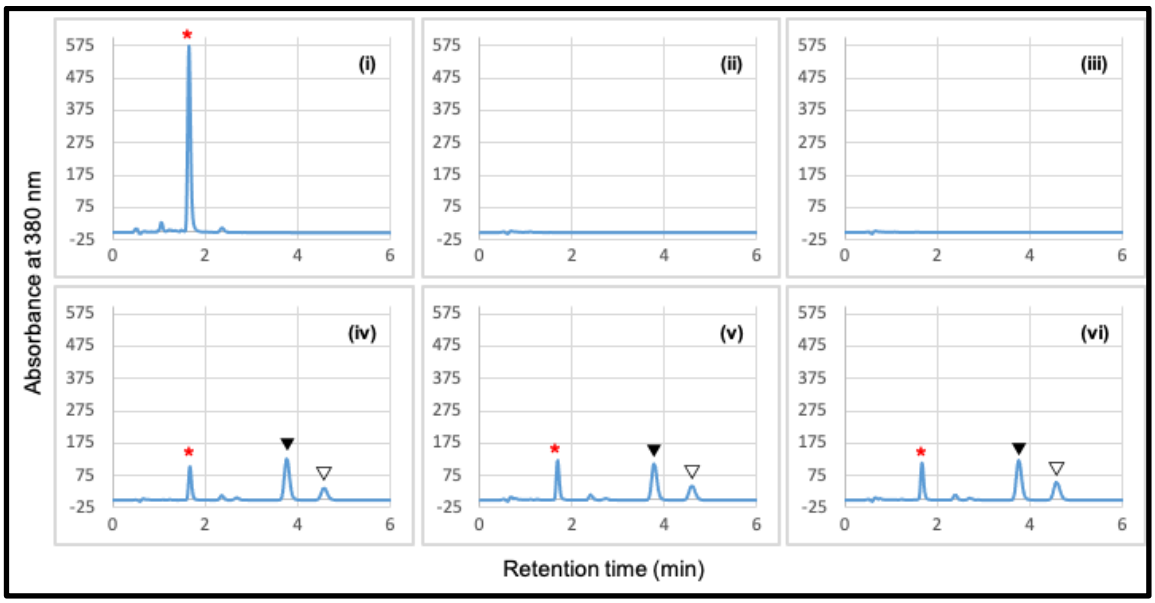

(A)
S. scabiei 87.22 (wild type)
194 bp
PL36 PL35
txtH
PL75
PL76
588 bp
S. scabiei ΔtxtH, Δmlp_lipo/ΔtxtH
FRT
aac(3)IV-oriT
FRT
PL75
PL76
1765 bp
(B)
PL75/PL76 primer set
1 2 3 4 5 6 7
2000 bp
1500 bp
750 bp
500 bp
PL35/PL36 primer set
1 2 3 4 5 6 7
250 bp
(C)
PL75/PL76 primer set
1 2 3 4 5 6 7
2000 bp
1500 bp
750 bp
500 bp
PL35/PL36 primer set
1 2 3 4 5 6 7
250 bp

Supplementary Figure 2.3 PCR verification of the *S. scabiei txtH* deletion mutants. (A) Schematic diagram showing the annealing sites of primers (indicated by the red arrows) used for the PCR verification. The expected product sizes for *S. scabiei* 87.22 (wild type) and the Δ*txtH* or Δmlp_{lipo}/Δ*txtH* mutant strains are indicated. FRT, Flip recombinase recognition sites. (B) Agarose gel electrophoresis of the PCR products generated using genomic DNA from *S. scabiei* 87.22 (lane 7) and from the Δ*txtH* mutant isolates (lanes 3-6). A negative control reaction (lane 2) was conducted for each primer set using water in place of template DNA. The size (kb) of each product was estimated by comparison with the 1kb ladder (lane 1). (C) Agarose gel electrophoresis of the PCR products generated using genomic DNA from *S. scabiei* 87.22 (lane 7) and from the Δmlp_{lipo}/Δ*txtH* mutant isolates (lanes 3-6). A negative control reaction (lane 2) was conducted for each primer set using water in place of template DNA. The size (kb) of each product was estimated by comparison with the 1kb ladder (lane 1).



Supplementary Figure 2.4 PCR verification of the *S. scabiei* Δ*mlp*$_{scab}$ deletion mutants. (A) Schematic diagram showing the annealing sites of the primers (indicated by the red arrows) used for the PCR verification. The expected product sizes for *S. scabiei* 87.22 (wild type), Δ*txtH*/Δ*mlp*$_{scab}$ or Δ*mlp*$_{lipo}$/Δ*txtH*/Δ*mlp*$_{scab}$ mutant isolates are indicated. FRT, Flip recombinase recognition sites. (B) Agarose gel electrophoresis of the PCR products generated using genomic DNA from *S. scabiei* 87.22 (lane 6) and from the Δ*txtH*/Δ*mlp*$_{scab}$

mutant isolates 1-2 (lanes 3-4). A negative control reaction was conducted for each primer set using water (lane 2) in place of template DNA, and a positive control was included for the PL71/PL72 primer set using Cosmid 57/Δ*mlp*$_{scab}$ as template (lane 5). The size (kb) of each product was estimated by comparison with the 1kb ladder for the PL71/PL72 primer set (lane 1) and with the 100 bp ladder for the PL155/PL156 primer set (lane 1). (C) Agarose gel electrophoresis of the PCR products generated using genomic DNA from *S. scabiei* 87.22 (lane 8, left image; lane 7, right image) and from the Δ*mlp*$_{lipo}$/Δ*txtH*/Δ*mlp*$_{scab}$ mutant isolates (lanes 3-6). A negative control reaction was conducted for each primer set using water in place of template DNA (lane 2), and a positive control was included for the PL71/PL72 primer set using Cosmid 57/Δ*mlp*$_{scab}$ as template (lane 7, left image). The size (kb) of each product was estimated by comparison with the 1kb ladder for the PL71/PL72 primer set (lane 1) and with the 100 bp ladder for the PL155/PL156 primer set (lane 1).



Supplementary Figure 2.5 Heterologous complementation of the *S. scabiei* MLP triple mutant. HPLC chromatograms of culture extracts from wild-type *S. scabiei* 87.22 (i), the triple MLP mutant ($\Delta mlp_{lipo}/\Delta txtH/\Delta mlp_{scab}$) (ii), the triple MLP mutant containing plasmid pRLDB50-1a (iii), the triple MLP mutant containing the *txtH* expression plasmid (iv), the triple MLP mutant containing the mlp_{lipo} expression plasmid (v) and the triple MLP mutant containing the mlp_{scab} expression plasmid (vi). The peak corresponding to thaxtomin A in each chromatogram is indicated with the red asterisks, and the peaks corresponding to the thaxtomin B and thaxtomin D intermediates are indicated with ▼ and ∇, respectively.

CHAPTER 3

Functional cross-talk of MbtH-like proteins during thaxtomin biosynthesis in the potato common scab pathogen *Streptomyces scabiei*

Yuting Li, Kapil Tahlan and Dawn R.D. Bignell

© *Frontiers in Microbiology*, 11:585456. doi: 10.3389/fmicb.2020.585456.

3.1 Abstract

Thaxtomin A is a potent phytotoxin that serves as the principle pathogenicity determinant of the common scab pathogen, *Streptomyces scabiei,* and is also a promising natural herbicide for agricultural applications. The biosynthesis of thaxtomin A involves the non-ribosomal peptide synthetases (NRPSs) TxtA and TxtB, and an MbtH-like protein (MLP), TxtH, which may function as a chaperone by promoting the proper folding of the two NRPS enzymes in *S. scabiei*. MLPs are required for the proper function of many NRPS enzymes in bacteria, and they are often capable of interacting with NRPSs from different biosynthetic pathways, though the mechanism by which this occurs is still poorly understood. To gain additional insights into MLP functional cross-talk, we conducted a broad survey of MLPs from diverse phylogenetic lineages to determine if they could functionally replace TxtH. The MLPs were assessed using a protein solubility assay to determine whether they could promote the soluble expression of the TxtA and TxtB adenylation domains. In addition, the MLPs were tested for their ability to restore

thaxtomin production in a *S. scabiei* mutant that lacked TxtH and other endogenous MLPs. Our results showed that the MLPs investigated vary in their ability to exhibit functional cross-talk with TxtH, with two of the MLPs being unable to compensate for the loss of TxtH in the assays performed. The ability of an MLP to serve as a functional partner for the thaxtomin NRPS was not obviously correlated with its overall amino acid similarity with TxtH, but instead with the presence of highly conserved residues. *In silico* structural analysis of TxtH in association with the TxtA and TxtB adenylation domains revealed that several such residues are situated at the predicted interaction interface, suggesting that they might be critical for promoting functional interactions between MLPs and the thaxtomin NRPS enzymes. Overall, our study provides additional insights into the mechanism of MLP cross-talk, and it enhances our understanding of the thaxtomin biosynthetic machinery. It is anticipated that our findings will have useful applications for both the control of common scab disease and the commercial production of thaxtomin A for agricultural use.

3.2 Introduction

Non-ribosomal peptides (NRPs) are a major class of specialized metabolites produced by certain bacteria and filamentous fungi (Marahiel et al., 1997). The biosynthesis of NRPs is performed by non-ribosomal peptide synthetases (NRPSs), which are large multienzyme complexes composed of modules that are each responsible for the incorporation of an amino acid into the growing peptide (Finking and Marahiel, 2004; Strieker et al., 2010). Each module typically constitutes three core domains: an adenylation (A-) domain, a peptidyl carrier protein (PCP-) domain and a condensation (C-) domain.

The A-domain selects a preferred amino acid substrate to initiate the adenylation reaction using Mg·ATP. The activated amino acyl-AMP intermediate is then covalently tethered to the downstream PCP-domain, which serves as the transport unit enabling the bound substrate to move between the different catalytic centers. The C-domain catalyzes the amide bond formation between adjacent PCP-bound intermediates. The biosynthesis of NRPs can involve additional domains that either incorporate modifications into the product or release it from the assembly line (Finking and Marahiel, 2004; Hur et al., 2012; Süssmuth and Mainz, 2017). Furthermore, some NRPSs require auxiliary proteins, including members of the MbtH-like protein (MLP) superfamily, for the optimal activity (Baltz, 2011).

MLPs are named after the MbtH protein, which is an integral component in the biosynthesis of the siderophore mycobactin in *Mycobacterium tuberculosis* (Quadri et al., 1998; McMahon et al., 2012). Proteins belonging to this family are generally small in size (approximately 60-70 amino acids) and are often found within NRP biosynthetic gene clusters (BGCs) that produce antibiotics or siderophores (Baltz, 2011). Several studies have demonstrated a role for these proteins as chaperones in the NRPS assembly line. In these reports, the soluble production of one or more NRPS A-domains in *Escherichia coli* was shown to be reduced or abolished in the absence of the MLP that is from the same biosynthetic pathway as the NRPS, suggesting that the MLP (called the cognate MLP) is required for the proper folding of the A-domain protein (Boll et al., 2011; Imker et al., 2010; Kaniusaite et al., 2020; McMahon et al., 2012; Zolova and Garneau-Tsodikova, 2012, 2014). Additionally, some MLPs have been shown to influence amino acid activation by the corresponding NRPS. In these instances, the NRPS can be heterologously

overexpressed in *E. coli* in soluble form in the absence of the cognate MLP, but the purified protein exhibits low or no activity for the target amino acid *in vitro* unless the purified cognate MLP is added to the reaction, or the MLP is co-expressed with the NRPS (Al-Mestarihi et al., 2014; Boll et al., 2011; Davidsen et al., 2013; Felnagle et al., 2010; Heemstra et al., 2009; Miller et al., 2016; Schomer et al., 2018; Zhang et al., 2010). Previous investigations also noted a 1:1 molar stoichiometry of the MLP-A-domain complex for optimal adenylation activity (Boll et al., 2011; Davidsen et al., 2013).

Several studies have reported that in bacteria containing multiple MLPs, the production of a particular NRP is only abolished in some cases when all of the MLP homologues are eliminated (Lautru et al., 2007; Wolpert et al., 2007). This suggests that MLPs from different NRP pathways can sometimes functionally replace one another, though the reason for this is currently not clear. In addition, some MLPs from other biosynthetic pathways (referred to as non-cognate MLPs) have been shown to be comparable or sometimes even more efficient in enhancing the solubility and/or adenylation activity of NRPS enzymes as compared to the cognate MLP (Boll et al., 2011; Mori et al., 2018a). For example, the *E. coli* enterobactin (ENT) biosynthetic pathway was used as a model to investigate the ability of different non-cognate MLPs to influence the function of the EntF NRPS in the absence of the cognate MLP, YbdZ. They found that non-cognate MLPs vary in their ability to compensate for the loss of YbdZ in the different assays performed, and that the interactions between MLPs and NRPSs are multifaceted and more complex than previously realized (Schomer and Thomas, 2017).

Recently, we examined the importance of MLPs in the biosynthesis of thaxtomin A, which is the principle pathogenicity determinant of the potato common scab pathogen

Streptomyces scabiei (syn. *S. scabies*). Thaxtomin A is a novel nitrated 2,5-diketopiperazine that exhibits potent phytotoxicity against both monocot and dicot plants (King et al., 2001), and it is considered a promising bioherbicide for the control of weed growth (Koivunen et al., 2013; Leep et al., 2010). Production of thaxtomin A in *S. scabiei* is mediated by a BGC that includes two NRPS-encoding genes, *txtA* and *txtB*, which generate the *N*-methylated cyclic dipeptide backbone, and a P450 monooxygenase-encoding gene, *txtC*, which is responsible for the post-cyclization hydroxylation steps (reviewed in Li et al., 2019b). Both TxtA and TxtB contain the three core domains (A-PCP-C) together with a methylation domain integrated into the C-terminal region of the A-domain (Huguet-Tapia et al., 2016). The arrangement of the core domains is unusual when compared to most other NRPSs, which typically have a C-A-PCP domain arrangement (Süssmuth and Mainz, 2017). Immediately downstream of *txtB* is the *txtH* gene, which encodes an MLP that is required for the soluble expression of the TxtA and TxtB A-domains (referred to herein as $TxtA^A$ and $TxtB^A$) in *E. coli,* suggesting that it exhibits a chaperone function in *S. scabiei* (Li et al., 2019a). Deletion of *txtH* in *S. scabiei* significantly reduced thaxtomin A production levels, though some production could still occur. In contrast, production was completely abolished when two non-cognate MLP-encoding genes (mlp_{lipo} and mlp_{scab}) located elsewhere on the chromosome were also deleted. The production of thaxtomin A in the MLP triple mutant could be restored by overexpression of *txtH,* mlp_{lipo} or mlp_{scab}, while overexpression of two non-cognate MLPs from other *Streptomyces* species failed to do so (Li et al., 2019a). Overall, our results showed that the TxtH MLP plays a key role in the biosynthesis of thaxtomin A, and that

some but not all non-cognate MLPs can functionally replace TxtH in the thaxtomin biosynthetic pathway.

In this study, we aimed to further investigate the mechanism of MLP cross-talk by examining the ability of various MLPs from different bacterial species to functionally replace TxtH during the biosynthesis of thaxtomin A. Using protein expression analysis in *E. coli* combined with thaxtomin A production assays in *S. scabiei,* we show that the different MLPs vary in their ability to exhibit functional overlaps with TxtH. Additionally, we conducted an *in silico* structural analysis of the protein complex involving the thaxtomin (Txt) A-domains with TxtH in order to identify potential residues that may play a key role in the Txt MLP-NRPS interaction. Our work not only provides additional insights into the mechanism of MLP functional cross-talk, but it also enhances our understanding of the thaxtomin biosynthetic machinery, and this in turn could have useful applications for both the control of common scab disease and the commercial production of thaxtomin A for agricultural use.

3.3 Materials and Methods

3.3.1 Bacterial strains, culture conditions and maintenance

E. coli strains used in this study are listed in Table 3.1. Strains were routinely cultivated at 37°C unless otherwise indicated. Liquid cultures were grown with shaking (200-250 rpm) in Luria-Bertani (LB) Lennox medium (Fisher Scientific, Ottawa, ON, Canada), low salt LB broth (1% w/v tryptone; 0.5% w/v yeast extract; 0.25% w/v NaCl), super optimal broth (SOB) or super optimal broth with catabolite repression (SOC) medium

(New England Biolabs, Whitby, ON, Canada), while solid cultures were grown on LB Lennox (or low salt LB) medium containing 1.5% w/v agar (NEOGEN, Michigan, US). When required, the solid or liquid growth media were supplemented with antibiotics as described before (Li et al., 2019a). *E. coli* strains were maintained at 4°C for short-term storage or at -80ºC in 20% v/v glycerol for long-term storage (Sambrook and Russell, 2001).

Table 3.1 Bacterial strains used in this study.

Strain	**Description**	**Resistance**[†]	**Reference or source**
***Escherichia coli* strains**			
DH5α	General cloning host	n/a	Gibco-BRL
NEB5α	DH5α derivative, high efficiency competent cells	n/a	New England Biolabs
BL21(DE3)	Source of genomic DNA for amplifying the *ybdZ* coding sequence	n/a	New England Biolabs
BL21(DE3)*ybdZ:aac(3)IV*	BL21(DE3) derivative, *ybdZ* replaced with an apramycin resistance cassette (*aac(3)IV*)	$Apra^R$	Herbst et al., 2013
ET12567/pUZ8002	*dam*$^-$, *dcm*$^-$, *hsdS*$^-$; nonmethylating conjugation host	Kan^R, Cml^R	Kieser et al., 2000
***Streptomyces* strains**			
Streptomyces scabiei 87.22	Wild-type strain	n/a	Loria et al., 1995
S. scabiei Δ*txtH*	87.22 derivative in which the *txtH* MLP-coding gene has been deleted	$Apra^R$	Li et al., 2019a
S. scabiei Δmlp_{lipo}/Δ*txtH*/Δmlp_{scab}	*S. scabiei* 87.22 derivative in which the *SCAB3331*(*mlp_{lipo}*), *txtH* and *SCAB85461*(*mlp_{scab}*) MLP-coding genes have been deleted	$Apra^R$, Hyg^R	Li et al., 2019a

Streptomyces coelicolor A3(2) M145	Source of genomic DNA for amplifying the *cdaX* and *cchK* coding sequences	n/a	Kieser et al., 2000
Streptomyces sp. 11-1-2	Source of genomic DNA for amplifying the *CGL27_RS10110* and *CGL27_RS02360* coding sequences	n/a	Bown and Bignell, 2017
Streptomyces europaeiscabiei 89-04	Source of genomic DNA for amplifying the *AWZ11_RS05060* coding sequence	n/a	Zhang et al., 2016
Streptomyces clavuligerus ATCC27064	Source of genomic DNA for amplifying the *SCLAV_p1293* coding sequence	n/a	ATCC

† $Apra^R$, Kan^R, Cml^R and Hyg^R = apramycin, kanamycin, chloramphenicol and hygromycin resistance, respectively.
n/a = not applicable.

Streptomyces strains used in this study are listed in Table 3.1. Strains were routinely cultured at 28°C unless otherwise indicated. Liquid cultures were typically grown with shaking (200 rpm) in trypticase soy broth (TSB; BD Biosciences, Mississauga, ON, Canada) medium with stainless steel springs. *S. scabiei* cultures for analysis of thaxtomin production were prepared by inoculating oat bran broth containing 0.35% w/v cellobiose (OBBC) with TSB seed cultures of each strain and then incubating at 25°C for 7 days as described before (Li et al., 2019a). Plate cultures were grown on potato mash agar (PMA; Fyans et al., 2016), International *Streptomyces* Project Medium 4 (ISP-4; BD Biosciences), nutrient agar (BD Biosciences, 1.5 % w/v agar) and soy flour mannitol agar (SFMA; Kieser et al., 2000). When required, the growth medium was supplemented with apramycin, nalidixic acid, kanamycin or hygromycin B (50 µg/mL final concentration; Millipore Sigma, Oakville, ON, Canada).

3.3.2 Plasmids, primers and DNA manipulation

Plasmids used in this study are listed in Table 3.2. Standard molecular biology procedures were implemented for all DNA manipulations performed (Sambrook and Russell, 2001). *Streptomyces* genomic DNA was isolated from mycelia harvested from TSB cultures using the DNeasy Blood & Tissue Kit as per the manufacturer's protocol (QIAgen Inc, Canada). The nucleotide sequences of the MLP-encoding genes *MXAN_3118* (from *Myxococcus xanthus* DK1622), *RHA1_ro04717* (from *Rhodococcus jostii* RHA1), *PA2412* (from *Pseudomonas aeruginosa* PA01) and *ybdZ* [from *Escherichia coli* BL21(DE3)] were codon optimized for expression in *Streptomyces* using a webserver (https://www.idtdna.com/codonopt) from Integrated DNA Technologies (Coralville, IA, USA). The codon optimized sequences along with *cloY* (from *Streptomyces roseochromogenes* subsp. *oscitans* DS12.976) and *comB* (from *Streptomyces lavendulae*) were then synthesized with 30-60 bp flanking regions by TWIST BIOSCIENCE (South San Francisco, CA, USA) (Supplementary Data File 3.1). All oligonucleotide primers used for cloning, PCR and sequencing were purchased from Integrated DNA Technologies and are listed in Supplementary Table 3.1. Restriction enzymes were purchased from New England Biolabs. PCR was routinely performed using Phusion or *Taq* DNA polymerase (New England Biolabs) according to the manufacturer's instructions, except that 5% v/v DMSO was included in the reactions. DNA sequencing was performed by The Centre for Applied Genomics (Toronto, ON, Canada).

Table 3.2 Plasmids used in this study.

Plasmid	**Description**	**Resistance**[†]	**Reference or source**
pGEM-T EASY	General cloning vector	Amp^R	Promega Corporation
pGEM-T EASY/*comB*	pGEM-T EASY derivative containing a 312 bp insert of the *comB* gene with flanking regions	Amp^R	This study
pGEM-T EASY/*cloY*	pGEM-T EASY derivative containing a 306 bp insert of the *cloY* gene with flanking regions	Amp^R	This study
pGEM-T EASY/*MXAN_3118*	pGEM-T EASY derivative containing a 306 bp insert of the *MXAN_3118* gene‡ with flanking regions	Amp^R	This study
pGEM-T EASY/*PA2412*	pGEM-T EASY derivative containing a 309 bp insert of the *PA2412* gene‡ with flanking regions	Amp^R	This study
pGEM-T EASY/*RHA1_ro04717*	pGEM-T EASY derivative containing a 342 bp insert of the *RHA1_ro04717* gene‡ with flanking regions	Amp^R	This study
pGEM-T EASY/*ybdZ*	pGEM-T EASY derivative containing a 300 bp insert of the *ybdZ* gene‡ with flanking regions	Amp^R	This study
pET28b	N- or C- terminal 6×histidine fusion tag protein expression vector with T7 promoter and *lac* operator	Kan^R	Novagen
pET28b/HIS_6-*txtH*	pET28b derivative containing a DNA	Kan^R	Li et al., 2019b

	fragment for expression of the HIS_6-TxtH protein		
pET28b/HIS_6-*cdaX*	pET28b derivative containing a DNA fragment for expression of the HIS_6-CdaX protein	Kan^R	This study
pET28b/HIS_6-*cchK*	pET28b derivative containing a DNA fragment for expression of the HIS_6-CchK protein	Kan^R	This study
pET28b/HIS_6-*SCLAV_p1293*	pET28b derivative containing a DNA fragment for expression of the HIS_6-SCLAV_p1293 protein	Kan^R	This study
pET28b/HIS_6-*ybdZ*	pET28b derivative containing a DNA fragment for expression of the HIS_6-YbdZ protein	Kan^R	This study
pET28b/HIS_6-*CGL27_RS10110*	pET28b derivative containing a DNA fragment for expression of the HIS_6-CGL27_RS10110 protein	Kan^R	This study
pET28b/HIS_6-*CGL27_RS02360*	pET28b derivative containing a DNA fragment for expression of the HIS_6-CGL27_RS02360 protein	Kan^R	This study
pET28b/HIS_6-*AWZ11_RS05060*	pET28b derivative containing a DNA fragment for expression of the HIS_6-AWZ11_RS05060 protein	Kan^R	This study
pET28b/HIS_6-*comB*	pET28b derivative containing a DNA fragment for expression of the HIS_6-ComB protein	Kan^R	This study
pET28b/HIS_6-*cloY*	pET28b derivative containing a DNA fragment for expression of the HIS_6-CloY protein	Kan^R	This study

pET28b/HIS$_6$-*MXAN_3118*	pET28b derivative containing a DNA fragment‡ for expression of the HIS$_6$-MXAN_3118 protein	Kan^R	This study
pET28b/HIS$_6$-*PA2412*	pET28b derivative containing a DNA fragment‡ for expression of the HIS$_6$-PA2412 protein	Kan^R	This study
pET28b/HIS$_6$-*RHA1_ro04717*	pET28b derivative containing a DNA fragment‡ for expression of the HIS$_6$-RHA1_ro04717 protein	Kan^R	This study
pACYCDuet-1	N-terminal 6×histidine fusion tag expression vector with T7 promoter and *lac* operator	Cml^R	Novagen
pACYCDuet-1/HIS$_6$-*txtA^A*	pACYCDuet-1 derivative containing a DNA fragment for expression of the HIS$_6$-TxtA^A protein	Cml^R	Li et al., 2019b
pACYCDuet-1/HIS$_6$-*txtB^A*	pACYCDuet-1 derivative containing a DNA fragment for expression of the HIS$_6$-TxtB^A protein	Cml^R	Li et al., 2019b
pRFSRL16	Harbours the *egfp* gene downstream of the *ermE*p* promoter and an RBS; integrates into the ΦC31 *attB* site	Apra^R, Kan^R	Joshi *et al.*, 2010
pRFSRL16/*txtH*	pRFSRL16 derivative in which *egfp* is replaced with the *S. scabiei txtH* gene	Apra^R, Kan^R	This study
pRFSRL16/*mlp$_{lipo}$*	pRFSRL16 derivative in which *egfp* is replaced with the *S. scabiei mlp$_{lipo}$* gene	Apra^R, Kan^R	This study
pRFSRL16/*cdaX*	pRFSRL16 derivative in which *egfp* is replaced with the *S. coelicolor cdaX* gene	Apra^R, Kan^R	This study

pRFSRL16/*cchK*	pRFSRL16 derivative in which *egfp* is replaced with the *S. coelicolor cchK* gene	Apra^R, Kan^R	This study
pRFSRL16/*SCLAV_p1293*	pRFSRL16 derivative in which *egfp* is replaced with the *S. clavuligerus SCLAV_p1293* gene	Apra^R, Kan^R	This study
pRFSRL16/*ybdZ*	pRFSRL16 derivative in which *egfp* is replaced with the *E. coli* BL21(DE3) *ybdZ* gene‡	Apra^R, Kan^R	This study
pRFSRL16/*CGL27_RS10110*	pRFSRL16 derivative in which *egfp* is replaced with the *Streptomyces* sp. 11-1-2 *CGL27_RS10110* gene	Apra^R, Kan^R	This study
pRFSRL16/*CGL27_RS02360*	pRFSRL16 derivative in which *egfp* is replaced with the *Streptomyces* sp. 11-1-2 *CGL27_RS02360* gene	Apra^R, Kan^R	This study
pRFSRL16/*AWZ11_RS05060*	pRFSRL16 derivative in which *egfp* is replaced with the *S. europaeiscabiei* 89-04 *AWZ11_RS05060* gene	Apra^R, Kan^R	This study
pRFSRL16/*comB*	pRFSRL16 derivative in which *egfp* is replaced with the *S. lavendulae comB* gene	Apra^R, Kan^R	This study
pRFSRL16/*cloY*	pRFSRL16 derivative in which *egfp* is replaced with the *S. roseochromogenes subsp. oscitans* DS12.976 *cloY* gene	Apra^R, Kan^R	This study
pRFSRL16/*MXAN_3118*	pRFSRL16 derivative in which *egfp* is replaced with *M. xanthus* DK1622 *MXAN_3118* gene‡	Apra^R, Kan^R	This study
pRFSRL16/*PA2412*	pRFSRL16 derivative in which *egfp* is replaced	Apra^R, Kan^R	This study

	with the *P. aeruginosa* PA01 *PA2412* gene‡		
pRFSRL16/*RHA1_ro04717*	pRFSRL16 derivative in which *egfp* is replaced with the *R. jostii* RHA1 *RHA1_ro04717* gene‡	$Apra^R$, Kan^R	This study

† Amp^R, $Apra^R$, Kan^R and Cml^R= ampicillin, apramycin, kanamycin and chloramphenicol resistance, respectively.
‡ Gene sequence was codon optimized for expression in *Streptomyces* spp.

3.3.3 Construction of *E. coli* protein expression plasmids

Construction of the expression plasmids pACYCDuet-1/HIS_6-*txtA*^{A}, pACYCDuet-1/HIS_6-*txtB*^{A} and pET28b/HIS_6-*txtH* was described in Li et al. (2019a). The MLP-encoding genes *CGL27_RS10110* and *CGL27_RS02360* from *Streptomyces* sp. 11-1-2, *cdaX* and *cchK* gene from *S. coelicolor*, *SCLAV_p1293* from *S. clavuligerus*, *AWZ11_RS05060* from *S. europaeiscabiei* and *ybdZ* from *E. coli* (Table 3.1) were PCR-amplified using genomic DNA as template and using primers with *Nde*I and *Eco*RI restriction sites added. The PCR products were purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Canada) and were then digested with *Nde*I and *Eco*RI and ligated into similarly digested pET28b (Table 3.2). The synthetic gene fragments for *comB*, *cloY*, *MXAN3118*, *PA2412* and *RHA1_ro04717* were cloned into the pGEM-T EASY vector (Promega North America, USA) as per the manufacturer's instructions (Table 3.2). The resulting plasmids were then used as templates for PCR amplification using primers listed in Supplementary Table 3.1. The gene products were each purified and then cloned into the *Nde*I/*Eco*RI restriction sites of pET28b except for *comB,* which was cloned into the *Nde*I/*Bam*HI vector restriction sites (due to the presence of an *Eco*RI site within the gene sequence). The cloned inserts in all constructed expression vectors were then verified by DNA sequencing.

3.3.4 Co-expression of HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$ with HIS_6-tagged MLPs

The co-expression of HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$ with HIS_6-tagged MLPs was conducted as previously described (Li et al., 2019a). Briefly, the expression strain *E. coli* BL21(DE3)*ybdZ:aac(3)IV* (Table 3.1) containing either pACYCDuet-1/HIS_6-$txtA^A$ or pACYCDuet-1/HIS_6-$txtB^A$, with and without a pET28b-derived MLP expression plasmid (Table 3.2), was cultured overnight in 3 mL of LB medium supplemented with 1% w/v glucose and the appropriate antibiotics. The overnight cultures were subcultured into fresh LB medium containing appropriate antibiotics, and the cultures were incubated at 37°C and 200 rpm until the OD_{600} reached 0.4–0.6. The production of the HIS_6-tagged proteins was induced by adding 1 mM isopropyl β-d-thiogalactopyranoside (IPTG) and then incubating the cultures at 16°C and 200 rpm for 48 hours. Cells from 1 mL of culture were harvested and were resuspended in 200 µL of 50 mM Tris-HCl (pH 8.0) containing 1 × cOmplete EDTA-free protease inhibitor. The cells were lysed by sonication and the cell debris was removed by centrifugation. The soluble proteins were collected, and the protein concentration was quantified using a Bradford protein assay kit (Fisher Scientific).

3.3.5 Western blot analysis

Equal amounts (10 µg) of total soluble protein extracts were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) on a 15% w/v gel before being transferred to an Amersham™ Hybond™ ECL membrane (GE Healthcare Canada Inc., Canada) as described by the manufacturer's instructions. To ensure equal loading of

each protein sample, separate polyacrylamide gels were prepared and then stained with Coomassie Brilliant Blue stain (50% v/v methanol, 10% v/v glacial acetic acid, 0.1% w/v Coomassie Blue) (Supplementary Figure 3.1). Membranes were blocked overnight in TBS-T buffer (50 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.05% v/v Tween 20) containing 5% w/v skim milk, and were then incubated with 6 × His Epitope Tag Antibody (mouse IgG2b) (Fisher Scientific) at a 1:2000 dilution. The membranes were washed several times with TBS-T buffer and were then incubated with the secondary antibody (Goat anti-mouse IgG2b, HRP conjugate) (Fisher Scientific) at a 1:2000 dilution. The membranes were processed using the ECL™ western blotting high sensitivity detection reagent (GE Healthcare) and were visualized by ImageQuant LAS4000 Biomolecular Imager (GE Healthcare). The intensity of the HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$ protein bands was quantified using ImageJ (Schneider et al., 2012) and the average % band intensity relative to the appropriate control (HIS_6-$TxtA^A$ or HIS_6-$TxtB^A$ co-expressed with HIS_6-TxtH) was calculated from triplicate membranes (Supplementary Figure 3.2) that were prepared using protein extracts from three independent cultures for each strain. Statistical analysis of the results was conducted in Minitab 19 (Minitab LLC, State College, PA, USA) using one-way ANOVAs with *a posteriori* multiple comparisons of least squared means performed using the Tukey test. P values ≤ 0.05 were considered as statistically significant in all analyses.

3.3.6 Construction of plasmids for overexpression of MLPs in *S. scabiei*

The MLP-encoding genes were PCR-amplified using the corresponding pET28b plasmid clone (for *txtH, mlp*$_{lipo}$*, cdaX, cchK, SCLAV_p1293, CGL_RS10110, CGL27_RS02360* and *AWZ11_RS05060*) or the pGEM-T EASY clone (for *comB, cloY, MXAN3118, PA2412, RHA1_ro04717* and *ybdZ*) as template (Table 3.2) and using gene-specific primers (Supplementary Table 3.1) with *Nde*I and *Not*I restriction sites added. The PCR products were digested with *Nde*I and *Not*I and then ligated into similarly digested pRFSRL16 (Joshi et al., 2010). The resulting plasmids (Table 3.2) contained the cloned MLP-encoding gene in place of the *egfp* gene in pRFSRL16, and each were verified by sequencing. The plasmids along with the control vector (pRFSRL16) were then introduced into the *S. scabiei* Δ*mlp*$_{lipo}$/Δ*txtH*/Δ*mlp*$_{scab}$ mutant (Table 3.1) by intergeneric conjugation with *E. coli* as described before (Kieser et al., 2000).

3.3.7 Analysis of thaxtomin production

Thaxtomins were extracted from *S. scabiei* OBBC cultures and were detected by reverse phase HPLC as described before (Li et al., 2019a). Briefly, each strain was cultured in triplicate, and in the case of the MLP overexpression strains, two different isolates per strain were cultured in triplicate for a total of six cultures. Culture extracts were prepared by extracting the culture supernatants with ethyl acetate, drying the extracts by evaporation, and resuspending the residual material in 100% v/v HPLC-grade methanol. The extracts were analyzed using an Agilent 1260 Infinity Quaternary LC system (Agilent Technologies Canada Inc.) with a Poroshell 120 EC-C18 column (4.6 × 50 mm, 2.7 μm particle size;

Agilent Technologies Canada, Inc.) held at a constant temperature of 40°C. An isocratic mobile phase consisting of 30% acetonitrile and 70% water at a constant flow rate of 1.0 mL/min was used for metabolite separation, and metabolites were monitored using a detection wavelength of 380 nm. The normalized total thaxtomin production level for each culture was determined by summing the measured peak area for thaxtomin A, thaxtomin B and thaxtomin D and then dividing the total area by the measured dry cell weight of the culture. The results for each strain were then averaged among the replicate samples and were reported as the percent thaxtomin production relative to wild-type *S. scabiei* 87.22. Statistical analysis of the results was conducted in Minitab 19 using one-way ANOVAs with *a posteriori* multiple comparisons of least squared means performed using the Tukey test. P values ≤ 0.05 were denoted as statistically significant in all analyses.

3.3.8 Bioinformatics analysis and structural modeling

Identification of the adenylation domain within the TxtA and TxtB amino acid sequences was performed as described previously (Li et al., 2019a). The homologues of TxtH were identified using the NCBI Protein Basic Local Alignment Search Tool (BLASTP) (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The cutoff used to select MLPs for analysis was 39% end to end amino acid identity with the *S. scabiei* TxtH. In total, 133 MLPs were chosen from different phyla, and the accession numbers for the proteins used are listed in Supplementary Table 3.2. Amino acid sequence alignments were generated using ClustalW within the Geneious version 6.1.2 software (Biomatters Ltd.). Phylogenetic trees were constructed from the alignments using the maximum likelihood method in the

MEGA X software (Kumar et al., 2018) and using the Whelan and Goldman plus gamma (WAG + G) substitution model (Whelan and Goldman, 2001). Bootstrap analyses were performed with 1000 replicates and the Interactive Tree of Life (iTOL) was used to visualize the tree (Letunic and Bork, 2007; https://itol.embl.de/).

The *in silico* 3-dimentional structures of the *S. scabiei* $TxtA^A$, $TxtB^A$ and TxtH were prepared using SWISS-MODEL (Biasini et al., 2014). The crystal structure of the TioS NRPS (PDB ID: 5wmm_1; Mori et al., 2018b) from *Micromonospora* sp. ML1 was used as the template for both the $TxtA^A$ and $TxtB^A$ models. The model of TxtH was generated based on the crystal structure of the FscK MLP from *Thermobifida fusca* (PDB ID: 6ea3_1; Bruner and Zagulyaeva, unpublished). The generated models (Supplementary Data 3.2, 3.3, 3.4) were evaluated by different parameters using the SWISS-MODEL webserver (Supplementary Table 3.3; https://swissmodel.expasy.org/) and were visualized using PyMOL (DeLano, 2002). The interface between the TioT-TioS complex (PDB ID: 5wmm) and the FscK-FscH complex (PDB ID: 6ea3) was analyzed using the Proteins, Interfaces, Structures, Assemblies software (PISA) server (Krissinel and Henrick, 2007; https://www.ebi.ac.uk/pdbe/pisa/) for use in homology modelling analysis. The TxtH model was docked with the $TxtA^A$ or $TxtB^A$ model in PyMOL based on the location of TioT in the TioT-TioS complex.

3.4 Results and Discussion

3.4.1 Selection of non-cognate MLPs for functional studies

In order to investigate the ability of different non-cognate MLPs to functionally replace TxtH in the thaxtomin biosynthetic pathway, we first conducted a phylogenetic analysis of 133 MLPs from the database, which included TxtH homologues from known or predicted thaxtomin producers, and other previously studied MLPs (Figure 3.1). This led to the identification of 12 candidate MLPs from diverse phylogenetic clades (Figure 3.1) that exhibited between 39-59% amino acid identity with TxtH (Table 3.3). Three of the MLPs originate from different species within the Proteobacteria, while the remaining nine MLPs originate from Actinobacteria, including different species of *Streptomyces* and a strain of *Rhodococcus jostii* (Figure 3.1). Eleven of the MLPs are associated with BGCs that are known or predicted to produce different types of NRP metabolites (Table 3.3), and six are encoded immediately next to a NRPS-encoding gene within BGCs (Supplementary Figure 3.3). Only one MLP is not encoded within a specific gene cluster and is therefore considered an orphan MLP (Table 3.3).

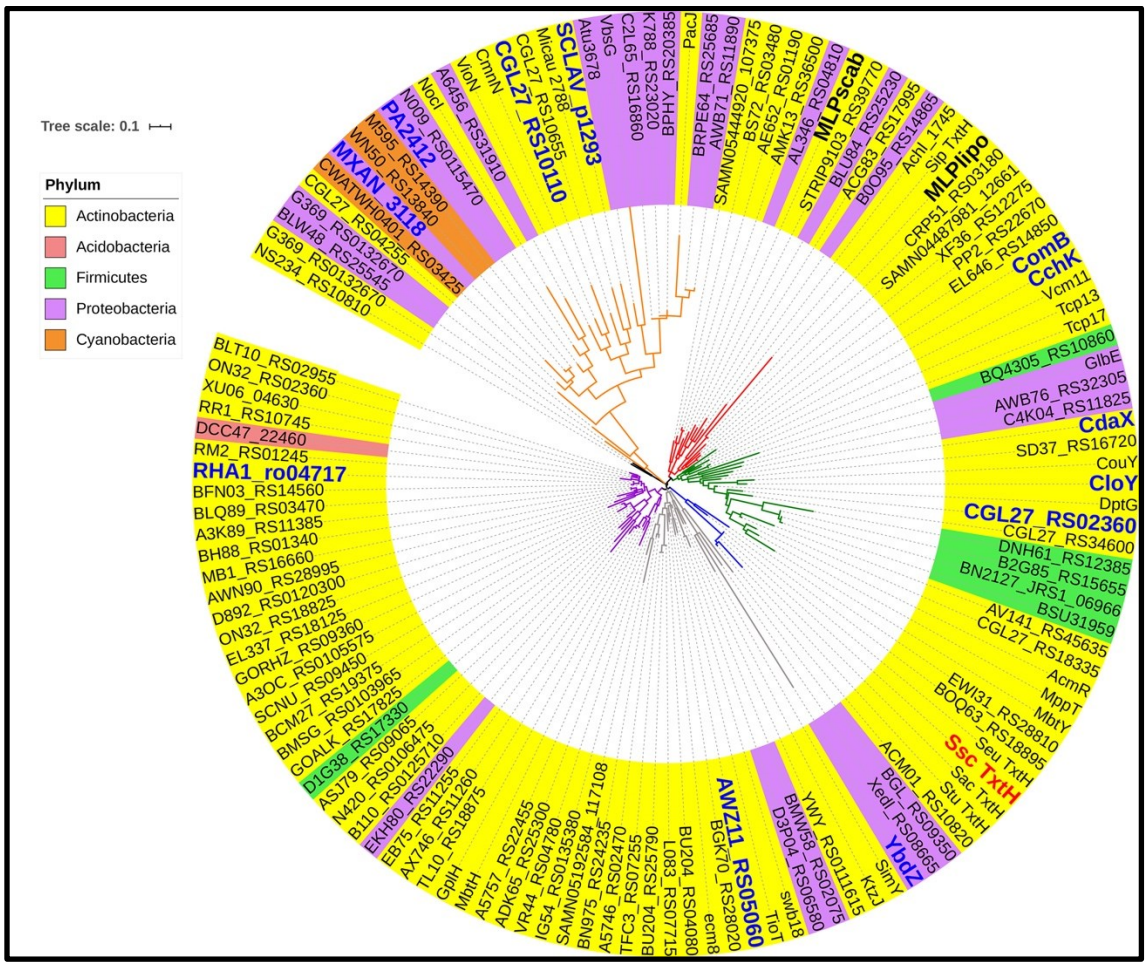


Figure 3.1 Phylogenetic analysis of MLPs. The phylogeny was generated from the amino acid sequences of 133 MLPs from the database. The MLPs originate from the phyla Actinobacteria (yellow), Acidobacteria (fuchsia), Proteobacteria (purple), Firmicutes (green) and Cyanobacteria (orange). TxtH from *S. scabiei* is highlighted in red, and the two other MLPs encoded in the *S. scabiei* genome (MLP_{lipo}, MLP_{scab}) are indicated in bold. MLPs that were chosen for functional analysis in this study are labeled in blue. Diverse lineages are shown in different colors, and the scale bar indicates the number of amino acid substitutions per site. Information about each MLP is provided in Supplementary Table 3.2. Ssc, *Streptomyces scabiei*; Sac, *Streptomyces acidiscabies*; Stu, *Streptomyces turgidiscabies*; Seu, *Streptomyces europaeiscabiei*.

Table 3.3 Overview of non-cognate MLPs tested in this study and their amino acid sequence identity/similarity to *S. scabiei* TxtH.

Bacterial Strain	MLP	Product	Product Class	Identity/ Similarity to TxtH (%)
Streptomyces coelicolor A3(2)	CdaX	Calcium-dependent antibiotic	Cyclic lipodepsipeptide	55/73
Streptomyces clavuligerus ATCC 27064	SCLAV_p1293	Putative maduropeptin	NRPS, T1PKS, ectoine, phosphoglycolipid	40/61
Streptomyces sp. 11-1-2	CGL27_RS02360	Putative skyllamycin	NRPS, arylpolyene, ladderane	52/73
Streptomyces sp. 11-1-2	CGL27_RS10110	Putative toyocamycin	NRPS, nucleoside	41/62
Escherichia coli BL21(DE3)	YbdZ	Enterobactin	Siderophore	40/60
Streptomyces europaeiscabiei 89-04	AWZ11_RS05060	Putative thiocoraline	NRPS, terpene	57/80
Streptomyces coelicolor A3(2)	CchK	Coelichelin	Peptide siderophore	54/70
Myxococcus xanthus DK 1622	MXAN_3118	†	†	39/59
Pseudomonas aeruginosa PAO1	PA2412	Pyoverdine	Siderophore	46/63
Rhodococcus jostii RHA1	RHA1_ro04717	Putative erythrochelin	NRPS	59/77
Streptomyces roseochromogenes subsp. oscitans DS12.976	CloY	Clorobiosin	Aminocoumarin	48/70
Streptomyces lavendulae	ComB	Complestatin	Glycopeptide-like	54/71

† MXAN_3118 is not associated with a specific NRP biosynthetic gene cluster

Among the chosen MLPs candidates, the importance of several in NRP biosynthesis has been demonstrated in previous studies. For instance, CdaX is encoded by a gene from the known calcium-dependent peptide antibiotic (CDA) BGC in *S. coelicolor* (Table 3.3) and can functionally replace CchK, which is encoded in the gene cluster responsible for producing the siderophore coelichelin in the same organism. The deletion of either *cdaX* or *cchK* reduces but does not abolish the production of the respective NRP products, while the disruption of both genes completely eliminates the production of both metabolites (Lautru et al., 2007). Additionally, CdaX has been shown to stimulate the activities of L-tyrosine-adenylating enzymes from different NRPS biosynthetic pathways (Boll et al., 2011). In contrast, results from our previous study suggested that CdaX is unable to functionally replace TxtH in the thaxtomin biosynthetic pathway (Li et al., 2019a). CloY from the clorobiocin BGC of *S. roseochromogenes* (Table 3.3; Pojer et al., 2002) is essential for production of the aminocoumarin antibiotic (Wolper et al., 2007), as it is required for the solubility and adenylation activity of its corresponding NRPS partner, CloH (Boll et al., 2011). The ComB-encoding gene is situated within a glycopeptide-like complestatin NRP BGC from *S. lavendulae* (Chiu et al., 2001) and was recently shown to stimulate the production of several NRPs in the mold *Penicillium chrysogenum,* which does not harbor any MLP-encoding genes in its genome (Zwahlen et al., 2019). YbdZ has been extensively investigated in recent studies and is required for the biosynthesis of the ENT siderophore in *E. coli* (Schomer and Thomas, 2017; Schomer et al., 2018). The deletion of *ybdZ* abolishes ENT production even though its NRPS partner (EntF) is not dependent on the presence of YbdZ for soluble protein production, and biochemical analyses have shown that the solubility and catalytic activity of EntF is significantly enhanced by YbdZ (Felnagle

et al., 2010). PA2412 is the MLP associated with the biosynthesis of the siderophore pyoverdine in *P. aeruginosa,* and strains without PA2412 cannot produce pyoverdine or grow under iron-restricted conditions (Drake et al., 2007). Furthermore, PA2412 has the ability to promote ENT biosynthesis in *E. coli* in the absence of YbdZ (Schomer and Thomas, 2017). Intriguingly, the orphan MLP MXAN_3118 from *Myxococcus xanthus* is not encoded within any NRP BGC, but it is able to interact with seven different NRPSs that are encoded elsewhere in the genome of this organism (Esquilín-Lebrón et al., 2018). In addition, MXAN_3118 can functionally replace YbdZ in multiple assays conducted in *E. coli* (Schomer and Thomas, 2017) and is therefore thought to be a promising "universal" MLP for promoting heterologous expression of NRPSs in bacterial and fungal strains in order to improve metabolite production.

In addition to MLPs with known function, we chose MLP candidates for our study that have not been previously characterized and which are associated with predicted NRP BCGs (Supplementary Figure 3.3). Three (CGL27_RS10110, CGL27_RS02360, AWZ11_RS05060) are from the plant pathogenic species *S. europaeiscabiei* (Zhang et al., 2016) and *Streptomyces* sp. 11-1-2 (Bown and Bignell, 2017), and one (RHA1_ro04717) is from the actinobacterium *Rhodococcus jostii,* which is known for its ability to transform a variety of organic compounds and pollutants (Martínková et al., 2009). In addition, we included SCLAV_p1293, which is associated with a predicted BGC on the linear plasmid of *S. clavuligerus* and was previously found to be unable to promote thaxtomin production in the *S. scabiei* MLP triple mutant (Li et al., 2019a).

3.4.2 Non-cognate MLPs from different bacteria can promote the solubility of the TxtA and TxtB A-domains to varying degrees

Previously, we showed that TxtH is required for the soluble production of both $TxtA^A$ and $TxtB^A$ in *E. coli*, suggesting that it functions as a chaperone to promote the proper folding of the NRPS adenylating enzymes. Two non-cognate MLPs encoded elsewhere on the *S. scabiei* chromosome were also shown to be able to promote the soluble production of $TxtA^A$ and $TxtB^A$, suggesting that some MLPs can exhibit functional redundancy with TxtH (Li et al., 2019a). To determine whether non-cognate MLPs from other bacterial species are able to exhibit functional cross-talk with TxtH, we expressed each A-domain with an N-terminal HIS_6 tag together or without an MLP, which also harbored an N-terminal HIS_6 tag. The amount of HIS_6-tagged $TxtA^A$ and $TxtB^A$ when co-expressed with each MLP was then assessed in soluble protein fractions by western blot analysis using antibodies against the HIS_6 tag.

Compared to TxtH, the non-cognate MLPs promoted the production of the two A-domains in soluble form with varying efficiencies (Figure 3.2A). In the case of $TxtA^A$, co-expression with SCLAV_p1293, YbdZ, CGL27_RS10110 and MXAN_3118 resulted in reduced soluble protein production, though the observed differences were not statistically significant when compared with the TxtH co-expression (Figure 3.2B). In contrast, the remaining MLPs promoted similar or higher soluble $TxtA^A$ protein levels (Figure 3.2A, B). For $TxtB^A$, co-expression with CdaX, CchK and CGL27_RS02360 resulted in similar or higher amounts of soluble protein production when compared to the TxtH co-expression. However, the remaining MLPs failed, or promoted reduced levels of soluble $TxtB^A$ production, with most resulting in statistically significant differences in protein levels when

compared to TxtH (Figure 3.2A, B). Production of both A-domains in soluble form was most severely impacted by co-expression with YbdZ and CGL27_RS10110, followed by SCLAV_p1293 and MXAN_3118. Of the two Txt NRPS A-domains, the soluble production of $TxtB^A$ was more strongly impacted by the different co-expressed MLP partners (Figure 3.2A, B). This is in accordance with previous reports showing differences in MLP-NRPS A-domain interactions, even within the same NRP biosynthetic pathway involving multiple NRPS enzymes (Davidsen et al., 2013; Felnagle et al., 2010; McMahon et al. 2012). Although there was some variability in the relative expression level of the MLPs in the *E. coli* strain based on SDS-PAGE analysis of the total soluble protein extracts (Supplementary Figure 3.1), we found no correlation between the amount of MLP detected and the amount of soluble A-domain protein produced when co-expressed with the MLP. For example, SCLAV_p1293 and MXAN_3118 were both detected at higher levels than TxtH in the total protein extracts, but neither were able to promote efficient production of soluble $TxtB^A$. In contrast, CGL27_RS02360 was not readily detectable in the total extracts, but it was able to promote the soluble production of both A-domains to levels comparable to those observed with TxtH. Our observations are consistent with other studies that also found no correlation between the detectable level of an MLP and its ability to promote soluble A-domain protein production in *E. coli* (Schomer and Thomas, 2017; Schomer et al., 2018). Overall, our results show that several phylogenetically distinct MLPs have the ability to functionally replace TxtH in promoting the soluble production of the Txt NRPS adenylating enzymes in *E. coli* to varying degrees, though not all MLPs are able to do so.

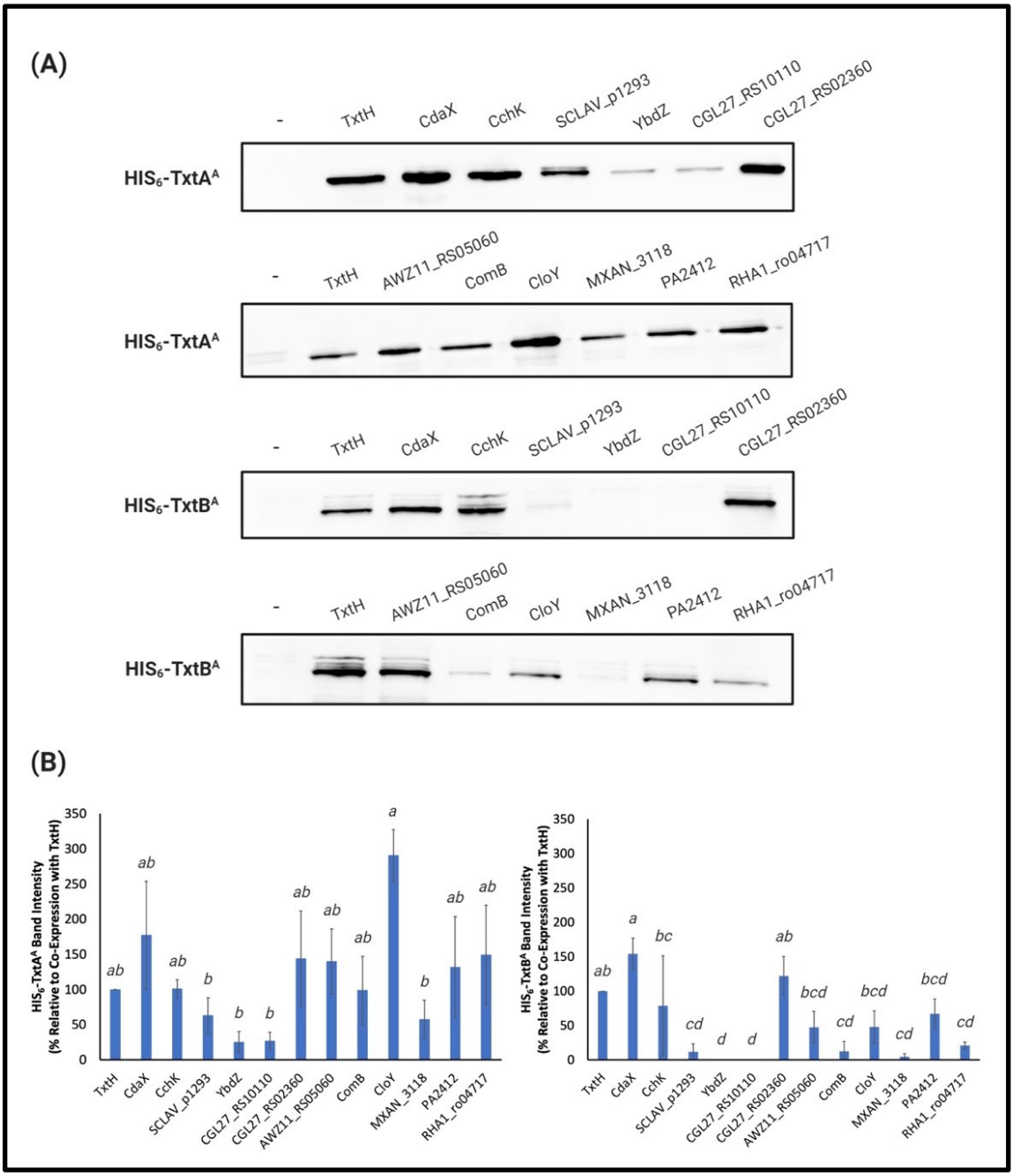


Figure 3.2 (A) Western blot analysis of soluble HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$ proteins expressed in the presence and absence (-) of different HIS_6-tagged MLPs in *E. coli* BL21(DE3)*ybdZ:aac(3)IV*. The analysis was conducted three times, and one representative set of blots is shown. (B) Quantification of the HIS_6-$TxtA^A$ (left) and HIS_6-$TxtB^A$ (right) protein band intensities following co-expression with different HIS_6-tagged MLPs. The bars represent the mean percent band intensity from triplicate western blots relative to the control (co-expression with HIS_6-TxtH; set to 100%) and was determined using ImageJ.

Error bars represent the standard deviation from the mean. Means with different letters (*a*, *b*, *c*, *d*) were determined to be significantly different ($P \leqslant 0.05$).

3.4.3 Influence of non-cognate MLPs on thaxtomin production in *S. scabiei*

In addition to examining the impact of the non-cognate MLPs on Txt NRPS A-domain solubility, we assessed their ability to promote the production of thaxtomin A in the absence of the native MLPs in *S. scabiei.* This was accomplished by overexpressing each MLP in a *S. scabiei* mutant that lacks all three endogenous MLP-encoding genes, including *txtH,* and is unable to produce thaxtomin (Li et al., 2019a). As reported previously, overexpression of *txtH* restores thaxtomin A production in the mutant, though not to levels observed in the wild-type strain (Figure 3.3A, B). This is due to polar effects of the *txtH* mutation on expression of the downstream *txtC* gene (Li et al., 2019a), which encodes the P450 monooxygenase that hydroxylates the thaxtomin backbone at the α- and/or ring carbon of the phenylalanine moiety (Alkhalaf et al., 2019; Healy et al., 2002). In addition, the thaxtomin B and D intermediates, which differ from thaxtomin A in the absence of one or both of the TxtC-dependent hydroxyl groups, were found to accumulate in the *S. scabiei* MLP triple mutant when *txtH* was overexpressed (Figure 3.3B), which is consistent with the observed polar effects of the *txtH* mutation on *txtC* gene expression. Therefore, in order to evaluate the efficiency of the different MLPs to exhibit functional redundancy with TxtH, the combined production of thaxtomins (thaxtomin A, B, D) was assessed in each of the MLP overexpression strains to account for any polar effects.

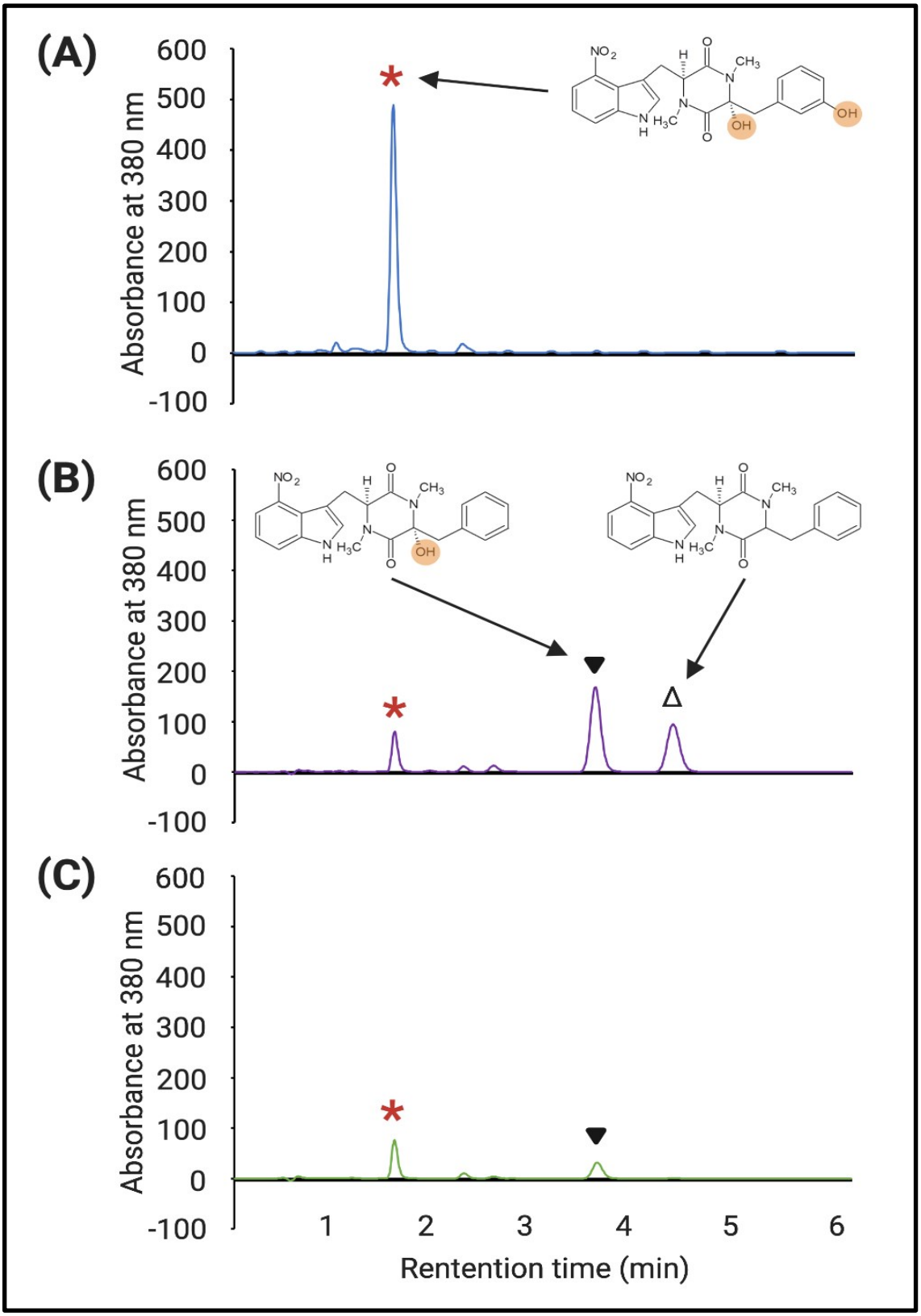

(A)
(B)
(C)
Absorbance at 380 nm
600
500
400
300
200
100
0
-100
1
2
3
4
5
6
Rentention time (min)
NO2
CH3
OH
H3C

Figure 3.3 HPLC analysis of culture extracts from wild-type *S. scabiei* 87.22 (A), the *S. scabiei* triple MLP deletion mutant ($\Delta mlp_{lipo}/\Delta txtH/\Delta mlp_{scab}$) containing the *txtH* expression vector (B) and the triple MLP deletion mutant containing the *MXAN_3118* expression vector (C). The peak corresponding to thaxtomin A (retention time=1.65 min) in each chromatogram is indicated with the red asterisks, and the peaks corresponding to thaxtomin B (retention time=3.81 min) and thaxtomin D (retention time=4.61 min) are indicated with ▼and Δ, respectively. The chemical structures of thaxtomin A, B and D are also shown next to the corresponding peaks, and the hydroxyl groups of thaxtomin A and B are highlighted.

As shown in Figure 3.4, all but two of the non-cognate MLPs were able to restore thaxtomin production in the MLP triple mutant to varying degrees. Overexpression of RHA1_ro04717 was most effective at restoring production to levels similar to that observed for TxtH, while overexpression of AWZ11_RS05060 and ComB restored production to levels similar to that observed for $MLP_{lipo,}$ a non-cognate MLP in *S. scabiei* that was previously shown to exhibit functional cross-talk with TxtH (Li et al., 2019a). The overexpression of CloY, MXAN_3118, CdaX, CchK, SCLAV_p1293, CGL27_RS02360 and PA2412 led to partial complementation of thaxtomin production, with levels ranging from 14-51% of that observed for TxtH (Figure 3.4). Among the MLPs tested, only YbdZ and CGL27_RS10110 were unable to restore detectable thaxtomin production when overexpressed in the MLP triple mutant. Interestingly, all three thaxtomins (thaxtomin A, B, D) were present in culture extracts of successfully complemented MLP strains with the exception of the MXAN_3118 overexpression strain, which did not accumulate detectable levels of thaxtomin D (Figure 3.3C). The reason for this is currently unclear, but it warrants further investigation.

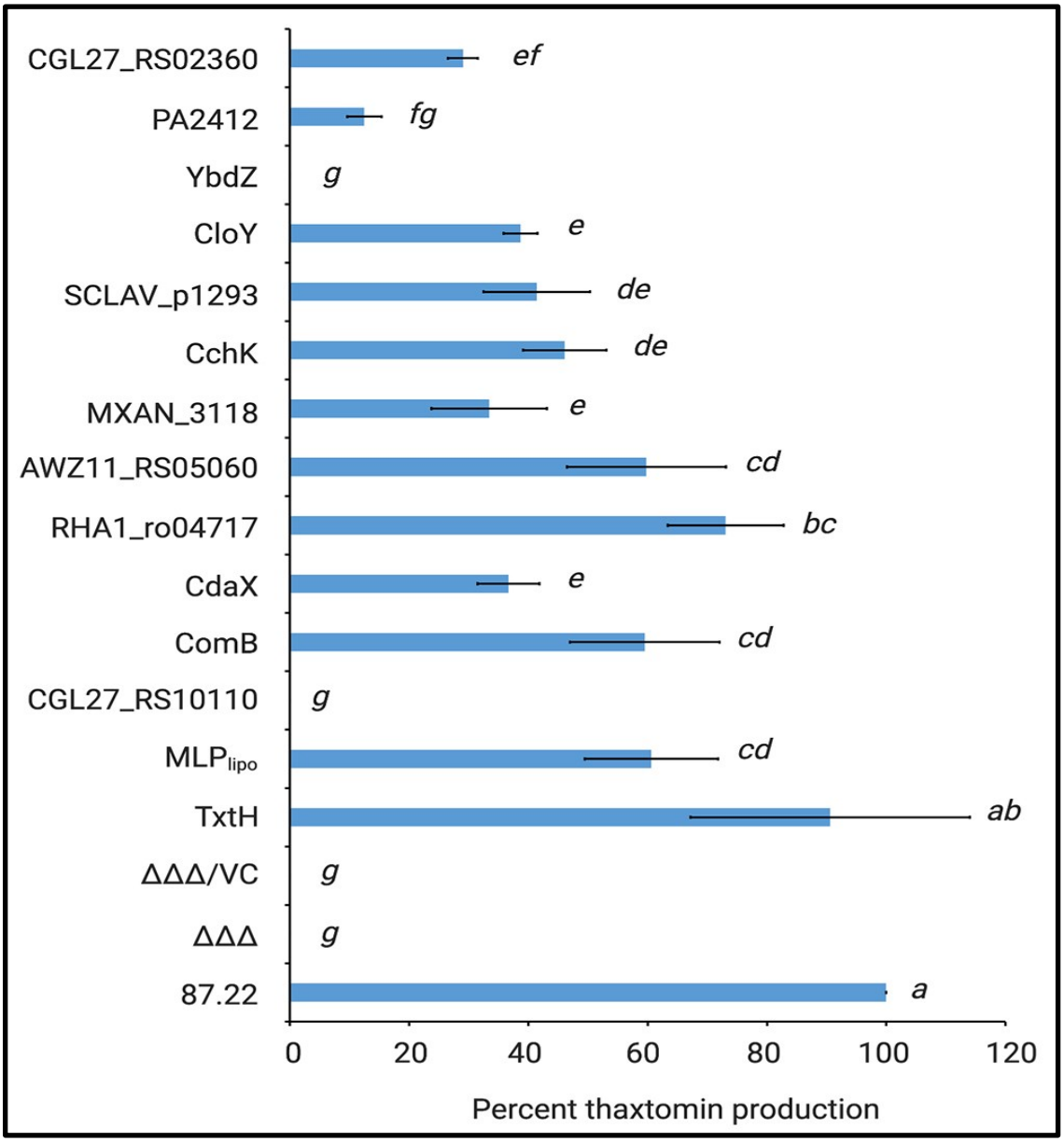


Figure 3.4 Relative quantification of thaxtomin production in the *S. scabiei* MLP triple mutant ($\Delta mlp_{lipo}/\Delta txtH/\Delta mlp_{scab}$; ΔΔΔ) expressing different non-cognate MLPs. The production levels are represented as the average % production of thaxtomins (thaxtomin A, thaxtomin B and thaxtomin D) relative to wild-type *S. scabiei* 87.22 (±SD). n=3 biological replicates for 87.22, ΔΔΔ, ΔΔΔ/VC (vector control); n=5 biological replicates for SCLAV_p1293; n=6 biological replicates for all other strains. Means with different letters (*a*, *b*, *c*, *d*, *e*, *f*, *g*) were determined to be significantly different ($P \leqslant 0.05$).

It is noteworthy that the results observed for CdaX and SCLAV_p1293 are contradictory to the results of our previous study, which found that overexpression of both genes failed to complement thaxtomin production in the *S. scabiei* MLP triple mutant (Li et al., 2019a). The reason behind this discrepancy is not clear, but it could be due to differences in the *Streptomyces* expression vectors that were used. In the current study, we used pRFSRL16, which harbors the *ermE*p* promoter as well as a Shine-Dalgarno (SD) sequence (AAAGGAGG) for expression of the cloned gene. In contrast, the expression vector used in our previous study (pRLDB50-1a) contains the *ermE*p* promoter but no SD sequence, and thus the native SD sequence was cloned along with the coding sequence of the gene to be expressed. As translation initiation is considered the rate limiting step of protein synthesis in bacteria, and there is evidence that the SD sequence and context play an important role in the initiation of translation of many mRNA transcripts (Guarlerzi and Pon, 2015), it is possible that the different expression vectors used in the current and previous study contributed to differences in levels of the CdaX and SCLAV_p1293 proteins produced in *S. scabiei,* though further investigations are required to verify this.

The results of the thaxtomin analysis together with the protein solubility assay are summarized in Figure 3.5. In general, the ability of an MLP to promote the soluble production of the Txt NRPS A-domains in *E. coli* corresponded with its ability to promote thaxtomin production in *S. scabiei.* In other words, only MLPs that enabled the soluble production of both A-domains, even in low amounts, were also found to promote the detectable production of thaxtomins. The ability of an MLP to serve as a functional partner had no relationship with amino acid similarity, since the two MLPs (YbdZ and CGL27_RS10110) that were unable to exhibit functional cross-talk with TxtH were just as

similar to TxtH as MLPs that could exhibit functional cross-talk (Table 3.3). A similar phenomenon was reported by Schomer and Thomas (2017), who found that the ability of non-cognate MLPs to compensate for the loss of YbdZ in *E. coli* did not correlate with the similarity of the MLP to YbdZ.

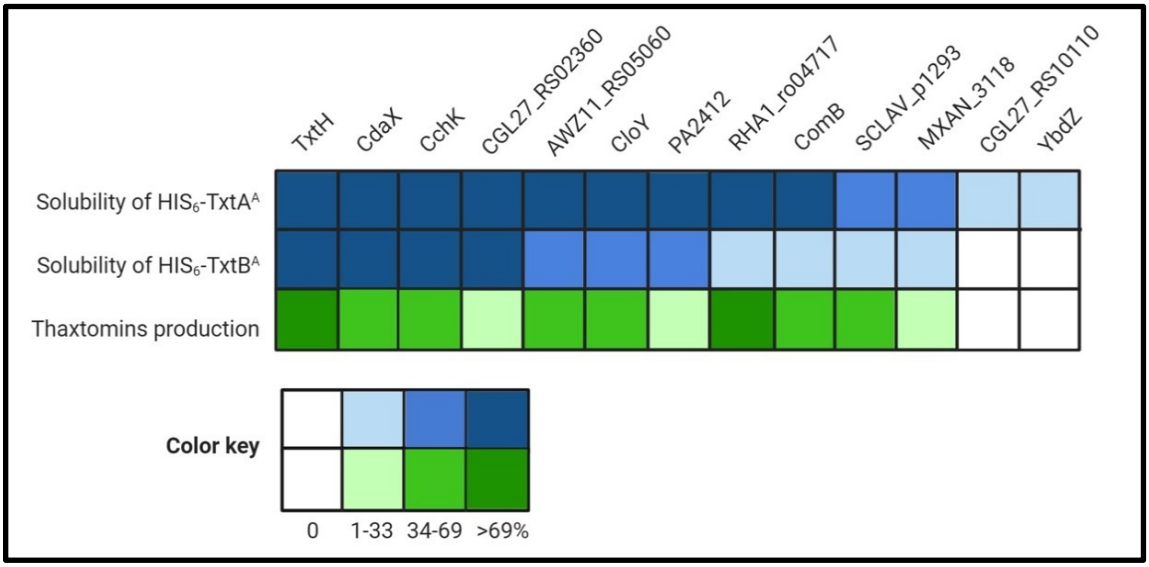


Figure 3.5 Summary of the results of the different assays examining the interaction between the thaxtomin NRPSs and the non-cognate MLPs. The heat map illustrates the relative amount of soluble HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$ proteins produced in the presence of the different MLPs as compared to TxtH (set to 100%), as well as the relative thaxtomin production levels in the presence of different MLPS as compared to TxtH (set to 100%).

It is notable that the relative efficiency of soluble protein production by an MLP did not appear to correlate with the relative efficiency of thaxtomin production in our study. For example, CdaX, CchK and CGL27_RS02360 were all able to promote the production of soluble protein for both of the Txt NRPS A-domains at levels similar to or great than that observed in the presence of TxtH, and yet none were able to fully complement thaxtomin production in the *S. scabiei* MLP triple mutant. Similarly, PA2412, CloY and

AWZ11_RS05060 exhibited somewhat comparable protein solubility profiles for both A-domains, but PA2412 was significantly less efficient at promoting thaxtomin production. PA2412 was also less efficient at promoting thaxtomin production than MXAN_3118, but it was more efficient at promoting the soluble production of both of the Txt A-domains than MXAN_3118. In addition, RHA1_ro04717 was the only non-cognate MLP that was able to fully complement thaxtomin production in *S. scabiei,* but it was much less efficient at promoting the soluble production of $TxtB^A$ compared to some other MLPs. While it is plausible that the solubility-promoting activity of some MLPs in our co-expression assay may have been influenced by the presence of the N-terminal HIS_6 tag, we previously showed that the HIS_6 tag does not impact this activity in the case of TxtH (Li et al., 2019a). Overall, our results suggest that the efficiency at which an MLP is able to promote NRPS A-domain solubility is not always a reliable indicator of the relative functionality of the MLP-NRPS pair *in vivo.* This may be due to effects of the MLP on the folding of the entire NRPS machinery that are not revealed when examining the individual A-domains alone. In addition, other studies have found that MLPs have a broader impact on NRPSs beyond protein solubility (Boll et al., 2011; Felnagle et al., 2010; Heemstra et al., 2009; Miller et al., 2016; Mori et al., 2018a; Schomer and Thomas, 2017; Schomer et al., 2018; Zhang et al., 2010). Schomer and Thomas (2017) showed that non-cognate MLPs can influence the solubility and catalysis of the EntF NRPS, including aminoacyl-*S*-PCP formation, and that these effects are separable. PA2412, for example, can enhance the catalysis of EntF but has no impact on EntF solubility, whereas two other non-cognate MLPs (CmnN, VioN) can enhance EntF solubility but do not influence catalysis. To date, we have been unable to detect the production of soluble Txt A-domain protein in the absence of TxtH, and so the

effect of TxtH or other MLPs on the adenylation or other activities of the TxtA and/or TxtB NRPS enzymes is currently unknown.

3.4.4 *In silico* analysis of the MLP-NRPS interface involved in thaxtomin biosynthesis

Although the degree of amino acid similarity between non-cognate MLPs and TxtH is unable to fully explain why some MLPs are capable of exhibiting functional cross-talk with TxtH while others are not, the overall topology of the Txt MLP-NRPS protein complex interface could provide some insights. Therefore, we utilized SWISS-MODEL to create *in silico* models for $TxtA^A$, $TxtB^A$ and TxtH using the structures of protein templates (Supplementary Table 3.3) that exhibited the best scores for GMQE (Global Model Quality Estimation) and QMEAN (Qualitative Model Energy Analysis) (Benkert et al., 2011; Waterhouse et al., 2018). Specifically, the structural models of $TxtA^A$ and $TxtB^A$ (Supplementary Figure 3.4) were computationally generated using the solved structure of the TioS NRPS from the thiocoraline biosynthetic pathway of *Micromonospora* sp. ML1 (PDB ID: 5wmm_1) as the template. The TioS NRPS requires its cognate MLP TioT for soluble production in *E. coli,* and the structure of the protein complex (PDB ID: 5wmm) revealed that TioT interacts with helix 10 and beta strands 18 and 19 from the A-domain of TioS (Mori et al., 2018b). TxtH was modeled using the crystal structure of the FscK MLP from *Thermobifida fusca* (PDB ID: 6ea3_1) as the template. The predicted TxtH structure is composed of three stranded anti-parallel beta sheets, one alpha helix and two single turn helices at its two termini (Supplementary Figure 3.4), which resembles the typical MLP monomers of solved structures (Drake et al., 2007; Miller et al., 2016; Tarry et al., 2017). During the modeling analysis, TxtH was docked with $TxtA^A$ or $TxtB^A$ based on the location

of TioT in the TioT-TioS complex (Figure 3.6A; Supplementary Figure 3.4). The predicted TxtH-TxtA^{A}/B^{A} interface is highly similar to that seen with other reported MLP-NRPS complexes (Mori et al., 2018b; Herbst et al., 2013; Miller et al., 2016; Tarry et al., 2017), where residues S23 and L24 of TxtH are predicted to hydrogen bond with A383 and A378 of TxtA^{A}, and with A410 and A405 of TxtB^{A} (Figure 3.6A). The same interaction is also observed in the adenylating enzyme SlgN1 from *Streptomyces lydicus*, which contains an MLP domain at its N-terminus (Herbst et al., 2013). Notably, the importance of residues S23 and L24 for the solubility-promoting activity of TxtH has been substantiated by site-directed mutagenesis (Li et al., 2019a).

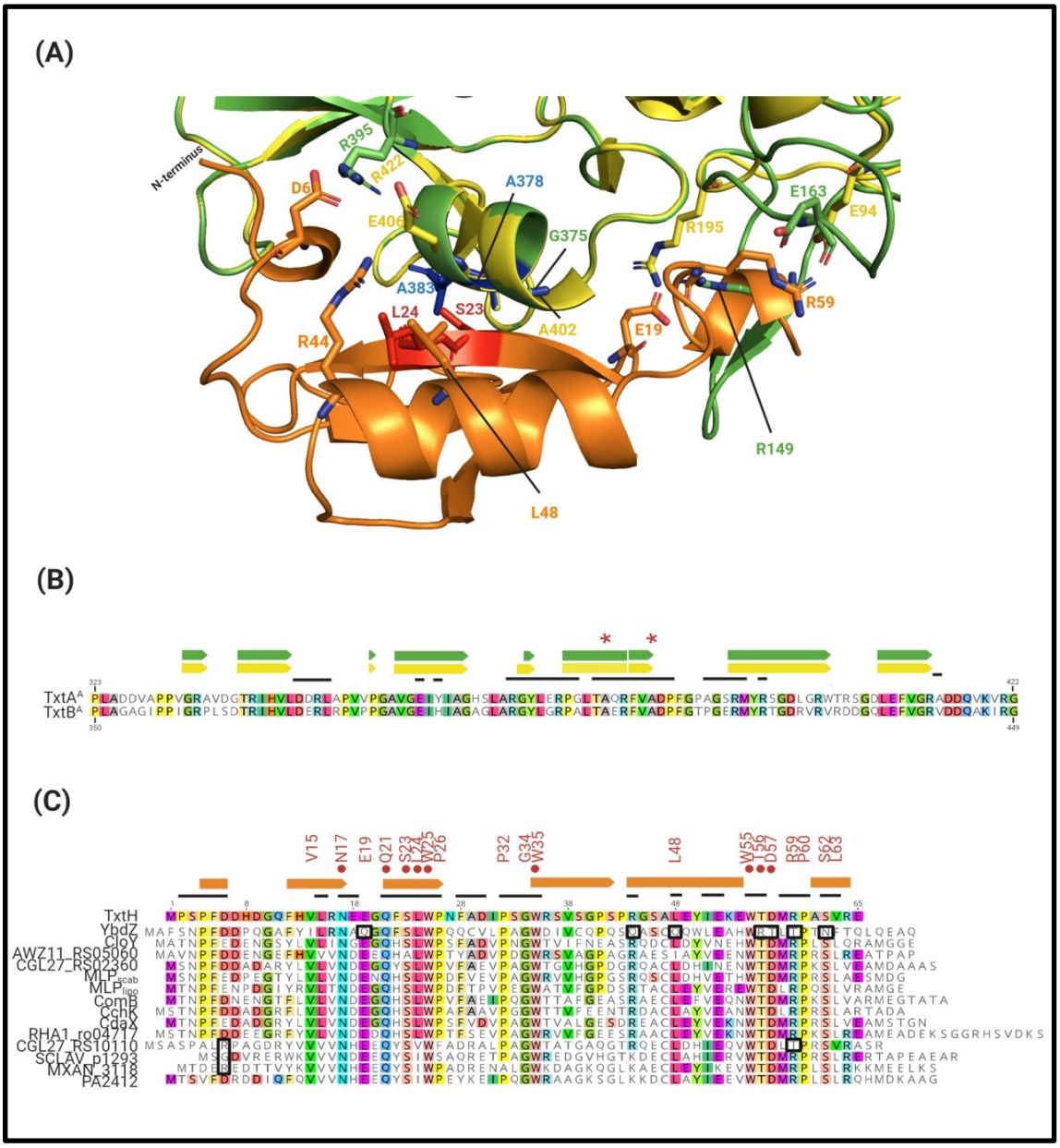


Figure 3.6 (A) Predicted interaction interface between the *S. scabiei* Txt A-domains and TxtH. $TxtA^A$ is shown in green, $TxtB^A$ is shown in yellow, and TxtH is shown in orange. The strictly conserved serine and leucine residues (red) of TxtH (S23 and L24) and two possible interacting alanine residues (blue) of $TxtA^A$ (A378 and A383) are highlighted. The corresponding alanine residues of $TxtB^A$ (A405 and A410) are not labeled. The residues that are associated with interaction interface are shown as sticks. (B) Partial amino acid sequence alignment of the *S. scabiei* $TxtA^A$ and $TxtB^A$. The residues involved in the formation of α-helix (box) and β-sheets (arrows) within the predicted structures are indicated above the alignment in green (for $TxtA^A$) and yellow (for $TxtB^A$). Residues in $TxtA^A$ and $TxtB^A$ that fall within the interaction interface between the TioS/FscH NRPSs

and their cognate MLPs (TioT/FscK) are indicated by the black lines above the amino acid alignment. The two alanine residues of $TxtA^A$ and $TxtB^A$ that are predicted to interact with S23 and L24 of TxtH are indicated by the astericks. (C) Amino acid sequence alignment of TxtH from *S. scabiei* and the non-cognate MLPs from other bacteria that were analyzed in this study. Residues within the non-cognate MLPs that match the amino acid residue in TxtH at the same position are coloured. The consensus sequence (VxxNxExQxSLWP-x5-PxGW-x12-L-x6-WTDxRPxSL) appearing in more than 85% of the 133 MLPs used in the phylogenetic analysis are indicated above the alignment. The residues shown to be important for the soluble production of $TxtA^A$ and/or $TxtB^A$ by TxtH (Li et al., 2019a) are indicated by the red circles. Variant residues in the non-cognate MLPs that may have a negative impact on the interaction with the thaxtomin NRPSs are highlighted in black boxes. Extracted secondary structures for TxtH are shown using orange boxes (helixes) and arrows (β-sheets) above the alignment. Residues in TxtH that fall within the interaction interface between TioT/FscK and their corresponding NRPSs (TioS/FscH) are indicated by the black lines above the amino acid alignment.

The predicted $TxtA^A$ and $TxtB^A$ structures display some differences, however, both models can make contacts with TxtH (Figure 3.6A; Supplementary Figure 3.4). The predicted TxtH binding interface region involves residues from helix 16 and beta sheets 19-22 of $TxtA^A$, and helix 15 and beta sheets 17-20 of $TxtB^A$ (Figure 3.6A, B). Several variable residues are present within the interface region of $TxtA^A$ and $TxtB^A$ (Figure 3.6B), suggesting that the two Txt NRPSs may interact differently with MLP partners, including TxtH. This is in line with results from the current study, where the solubility of $TxtB^A$ was impacted more than that of $TxtA^A$ by the MLP partner that it was co-expressed with (Figure 3.2A, B). In addition, our previous work showed that the solubility of $TxtB^A$ was affected to a greater extent than $TxtA^A$ during co-expression with various TxtH point mutants (Li et al., 2019a). Therefore, our results suggest that the formation of an MLP-NRPS functioning pair involves a more stringent interaction in the case of TxtB than it does for TxtA.

More detailed analysis of the amino acid sequences of the 133 MLP proteins used in the phylogenetic analysis (Figure 3.1) indicated the presence of a sequence/motif

(VxxNxExQxSLWP-x5-PxGW-x12-L-x6-WTDxRPxSL) in >85% of the proteins. The motif is similar to the signature sequence that was previously proposed by Baltz (2011) for predicting functional MLP homologues in sequenced genomes. FscK, the MLP whose structure (PDB ID: 6ea3_1) was used as a template to model TxtH, also contains all the residues from the motif. Many of these residues (except for V15, E19 and D57) are situated at its NRPS interacting interface, suggesting their importance for MLP functionality. In addition, the motif is well conserved in TxtH (except for L63) and in some of the non-cognate MLPs examined in the current study (Figure 3.6C), whereas other proteins display some variations. It is possible that differences in the sequence of this motif along with differences at other positions might impact the interaction of MLPs with one or both of the Txt A-domains. For instance, the positively-charged R44 residue of TxtH is predicted to form a salt bridge with E406 in $TxtB^{A}$ based on homology modelling using the MLP-NRPS structures of TioT-TioS or SlgN1 as template (Mori et al., 2018b; Herbst et al., 2013). In the case of YbdZ, the corresponding residue is an uncharged Q (Figure 3.6C), which is not expected to be involved in salt bridge formation and could potentially impact the YbdZ-$TxtB^{A}$ interaction. Therefore, the R→Q substitution in YbdZ might explain why this non-cognate MLP failed to promote the soluble expression of $TxtB^{A}$ (Figure 3.2). On the other hand, SCLAV_p1293, MXAN_3118 and PA2412 contain a positively charged K residue at the same position (Figure 3.6C) and promoted soluble $TxtB^{A}$ protein production, but not to the same extent as TxtH (Figure 3.2). It has been reported that RE salt bridges are more favorable for speeding up protein folding as compared to KE (Meuzelaar et al., 2016), but their relevance in the MLP-TxtA/B interaction requires further investigation.

Another potential interaction could involve the negatively-charged E19 residue of TxtH, which is predicted to be in close proximity to R149 in $TxtA^A$ and R195 in $TxtB^A$ (Figure 3.6A). The TxtH E19 residue is conserved in all of the non-cognate MLPs with the exception of YbdZ, which contains an uncharged Q at that position (Figure 3.6C), and could be another reason for the inability of YbdZ to promote Txt NRPS A-domain solubility (Figure 3.2A, B). In TxtH, D6 is predicted to contribute to hydrogen bonding and salt bridge formation with R395 in $TxtA^A$ and R422 in $TxtB^A$ to stabilize the MLP-NRPS interface (Figure 3.6A). A similar interaction is observed between D7 of TioT and R395 of TioS, between E6 of the MLP domain and R446 of the A-domain in SlgN1, as well as between D1324 of the MLP domain and R853 of the A-domain in ObiF1 (Mori et al., 2018b; Herbst et al., 2013; Kreitler et al., 2019). It should be noted that the corresponding negatively charged D or E residues are not present in either CGL27_RS10110, SCLAV_p1293 and MXAN_3118, which could in part explain why some of them failed or were not as efficient as TxtH in promoting Txt NRPS A-domain solubility or thaxtomin production (Figures 3.2 and 3.4).

In general, the C-terminal region of the conserved motif in YbdZ (Q-x6-WRTxTPxN) differs significantly from that present in TxtH (L-x6-WTDxRPxS) and other non-cognate MLPs (Figure 3.6C). In FscK and TioT, residues from this region (with the exception of D) are involved in binding with their cognate NRPS partners (Bruner and Zagulyaeva, unpublished; Mori et al., 2018b). In our model, the hydrophobic side chain of L48 from TxtH is closely packed with G375 in $TxtA^A$ and A402 in $TxtB^A$ towards the center of the interface, possibly contributing to nonpolar interactions (Figure 3.6A). The substitution of a polar Q residue at this position in YbdZ may further hinder its interaction

with the Txt A-domain proteins from the current study (Figure 3.6C). In addition, the solution structure of the Rv2377c MLP from *Mycobacteria tuberculosis* and of PA2412 from *P. aeruginosa* has demonstrated that the highly conserved WTDxRP portion of the motif is within an intrinsically disordered region in both proteins (Buchko et al., 2010). Disordered regions of proteins have been associated with functional diversity or with binding to multiple protein partners (Haynes et al., 2006; Xie et al., 2007). In our previous work, we showed that the WTD residues of TxtH are all important for promoting the solubility of $TxtA^A$ and $TxtB^A$ (Li et al., 2019a). The WTDxRP motif is absolutely conserved in all of the non-cognate MLPs examined in our studies except for YbdZ and CGL27_RS10110, both of which contain a T instead of the R residue (Figure 3.6C). R59 of TxtH is predicted to form a salt bridge with E163 of $TxtA^A$ and E94 of $TxtB^A$ (Figure 3.6A), and the substitution to an uncharged T may impact the ability of YbdZ and CGL27_RS10110 to bind efficiently to the A-domains, which could further explain they were not able to replace TxtH in the assays conducted (Figure 3.5). Overall, the structures of MLPs and their partners (including our *in silico* TxtH-$TxtA^A$/B^A models) provide important insights into the key residues that are involved in MLP/NRPS interactions and which may also account for the ability of MLPs from different biosynthetic pathways to exhibit functional redundancy. The question of why functional cross-talk occurs among different MLPs and its significance is one that remains to be addressed.

3.5 Conclusion

Here, we showed that phylogenetically distinct MLPs from different organisms vary in their ability to exhibit functional redundancy with TxtH from the thaxtomin biosynthetic pathway in *S. scabiei.* Except for YbdZ and CGL27_RS10110, all MLPs examined in this study were able to promote the soluble production of the Txt A-domains in *E. coli* and enabled thaxtomin production to varying degrees in a *S. scabiei* mutant lacking endogenous MLPs. *In silico* structural analysis of TxtH with its cognate NRPS A-domains revealed that the ability of different non-cognate MLPs to exhibit functional cross-talk with TxtH likely depends on the conservation of key residues at the MLP-NRPS interaction interface rather than the overall amino acid similarity shared between the proteins. In addition, the *in silico* analysis combined with our protein solubility assay results suggest that the two Txt NRPSs differ in their interactions with TxtH and with most of the non-cognate MLPs examined in this study. Overall, our study provides additional insights into the mechanism of MLP cross-talk and its impact on specialized metabolite biosynthesis in bacteria. Thaxtomin A is essential for common scab disease development by *S. scabiei* and other plant pathogenic *Streptomyces* spp., and thus our research on the thaxtomin biosynthetic machinery is expected to have useful applications for the development of strategies for effective disease management. Furthermore, the potent herbicidal activity exhibited by thaxtomin A (King et al., 2001) makes it an attractive bioherbicide for controlling the growth of weeds (Koivunen et al., 2013; Leep et al., 2010), and a better understanding of the thaxtomin biosynthetic pathway may facilitate the large-scale commercial production of this compound for agricultural applications. Currently, work is ongoing to determine whether TxtH and the non-cognate MLPs examined in this study can

influence the catalytic activity of either or both of the Txt NRPSs. In addition, the crystal structure of the TxtH-TxtA(B) complexes will be useful in better understanding the molecular basis for the interaction between TxtH and its two cognate NRPSs. Finally, the ability of TxtH and other non-cognate MLPs to influence the production of other NRPs in *S. scabiei* is the subject of on-going studies.

3.6 Acknowledgement

This work was supported by a Natural Sciences Engineering Research Council of Canada Discovery Grant (RGPIN-2018-05155) to D.R.D.B. and by a Natural Sciences Engineering Research Council of Canada Discovery Grant (RGPIN-2018–05949) to K.T. Y.L. was supported in part by the President's Doctoral Student Investment Fund at Memorial University of Newfoundland.

3.7 References

Alkhalaf, L.M., Barry, S.M., Rea, D., Gallo, A., Griffiths, D., Lewandowski, J.R., et al. (2018) Binding of distinct substrate conformations enables hydroxylation of remote sites in thaxtomin D by cytochrome P450 TxtC. *J Am Chem Soc.* 141, 216–222. doi: 10.1021/jacs.8b08864.

Al-Mestarihi, A.H., Villamizar, G., Fernández, J., Zolova, O. E., Lombó, F., and Garneau-Tsodikova, S. (2014) Adenylation and S-methylation of cysteine by the bifunctional enzyme TioN in thiocoraline biosynthesis. *J Am Chem Soc.* 136, 17350–17354. doi: 10.1021/ja510489j.

Baltz, R.H. (2011) Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J Ind Microbiol Biotechnol.* 38:1747. doi:10.1007/s10295-011-1022-8.

Benkert, P., Biasini, M., and Schwede, T. (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.* 27, 343–350. doi:10.1093/bioinformatics/btq662.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., et al. (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. 42, W252–W258. doi:10.1093/nar/gku340.

Boll, B., Taubitz, T., and Heide, L. (2011) Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. *J Biol Chem.* 286, 36281–36290. doi:10.1074/jbc.M111.288092.

Bown, L., and Bignell, D.R.D. (2017) Draft genome sequence of the plant pathogen *Streptomyces* sp. strain 11-1-2. *Genome Announc*. 5: e00968-17. doi:10.1128/genomeA.00968-17.

Buchko, G.W., Kim, C.Y., Terwilliger, T.C., and Myler, P.J. (2010) Solution structure of Rv2377c-founding member of the MbtH-like protein family. *Tuberculosis.* 90, 245–251. doi: 10.1016/j.tube.2010.04.002.

Chiu, H.T., Hubbard, B.K., Shah, A.N., Eide, J., Fredenburg, R.A., Walsh, C.T., et al. (2001) Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proc Nat. Acad Sci USA.* 98, 8548–8553. doi:10.1073/pnas.151246498.

Davidsen, J.M., Bartley, D.M., and Townsend, C.A. (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. doi: 10.1021/ja307710d.

DeLano, W.L. (2002) Pymol: An open-source molecular graphics tool. *CCP4 Newsl Protein Crystallogr.* 40, 82-92.

Drake, E.J., Cao, J., Qu, J., Shah, M.B., Straubinger, R.M., and Gulick, A.M. (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. doi:10.1074/jbc.M611833200.

Esquilín-Lebrón, K.J., Boynton, T.O., Shimkets, L.J., and Thomas, M.G. (2018) An Orphan MbtH-Like Protein Interacts with Multiple Nonribosomal Peptide Synthetases in *Myxococcus Xanthus* DK1622. *J Bacteriol.* 200, e00346-18. doi: 10.1128/JB.00346-18.

Felnagle, E.A., Barkei, J.J., Park, H., Podevels, A.M., McMahon, M.D., Drott, D.W., et al. (2010) MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry*. 49, 8815–8817. doi:10.1021/bi1012854.

Finking, R., and Marahiel, M.A. (2004) Biosynthesis of Nonribosomal Peptides. *Annu Rev Microbiol.* 58, 453-488. doi: 10.1146/annurev.micro.58.030603.123615.

Fyans, J.K., Bown, L., and Bignell, D.R.D. (2016) Isolation and characterization of plant-pathogenic *Streptomyces* species associated with common scab-infected potato tubers in Newfoundland. *Phytopathology.* 106, 123–131. doi:10.1094/PHYTO-05-15-0125-R.

Gualerzi, C.O. and Pon, C.L. (2015) Initiation of mRNA translation in bacteria: structural and dynamic aspects. *Cell Mol Life Sci*. 72, 4341–4367. doi:10.1007/s00018-015-2010-3.

Haynes, C., Oldfield, C.J., Ji, F., Klitgord, N., Cusick, M.E., Radivojac, P., et al. (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLOS Comput. Biol.* 2: e100. doi:10.1371/journal.pcbi.0020100.

Healy, F.G., Krasnoff, S.B., Wach, M., Gibson, D.M., and Loria, R. (2002) Involvement of a cytochrome P450 monooxygenase in thaxtomin A biosynthesis by *Streptomyces acidiscabies*. *J Bacteriol.* 184, 2019–2029. doi:10.1128/JB.184.7.2019-2029.2002.

Heemstra, J.R., Walsh, C.T., and Sattely, E.S. (2009) Enzymatic tailoring of ornithine in the biosynthesis of the Rhizobium cyclic trihydroxamate siderophore vicibactin. *J Am Chem Soc.* 131, 15317–15329. doi: 10.1021/ja9056008.

Herbst, D.A., Boll, B., Zocher, G., Stehle, T., and 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. doi:10.1074/jbc.M112.420182.

Huguet-Tapia, J.C., Lefebure, T., Badger, J.H., Guan, D., Pettis, G.S., Stanhope, M.J., et al. (2016) Genome content and phylogenomics reveal both ancestral and lateral evolutionary pathways in plant-pathogenic *Streptomyces* species. *Appl Environ Microbiol.* 82, e3504–e3515. doi: 10.1128/AEM.03504-15.

Hur, G.H., Vickery, C.R., and Burkart, M.D. (2012) Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep.* 29, 1074–1098. doi:10.1039/c2np20025b.

Imker, H.J., Krahn, D., Clerc, J., Kaiser, M., and Walsh, C.T. (2010) *N*-acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF. *Chem Biol.* 17, 1077–1083. doi: 10.1016/j.chembiol.2010.08.007

Joshi, M.V., Mann, S.G., Antelmann, H., Widdick, D.A., Fyans, J.K., Chandra, G., et al. (2010) The twin arginine protein transport pathway exports multiple virulence proteins in the plant pathogen *Streptomyces scabies*. *Mol Microbiol.* 77, 252–271. doi:10.1111/j.1365-2958.2010.07206.x.

Kaniusaite, M., Tailhades, J., Kittilä, T., Fage, C.D., Goode, R.J., Schittenhelm, R.B., et al. (2020) Understanding the early stages of peptide formation during the biosynthesis of teicoplanin and related glycopeptide antibiotics. *FEBS J.* doi: 10.1111/febs.15350. [Epub ahead of print].

Kieser, T, Bibb, M.J., Buttner, M.J., and Chater, K.F. (2000) Practical *Streptomyces* Genetics. 2nd ed. Norwich: John Innes Foundation.

King, R., Lawrence, C., and Gray, J. (2001) Herbicidal properties of the thaxtomin group of phytotoxins. *J Agric Food Chem.* 49, 2298–2301. doi: 10.1021/jf0012998.

Koivunen, M., and Marrone, P. (2013) *Uses of Thaxtomin and Thaxtomin Compositions as Herbicides.* U.S. Patent No. US8476195 B2.

Kreitler, D.F., Gemmell, E.M., Schaffer, J.E., Wencewicz, T.A., and Gulick, A.M. (2019) The structural basis of *N*-acyl-α-amino-β-lactone formation catalyzed by a

nonribosomal peptide synthetase. *Nat Commun.* 10, 1-13. doi: 10.1038/s41467-019-11383-7.

Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* 372, 774–797. doi: 10.1016/j.jmb.2007.05.022.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096.

Lautru, S., Oves-Costales, D., Pernodet, J.L., and Challis, G.L. (2007) MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. *Microbiology.* 153, 1405–1412. doi:10.1099/mic.0.2006/003145-0.

Leep D., Doricchi, L., Perez Baz, M.J., Millan, F.R., and Fernandez Chimeno, R.I. (2010) *Use of thaxtomin for selective control of rice and aquatic based weeds.* WO Patent No 2010/121079 A3.

Letunic, I., and Bork, P. (2007) Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics.* 23, 127–128. doi:10.1093/bioinformatics/btl529.

Li, Y., Liu, J., Adekunle, D., Bown, L., Tahlan, K., and Bignell, D.R.D. (2019a) TxtH is a key component of the thaxtomin biosynthetic machinery in the potato common scab pathogen *Streptomyces scabies*. *Mol Plant Pathol.* 20, 1379-1393. doi:10.1111/mpp.12843.

Li, Y., Liu, J., Díaz-Cruz, G., Cheng, Z., and Bignell, D.R.D. (2019b) Virulence mechanisms of plant-pathogenic *Streptomyces* species: An updated review. *Microbiology.* 165, 1025-1040. doi:10.1099/mic.0.000818.

Loria, R., Bukhalid, R.A., Creath, R.A., Leiner, R.H., Olivier, M. and Steffens, J.C. (1995) Differential production of thaxtomins by pathogenic *Streptomyces* species *in vitro*. *Phytopathology.* 85, 537-541. doi: 10.1094/Phyto-85-537.

Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem Rev.* 97, 2651–2674. doi:10.1021/cr960029e.

Martínková, L., Uhnáková, B., Pátek, M., Nešvera, J., and Křen, V. (2009) Biodegradation potential of the genus *Rhodococcus. Environ Int.* 35, 162–177. doi:10.1016/j.envint.2008.07.018.

McMahon, M.D., Rush, J.S., and Thomas, M.G. (2012) Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *J Bacteriol.* 194: e088-12. doi:10.1128/JB.00088-12.

Meuzelaar, H., Vreede, J., and Woutersen, S. (2016) Influence of Glu/Arg, Asp/Arg, and Glu/Lys salt bridges on α-helical stability and folding kinetics. *Biophys J.* 110, 2328–2341. doi: 10.1016/j.bpj.2016.04.015

Miller, B.R., Drake, E.J., Shi, C., Aldrich, C.C., and Gulick, A.M. (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. doi:10.1074/jbc.M116.746297.

Mori, S., Green, K.D., Choi, R., Buchko, G.W., Fried, M.G., and Garneau-Tsodikova, S. (2018a) Using MbtH-Like Proteins to Alter the Substrate Profile of a Nonribosomal Peptide Adenylation Enzyme. *ChemBioChem.* 19, 2186-2194. doi:10.1002/cbic.201800240.

Mori, S., Pang, A.H., Lundy, T.A., Garzan, A., Tsodikov, O.V., and Garneau-Tsodikova, S. (2018b) Structural basis for backbone *N*-methylation by an interrupted adenylation domain. *Nat Chem Biol.* 14, 428–430. doi:10.1038/s41589-018-0014-7.

Pojer, F., Li, S.M., and Heide, L. (2002) Molecular cloning and sequence analysis of the clorobiocin biosynthetic gene cluster: New insights into the biosynthesis of aminocoumarin antibiotics. *Microbiology.* 148, 3901–3911. doi:10.1099/00221287-148-12-3901.

Quadri, L.E.N., Sello, J., Keating, T.A., Weinreb, P.H., and Walsh, C.T. (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. doi:10.1016/S1074-5521(98)90291-5.

Sambrook, J.F., and Russell, D.W. (2001) Molecular cloning: a laboratory manual (3-volume set). *Cold Spring Harb Perspect Med.* doi:10.3724/SP.J.1141.2012.01075.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods.* 9, 671–675. doi:10.1038/nmeth.2089.

Schomer, R.A., Park, H., Barkei, J.J., and Thomas, M.G. (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. doi:10.1021/acs.biochem.8b00552.

Schomer, R.A., and Thomas, M.G. (2017) Characterization of the functional variance in MbtH-like protein interactions with a nonribosomal peptide synthetase. *Biochemistry*. 56, 5380–5390. doi:10.1021/acs.biochem.7b00517.

Strieker, M., Tanović, A., and Marahiel, M.A. (2010) Nonribosomal peptide synthetases: Structures and dynamics. *Curr Opin Struct Biol*. 20, 234–240. doi:10.1016/j.sbi.2010.01.009.

Süssmuth, R.D., and Mainz, A. (2017) Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew Chem*. 56, 3770–3821. doi:10.1002/anie.201609079.

Tarry, M.J., Haque, A.S., Bui, K.H., and Schmeing, T.M. (2017) X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure*. 25, 89–96. doi:10.1016/j.str.2017.03.014.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018) SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res Spec Publ.* 46:gky427. doi:10.1093/nar/gky427.

Whelan, S., and Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 18, 691–699. doi: 10.1093/oxfordjournals.molbev.a003851.

Wolpert, M., Gust, B., Kammerer, B., and Heide, L. (2007) Effects of deletions of mbtH-like genes on clorobiocin biosythesis in *Streptomyces coelicolor*. *Microbiology*. 153, 1413-1423. doi:10.1099/mic.0.2006/002998-0.

Xie, H., Vucetic, S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Obradovic, Z., et al. (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *Proteome Res*. 6, 1917–1932. doi:10.1021/pr060394e.

Zhang, Y., Bignell, D.R., Zuo, R., Fan, Q., Huguet-Tapia, J.C., Ding, Y., et al. (2016) Promiscuous Pathogenicity Islands and Phylogeny of Pathogenic *Streptomyces* spp. *Mol Plant Microbe Interact*. 29, 640–650. doi: 10.1094/MPMI-04-16-0068-R.

Zhang, W., Heemstra, J.R., Walsh, C.T., and Imker, H.J. (2010) Activation of the pacidamycin pacl adenylation domain by MbtH-like proteins. *Biochemistry*. 49, 9946–9947. doi:10.1021/bi101539b.

Zolova, O.E., and Garneau-Tsodikova, S. (2012) Importance of the MbtH-like protein TioT for production and activation of the thiocoraline adenylation domain of TioK. *MedChemComm*. 3, 950–955. doi:10.1039/c2md20131c.

Zolova, O.E., and Garneau-Tsodikova, S. (2014) KtzJ-dependent serine activation and O-methylation by KtzH for kutznerides biosynthesis. *J Antibiot*. 67: ja.2013.98. doi: 10.1038/ja.2013.98.

Zwahlen, R.D., Pohl, C., Bovenberg, R. A.L., and Driessen, A.J.M. (2019) Bacterial MbtH-like Proteins Stimulate Nonribosomal Peptide Synthetase-Derived Secondary Metabolism in Filamentous Fungi. *ACS Synth Biol*. 8, 1776–1787. doi:10.1021/acssynbio.9b00106.

3.8 Supplementary Information

Supplementary Table 3.1 Oligonucleotide primers used in this study.

Primer	Sequence (5′ - 3′)†	Use
PL169	GCGC**CATATG**ACCAATCCGTTCGAAGACGC	Forward primer for construction of pET28b/HIS$_6$-*cdaX*
PL170	GCGC**GAATTC**TCAGTTGCCGGTGCTCATCG	Reverse primer for construction of pET28b/HIS$_6$-*cdaX*
PL183	GCGC**CATATG**AGCACCAACCCCTTCGACGA	Forward primer for construction of pET28b/HIS$_6$-*cchK*
PL184	GCGC**GAATTC**TCAGGCGTCCGCGGTCCGGG	Reverse primer for construction of pET28b/HIS$_6$-*cchK*
PL167	GCGC**CATATG**AGCGGCGATGTGCGGGAGCG	Forward primer for construction of pET28b/HIS$_6$-*SCLAV_p1293*
PL168	GCGC**GAATTC**TCACCGGGCCTCCGCCTCCG	Reverse primer for construction of pET28b/HIS$_6$-*SCLAV_p1293*
PL175	GCGC**CATATG**AGCGCCTCACCCGCCCTGCG	Forward primer for construction of pET28b/HIS$_6$-*CGL27_RS10110*
PL176	GCGC**GAATTC**TCATCGTGAGGCTCGTACGGA	Reverse primer for construction of pET28b/HIS$_6$-*CGL27_RS10110*
PL177	GCGC**CATATG**AGCAACCCCTTCGACGACGC	Forward primer for construction of pET28b/HIS$_6$-*CGL27_RS02360*
PL178	GCGC**GAATTC**TCAGGAGGCGGCCGCGTCCA	Reverse primer for construction of pET28b/HIS$_6$-*CGL27_RS02360*
PL173	GCGC**CATATG**GCAGTGAACCCGTTCGACGA	Forward primer for construction of pET28b/HIS$_6$-*AWZ11_RS05060*
PL174	GCGC**GAATTC**TCAGGGCGCGGGGGTCGCCT	Reverse primer for construction of pET28b/HIS$_6$-*AWZ11_RS05060*
PL185	GCGC**CATATG**ACCGACGAACGGGAGGACAC	Forward primer for construction of pET28b/HIS$_6$-MXAN_3118
PL186	GCGC**GAATTC**CTAGCTCTTGAGTTCTTCCA	Reverse primer for construction of pET28b/HIS$_6$-MXAN_3118
PL187	GCGC**CATATG**TCCACCAACCCCTTCGACGA	Forward primer for construction of pET28b/HIS$_6$-RHA1_ro04717

PL188	GCGC**GAATTC**TCAGCTCTTGTCGACGCTGT	Reverse primer for construction of pET28b/HIS$_6$-RHA1_ro04717
PL189	GCGC**CATATG**GCGACGAACCCGTTCGAGGA	Forward primer for construction of pET28b/HIS$_6$-*cloY*
PL190	GCGC**GAATTC**CTACTCGCCACCCATCGCCC	Reverse primer for construction of pET28b/HIS$_6$-*cloY*
PL191	GCGC**CATATG**ACTAACCCTTTCGACAACGA	Forward primer for construction of pET28b/HIS$_6$-*comB*
PL192	GCGC**GGATCC**TCAGGCCGTGGCGGTGCCCT	Reverse primer for construction of pET28b/HIS$_6$-*comB*
PL193	GCGC**CATATG**ACTTCAGTGTTCGACCGTGA	Forward primer for construction of pET28b/HIS$_6$-*PA2412*
PL194	GCGC**GAATTC**TCAGCCGGCCGCCTTGTCCA	Reverse primer for construction of pET28b/HIS$_6$-*PA2412*
PL208	GCGC**CATATG**GCCTTCTCCAACCCCTTCGA	Forward primer for construction of pET28b/HIS$_6$-*ybdZ*
PL209	ATAT**GCGGCCGC**TCACTGCGCTTCCTGGAGCT	Reverse primer for construction of pET28b/HIS$_6$-*ybdZ*
PL35	GCGC**CATATG**CCCTCACCCTTCGACGAC	Forward primer for construction of pRFSRL16/*txtH*
PL195	ATAT**GCGGCCGC**TCATTCACGGACGGACGCCG	Reverse primer for construction of pRFSRL16/*txtH*
PL163	GCGC**CATATG**ACCAACCCCTTCGAGAAC	Forward primer for construction of pRFSRL16/*mlp$_{lipo}$*
PL196	ATAT**GCGGCCGC**TCACTCGCCCATGGCCCGGA	Reverse primer for construction of pRFSRL16/*mlp$_{lipo}$*
PL169	GCGC**CATATG**ACCAATCCGTTCGAAGACGC	Forward primer for construction of pRFSRL16/*cdaX*
PL198	ATAT**GCGGCCGC**TCAGTTGCCGGTGCTCATCG	Reverse primer for construction of pRFSRL16/*cdaX*
PL167	GCGC**CATATG**AGCGGCGATGTGCGGGAGCG	Forward primer for construction of pRFSRL16/*SCLAV_p1293*
PL199	ATAT**GCGGCCGC**TCACCGGGCCTCCGCCTCCG	Reverse primer for construction of pRFSRL16/*SCLAV_p1293*
PL175	GCGC**CATATG**AGCGCCTCACCCGCCCTGCG	Forward primer for construction of pRFSRL16/*CGL27_RS10110*
PL200	ATAT**GCGGCCGC**TCATCGTGAGGCTCGTACGGA	Reverse primer for construction of pRFSRL16/*CGL27_RS10110*
PL177	GCGC**CATATG**AGCAACCCCTTCGACGACGC	Forward primer for construction of pRFSRL16/*CGL27_RS02360*
PL214	ATAT**GCGGCCGC**TCAGGAGGCTGCCGCGTCCATGGCCTCGAC	Reverse primer for construction of pRFSRL16/*CGL27_RS02360*

PL173	GCGC**CAT**ATGGCAGTGAACCCG TTCGACGA	Forward primer for construction of pRFSRL16/*AWZ11_RS05060*
PL201	ATAT**GCGGCCGC**TCAGGGCGCG GGGGTCGCCT	Reverse primer for construction of pRFSRL16/*AWZ11_RS05060*
PL183	GCGC**CAT**ATGAGCACCAACCCC TTCGACGA	Forward primer for construction of pRFSRL16/*cchK*
PL202	ATAT**GCGGCCGC**TCAGGCGTCC GCGGTCCGGG	Reverse primer for construction of pRFSRL16/*cchK*
PL185	GCGC**CAT**ATGACCGACGAACGG GAGGACAC	Forward primer for construction of pRFSRL16/*MXAN3118*
PL203	ATAT**GCGGCCGC**CTAGCTCTTG AGTTCTTCCA	Reverse primer for construction of pRFSRL16/*MXAN3118*
PL187	GCGC**CAT**ATGTCCACCAACCCC TTCGACGA	Forward primer for construction of pRFSRL16/*RHA1_ro04717*
PL204	ATAT**GCGGCCGC**TCAGCTCTTG TCGACGCTGT	Reverse primer for construction of pRFSRL16/*RHA1_ro04717*
PL189	GCGC**CAT**ATGGCGACGAACCCG TTCGAGGA	Forward primer for construction of pRFSRL16/*cloY*
PL205	ATAT**GCGGCCGC**CTACTCGCCA CCCATCGCCC	Reverse primer for construction of pRFSRL16/*cloY*
PL191	GCGC**CAT**ATGACTAACCCTTTC GACAACGA	Forward primer for construction of pRFSRL16/*comB*
PL206	ATAT**GCGGCCGC**TCAGGCCGTG GCGGTGCCCT	Reverse primer for construction of pRFSRL16/*comB*
PL193	GCGC**CAT**ATGACTTCAGTGTTC GACCGTGA	Forward primer for construction of pRFSRL16/*PA2412*
PL207	ATAT**GCGGCCGC**TCAGCCGGCC GCCTTGTCCA	Reverse primer for construction of pRFSRL16/*PA2412*
PL208	GCGC**CAT**ATGGCCTTCTCCAAC CCCTTCGA	Forward primer for construction of pRFSRL16/*ybdZ*
PL209	ATAT**GCGGCCGC**TCACTGCGCT TCCTGGAGCT	Reverse primer for construction of pRFSRL16/*ybdZ*

† Non-homologous extensions are underlined, while engineered restriction sites are indicated in bold.

Supplementary Table 3.2 MLPs used in the phylogenetic analysis.

Bacterial species	Protein accession number	Locus tag(s) and/or protein name	Amino acid sequence
Acidobacteria bacterium isolate ATN1	RIK00317.1	DCC47_22460	mstnpfddedgrfyvlmndeeqyslwptfsevpqgwrvvfgeesraacveyveknwtdmrpkslrdameadekarnaga
Actinophytocola xanthii strain 11-183	WP_075124157.1	BU204_RS04080	mtnpfedpegtfhvlvndeeqhslwpsfaqvpagwrsvhgpagresclayveehwtdlrprslrermss
Actinophytocola xanthii strain 11-183	WP_075128341.1	BU204_RS25790	msnpfdqedgtflvlrneedqyslwpefadvppgwvlvhgpdtrtscldyvdrewtdmrprslvvamegrseeagke
Actinoplanes sp. N902-109	WP_015619633.1	L083_RS07715	mssnpfddengtfhvlvndeeqhslwpsfkeipsgwrsvfgpaarqealdyvdanwtdlrpkslrdsmaq
Actinoplanes sp. TFC3	WP_067497195.1	TFC3_RS07255	mssnpfddengtfhvlvndeeqhslwpsfkeipsgwrsvfgpaprqqaldyvdqnwtdlrpkslrdsmaq
Actinoplanes teichomyceticus ATCC 31121	CAE53354.1	Tcp13	mtnpfdnedgsflvlvngegqhslwpafaevpdgwtgvhgpasrqdclgyveqnwtdlrpkslisqisd
Actinoplanes teichomyceticus ATCC 31121	CAE53358.1	Tcp17	mtnpfdnedgsflvlvngegqhslwpafaevpdgwtgvhgpasrqdclgyveqnwtdlrprslveqada
Agrobacterium fabrum str. C58	WP_010973243.1	Atu3678	mssqtpaedlhynvvisdeerysiwpvykavpagwrlsgfsgskqacldhievewtdmrplslrrlmdgeaanitsaqe
Amycolatopsis orientalis strain B-37	WP_037306096.1	SD37_RS16720	mpnpfedpdakylvlvndegqhslwpvfadvpagwksvfgesgrqecldyieknwtdmrpkslieamektapas
Amycolatopsis orientalis strain KFCC10990P	KM232637.1	Vcm11	mtnpfdnedgsffvlvndegqhslwpafaevpagwttvhgeagrkeclayveenwtdlrpksliqeaga
Athrobacter chlorophenolicus A6	CP001341.1	Achl_1745	mtnpfddksatfsvlvneyqqhslwpafaavpegwvtmfgpdnreacldyvsrtwtdmaprkvselaasq
Bacillus cereus JRS1	CYHI01000652.1	BN2127_JRS1_06966	manpfenadgtylvlineegqyslwpgfidvpsgwtvvheqkgreacldyiqshwsdmrpnslkpvenv
Bacillus massiliglaciei strain Marseille-P2600	WP_110927797.1	BQ4305_RS10860	mtnpfenedslflvlmneegqyslwpafldvpagwvkkfgqssrvlcqqyiesnwrdmrpasikeelaasnk
Bacillus subtilis strain RC 25	WP_021480512.1	B2G85_RS15655	manpfenadgtylvlineegqyslwpsfidvpsgwtvvheqkgreacldyiqshwsdmrpnslktvenv
Bacillus subtilis subsp. subtilis strain 168	NC_000964.3	BSU_31959	manpfenadgtylvlvneegqyslwpgfidvpsgwtvvheqkgreacldyiqshwsdmrpnslktvenv
Burkholderia insecticola sp. RPE64	WP_044043681.1	BRPE64_RS25685	mtqqqnlaaddlvytvvineeeqfsiwptfrdvpagwreagvrgpkaeclayiektwtdmrpaslrrhmdavsaqakarln

Burkholderia plantarii PG1	WP_042624918.1	BGL_RS09350	mpnpfdndaaeftvlrnaegqhslwpafaaqpagwdrrhgpasradclrfvethwtdirplsgqrdaagatargdnprsr
Caballeronia peredens isolate LMG 29314	WP_087124204.1	AWB71_RS11890	mtqqqnlaaddlvytvvineeeqfsiwptfrdvpagwreagvrgpkadclayiektwtdmrpaslrrhmdavnaqakarln
Caballeronia temeraria isolate LMG 29319	WP_061164178.1	AWB76_RS32305	mnpfddetgeffvlrndegqhslwpsfaavpagwtsvfgvaarqacidyinenwtdirpaslrreap
Chelatococcus sp. CO-6	WP_019401813.1	AL346_RS04810	manpfddedgvflvlvndegqhslwpsfadvpagwqtvhgpaarqecldyvsahwtdmrprsliaaeet
Crocosphaera watsonii WH 0401	WP_007303638.1	CWATWH0401_RS03425	mkqderedttnyrvvvnheeqysiwpdyrdipagwrdtgksgqkedcleyikevwtdmrplslrkkme
Dyella sp. 4M-K27	WP_126687008.1	EKH80_RS22290	mtnpfddsngtffvlvnheeqyslwpefaqipagwtvkfgpdkrqecldyveqnwvdmrprslieameaera
Escherichia coli BL21(DE3)	ACT42431.1	ECD_00552, YbdZ	mafsnpfddpqgafyilrnaqgqfslwpqqcvlpagwdivcqpqsqascqqwleahwrtltptnftqlqeaq
Frankia sp. R43	WP_054567392.1	ACG83_RS17995	mtgspfddengqflallndegqfslwplfaqvpagwravhgpesrqacldyieaqwtdmrpasliertrrdrpggapg
Gordonia alkanivorans NBRC 16433	WP_006360234.1	GOALK_RS17825	mtnpfddedgrfyvlvndenqhslwptfadipagwtkvfgedsraacleyveqnwtdlrpkslieameadksgda
Gordonia amicalis CCMA-559	WP_024497698.1	BMSG_RS0103965	mtnpfddedgrfyvlvndenqhslwptfadipagwtkvfgedsraacleyveqnwtdlrpkslieameadkgags
Gordonia neofelifaecis NRRL B-59395	WP_009679141.1	SCNU_RS09450	mtnpfddengrfyvlvneenqhslwptfadipagwtkvfgeesreacleyveknwtdlrpqslidamaadqkna
Gordonia polyisoprenivorans HW436 A3OC	WP_006370451.1	A3OC_RS0105575	mtnpfddengrfyvlvndenqhslwptfadipagwtkvfgedsraacleyveknwtdlrpkslidameadkaareng
Gordonia rhizosphera NBRC 16068	WP_006332355.1	GORHZ_RS09360	mtnpfddengrfyvlvndenqhslwptfadipagwttvfgedsraacleyveknwtdlrpkslieameadaaepeaq
Gordonia terrae strain 3612	WP_004020535.1	BCM27_RS19375	mtnpfddedgrfyvlvndenqhslwptfadipagwtkvfgedsraacleyveqnwtdlrpkslieameadkssda
Kitasatospora sp. MY 5-36	WP_049649155.1	AE652_RS01190	mtnpfddqdgtflvlvneenqhslwpqfadvpegwtvvhgpdtnaacleyvekswtdmrprsladamdtrk
Kroppenstedtia sanguinis strain X0209	WP_124248777.1	D1G38_RS17330	mtnpfddedgrfyvlvndenqhslwptfaeipagwtkvfgedsrqaclayveenwtdlrpkslieameadadsk

Kutzneria sp. 744	ABV56590.1	KtzJ	msanpfddedgqfqvlvndedqhslwpafapvpdgwrvvfgadrrdrclayveqnwtdmrpkslreamaad
Limnoraphis robusta CS-951	WP_046276759.1	WN50_RS13840	mnseedttiyrvvineeeqysiwpdyreipfgwrdvgksglkqecldyikevwtdmrplslrrkmeeleks
Lyngbya aestuarii BL J	WP_023067057.1	M595_RS14390	mnseedttiyrvvvneeeqysiwpdyreipfgwrdvgksglkqecldyikevwtdmrplslrrkmeesqts
Lysobacter enzymogenes strain ATCC 29487	WP_074873762.1	BLU84_RS25230	msnpfddtngtflvlvndenqhslwpqfaeipagwravhgpterqecldyietnwtdmrpasliqameqdavqrna
Microbacterium oxydans strain NS234	WP_058631979.1	NS234_RS10810	mstnpfddedgvflalvndeeqyslwpefaevpsgwrivfgpanraatlefiektwtdlrprslreamaaeear
Micromonospora aurantiaca ATCC 27029	ADL46323.1	Micau_2788	maeprflvvrndeeqysiwsadrdlpagwhdtgfagsreeclahvdevwtdmrprsvreals
Micromonospora sp. ML1	CAJ34376.1	TioT	msvnpfddedgefyvlvndeeqhslwptfgdvpdgwrivfgpagraesvayveenwtdmrpkslreamsaa
Millisia brevis NBRC 105863	WP_066909728.1	MB1_RS16660	mstnpfddedgrffvlindedqhslwptfadvpegwrvvfgedsraacleyveknwtdmrprslreameadaaarkaqada
Mycobacterium aurum strain NCTC10437	WP_048633068.1	EL337_RS18125	mstnpfddehgtfhvlandeeqyslwptfaevptgwrvvfgdgsradcleyvektwtdlrprslrdams
Mycobacterium conceptionense strain IS-2586	WP_019345598.1	A5746_RS02470	mstnpfddengtfhvlvndegqhslwpafadvpagwqvvfgpagraecldhveanwtdlrpaslreamssst
Mycobacterium farcinogenes strain DSM 43637	WP_036393643.1	BN975_RS24235	mstnpfddengmfhvlvndegqhslwpafadvpagwqvvfgtagraecldyveanwtdlrpaslreamssst
Mycobacterium iranicum UM_TJL	WP_024445279.1	N420_RS0106475	mstnpfddedgifyvlsndeqqyslwpafadipagwqvvfgestrsdclayveenwtdmrprslreams
Mycobacterium llatzerense strain CLUC14	WP_043399539.1	TL10_RS18875	mstnpfdddngsffvlvndeeqhslwptfadvpagwrvvfgeadrascleyierewtdirpkslrdrlavgqrl
Mycobacterium mucogenicum strain CCH10-A2	WP_061001811.1	AX746_RS11260	mstnpfdddngsffvlindeeqhslwptfadvpagwrvvfgeadrascleyierewtdirpkslrdrlavgqql
Mycolicibacterium smegmatis MC2 155	YP_884812.1	MSMEG_0399, GplH	msinpfdddngsffvlvndeeqhslwpsfadvpagwrvvfgeasradclefieqnwtdirpkslrerlaqggaldg
Mycobacterium sp. 852013-51886	WP_066839109.1	A5757_RS22455	mstnpfdddsgsffvlvndegqhslwptfaevpagwravhgeapraecleyveqhwtdirpktlrerlpaggasdn
Mycobacterium sp. NAZ190054	WP_067954378.1	ASJ79_RS09065	mstnpfdddngtfyvlvneeeqyslwptfadvpagwrvvfgestradclayveetwtdlrprslreamnanp

Mycobacterium sp. ST-F2	WP_073695485.1	EB75_RS11255	mstnpfdddngtffvlindeeqhslwptfadipqgwrvvhgeadrascleyierewtdirpkslrdrlavgqrl
Mycobacterium tuberculosis strain HN-506	BAX49639.1	HN506_0250, MbtH	mstnpfddddgeffvlindeeqhslwptfadvpagwrvvhgeaeraacldyieqnwtdirpkslrerlatgqggsg
Myxococcus xanthus DK 1622	ABF91873.1	MXAN_3118	mtderedttvykvvvnheeqysiwpadrenalgwkdagkqglkaecleyikevwtdmrplslrkkmeelks
Nocardia paucivorans NBRC 100373	WP_040788378.1	ON32_RS02360	mstnpfddedgrfyvlvndeeqhslwptfaevpagwrvvfgedsraacveyveknwtdmrpkslreamaadeaarqaka
Nocardia paucivorans NBRC 100373	WP_040792270.1	ON32_RS18825	mstnpfddedgrffvlinneeqyslwptfaevplgwrvvfgednrascieyvekswtdmrpkslrdamaaddavrravrs
Nocardia sp. BMG51109	WP_024803007.1	D892_RS0120300	msknpfddedgrffvlvneedqhslwpvfaevpagwrivfgedtrsacieyveknwtdmrprslreameadlqnaatesa
Nocardia terpenica strain IFM 0406	WP_067589043.1	AWN90_RS28995	matstnpfddedgrfvvlvnaeeqhslwptfadvpegwrvvfgedtraacleyvernwtdmrpktlreamnsan
Nocardia uniformis subsp. tsuyamanensis	AAT09800.1	NocI	mlgenedsgefevvvnheeqysiwpadravpdgwrtagqrgakraclewidanwtdmrplslrealrgagdra
Nocardiopsis dassonvillei strain NCTC10488	WP_013155306.1	EL646_RS14850	msnpfddedarflvlvndegqhslwpafaevprgwrvaqgetsraealeyverewtdlrpasliaaqeg
Nonomuraea solani strain CGMCC 4.7037	SEG91870.1	SAMN05444920_107375	mtnpfddengtflvlvndegqhslwpdfadvpagwetvfgpgthaaaldyveqnwtdmrplslqramge
Paenibacillus sp. SMB1 366	WP_111146967.1	DNH61_RS12385	msnpfenpegsyhvlvneegqfslwpsfigvpqgwkivcerqsrqasldfinagwtdmrpnsirplesagerrg
Paraburkholderia caribensis MBA4	WP_035999678.1	K788_RS23020	mtqnthqdaqhdvqytvvindeeqysiwptfrdvpagwrevgvrgpkaaclehiesvwtdmrpaslrrhmdaapntr
Paraburkholderia endofungorum strain HKI456	WP_104078542.1	B0O95_RS14865	msnpfddpngtflvlvndenqhslwpnfvevpagwravhgpdarqacleyiernwtdmrpasliesmgklsvieperka
Paraburkholderia phymatum STM815	WP_012403342.1	BPHY_RS20385	mtqnthqdaqddvqytvvvndeeqysiwptfrdvpagwrevgvrgpkaaclehieavwtdmrpaslrrhmdaapntr
Paraburkholderia terrae strain DSM 17804	WP_042304274.1	C2L65_RS16860	mtqnthqdaqddvqytvvindeeqysiwptfrdvpagwrevgvrgakaaclehiesvwtdmrpaslrrhmdaapntr
Paraburkholderia tropica strain P-31	WP_065064659.1	A6456_RS31910	mswddenaefevvineeeqysiwpsykpipggwrtvgkkdkkaeclayieehwtdmrpaslrramdgdgaskdavipasqh

Paracoccus alcaliphilus strain DSM 8512	WP_090610398.1	BMW58_RS02075	msnpfddqdgiflvlvndenqhslwpefaavpegwrsvfgpdkrpacidyvetswtdmrpaslvametekaa
Paracoccus sp. 1011MAR3C25	WP_119747089.1	D3P04_RS06580	msnpfddqdgiflvlvndenqhslwpefaavpegwrtvfgpdkrpacisyveanwtdmrpaslvaaeteeaa
Polyangium brachysporum DSM 7029	AKJ29078.1	AAW51_2387, GlbE	mthllddvdgaflvlvnaegqhslwpaslqvpagweiafgsetrpacldyieahwtdlrprsvarrahaqpea
Pseudomonas aeruginosa PAO1	AAG05800.1	PA2412	mtsvfdrddiqfqvvvnheeqysiwpeykeipqgwraagksglkkdclayieevwtdmrplslrqhmdkaag
Pseudomonas chlororaphis strain B25	WP_124320129.1	C4K04_RS11825	mnpfdnedgqflvlcndegqhslwpafsqvpagwqilfgaaarqdcvtyiaqnwtdmrpaslrqrqg
Pseudomonas syringae UB246	WP_027900544.1	N009_RS0115470	mvfdredltfqvvcnheeqysiwpdykaipngwravgmkglkkdcleyieqhwtdmrplslrqkmdqdkvva
Pseudonocardia sp. HH130629-09	WP_060712088.1	XF36_RS12275	mstnpfdddngtfrvlvndegqhslwpdfadvpagwtsvhgpadrtscldyveriwtdlrprslrertna
Pseudonocardia sp. P2	WP_020625570.1	PP2_RS22670	mstnpfddengtfrvlvndegqhslwpdfaevpagwtsvhgpadrtscldyvernwtdlrprslrertda
Rhizobium leguminosarum strain Norway plasmid pRLN1	AUW45697.1	CUJ84_pRLN1000230, VbsG	mdnlephddlwiavidterrysiwpqdkripvgwepagfagsrqhclahirdvwadprplslrsaiagdarl
Rhodococcus erythropolis strain BG43	AKD96145.1	XU06_04630	mstnpfddedgrfyvlvndedqhslwptfsevpagwrvvfgednrkacleyveanwtdmrpkslreameadqaaggkhaves
Rhodococcus fascians 02-816c	WP_019663793.1	BH88_RS01340	mstnpfddedgrffvlvndedqhslwptfsdvpqgwrvvfgedsraacleyveknwtdmrprslreameadaaarqsaegtpea
Rhodococcus jostii RHA1	ABG96503.1	RHA1_ro04717	mstnpfddeegrfyvlvndedqhslwptfsevpagwrvvfgeesraacleyveknwtdmrpkslreameadeksggrhsvdks
Rhodococcus kyotonensis strain KB10	WP_068425313.1	A3K89_RS11385	mstnpfddedgrffvlvndedqhslwptfsdvpqgwrvvfgedsraacleyveknwtdmrprslreameadaaarqsaegkpea
Rhodococcus marinonascens NBRC 14363	WP_072687017.1	RM2_RS01245	mstnpfddeegrfyvlvndedqhslwptfsevpagwrvvfgeesraacleyveknwtdmrpkslreameadeksdgrhsvekn
Rhodococcus rhodnii NBRC 100604	WP_010840282.1	RR1_RS10745	mstnpfddedgrfyvlvndeeqyslwptfsevptgwrvvfgedsraacleyveknwtdmrpkslrdameadeaarraadas
Rhodococcus sp. WMMA185	WP_070379598.1	BFN03_RS14560	mstnpfddeegrfyvlvndeeqhslwptfsevpagwrvvfgeesraacleyveknwtdmrpkslreameadeksrvdns

Rhodococcus triatomae strain DSM 44892	WP_072736193.1	BLQ89_RS03470	mstnpfddedgrfyvlvndedqhslwptfsevpagwrvvfgedsraacleyveknwtdmrprslreameadetsggrhsvdk
Rhodococcus tukisamuensis strain JCM 11308	WP_072843117.1	BLT10_RS02955	mstnpfddedgrfyvlvndedqhslwptfsdvpqgwrvvfgedsrqacveyveknwtdmrpkslreamekdkaagnn
Saccharothrix mutabilis subsp. Capreolus	ABR67757.1	CmnN	mdtylvvvnheeqysvwpadrplpagwraegtsgdkeqclahietvwtdmrplsvrrraeav
Salinispora arenicola CNT005	WP_028678147.1	B110_RS0125710	mttnpfddedgqfyvlvndedqyslwptfadipagwtarygpatradsvsyveshwtdmrprslrevmdg
Streptomyces acidiscabies a10	GAQ51743.1	a10_01523, TxtH	mpspfddhdgqfhvlrneegqfslwpnfadipsgwrsvsgpsprgsaleyiekewtdmrpasvre
Streptomyces acidiscabies a10	WP_010357635.1	AV141_RS45635, Saa2_09136	mstnpfedddarylvltndegqhslwpafaevpdgwtvahpedtrqacleyvernwtdmrpkslveamga
Streptomyces agglomeratus strain 5-2-6	WP_069926457.1	BGK70_RS28020	mstnpfddengqfhvlvndedqhslwpafaevpagwrsvfgpaaraesvayveehwtdmrprslreamng
Streptomyces anulatus subsp. fumigatus	ADG27355.1	AcmR	mtnpfenpdgqylvlvnaegqyslwpafaevpagwtvalaetdrqscldhieahwtdmrplslvrr
Streptomyces antibioticus JCM 4620	AAG34186.1	SimY	manpfddqrgsflvlrnaeeqvslwpafagvptgwqvakgpnsraaclayieeawtdlrpkslidatdps
Streptomyces clavuligerus ATCC 27064	WP_003958107.1	SCLAV_p1293, SCLAV_RS33345	msgdvrerwkvvvndeeqysiwsaqretpagwredgvhgtkdeclahiervwtdmrprslrertapeaear
Streptomyces coelicolor A3(2) M145	NP_624806.1	SCO0489, CchK	mstnpfddadgrflvlvndegqhslwpafaavpggwttvfeentrdaclayveanwtdlrprslartada
Streptomyces coelicolor A3(2) M145	AAD18046.1	Cda-orfX, CdaX	mtnpfedadgrylvlvndegqhslwpsfvdvpagwtvalgesdreacleyveknwtdmrprslveamstgn
Streptomyces coeruleorubidus NRRL 18370	ADN26246.1	PacJ	mfddegrefcvvrngeeqysiwpsgrelpagwvevgtrgpkaeclayldrtwtdmrpaslrqslsadegsaa
Streptomyces europaeiscabiei 89-04	WP_046706407.1	AWZ11_RS05060	mavnpfddengefhvvvndeeqhalwptyadvpdgwrsvagpagraesiayveenwtdmrprslreatpap
Streptomyces europaeiscabiei 89-04	WP_010350602.1	AWZ11_RS28370, TxtH	mpspfddhdgqfhvlrneegqfslwpnfadipsgwrsvsgpsprgsaleyiekewtdmrpasvre
Streptomyces filamentosus NRRL 11379	AAX31560.1	DptG	manpfenndgsylvlvndegqyslwpafadvpagwtvtfgessrqecldhinenwtdmrpkslirqmendrttaa

Streptomyces hygroscopicus subsp. aabomyceticus	AAU34213.1	MppT	mgtnpfddpdgrylvlvneedqhslwpafaevpqgwtvalaetdrqsaldfitehwtdmrprslvrameea
Streptomyces ipomoeae 91-03	WP_009337597.1	STRIP9103_RS39770	mtdnpfddedgtflvlvndenqhslwplfadvpagwttvhgpapraacldyiecewtdmrpaslvramsaagegdr
Streptomyces ipomoeae 91-03	AEL30516.1	TxtH	msnpfestthpmfvllndeqqmslwpaftpvpdgwraafgpavrteclayierewgdlrpaslr
Streptomyces katrae strain NRRL ISP-5550	WP_045946089.1	VR44_RS04780	mstnpfddgdgrfhvlvndedqhslwpvfaevpagwrvvfgeadraacleyvernwtdlrprslreamaad
Streptomyces lasaliensis ATCC 31180	BAE98157.1	Ecm8	mstnpfddetgrfhvlvngedqhslwpafaevpagwrsvfgpaarteslayveehwtdmrprslreaadg
Streptomyces lavendulae	AAK81828.1	ComB	mtnpfdnengtflvlvndegqhslwpvfaeipqgwttafgeasraeclefveqnwtdmrpkslvarmegtata
Streptomyces pini strain PL19	SFL35986.1	SAMN05192584_117108	mtnpfddsegtfhvlvndegqhslwpnfvdipagwqavvndrprqecldyieenwtdmrpkslieamaahektaae
Streptomyces pristinaespiralis Pr11	CBH31049.1	MbtY	msnpfedaegtflvlvnhegqyslwpsfaevpagwtvalpatdresalahitdrwtdmrpqslidamngtaa
Streptomyces rishiriensis strain DSM 40489	AAG29779.1	CouY	atnpfddengvylllvndegqhslwpsfaalpkgwtvildeasrqecldyvnehwtdmrplslqpamgde
Streptomyces roseochromogenes subsp. oscitans DS12.976	AAN65223.1	CloY	matnpfedengsylvlingegqhslwpsfadvpngwtvifneasrqdcldyvnehwtdmrplslqramgge
Streptomyces scabiei 87.22	CBG70277.1	SCAB_31771, TxtH	mpspfddhdgqfhvlrneegqfslwpnfadipsgwrsvsgpsprgsaleyiekewtdmrpasvre
Streptomyces scabiei 87.22	WP_012998279.1	SCAB_3331, MLPlipo	mtnpfenpdgiyrvltndegqhslwpdftpvpegwatvfgpdsrtacleyverewtdlrpqslvramge
Streptomyces scabiei 87.22	WP_013005929.1	SCAB_85461, MLPscab	msnpfedpegtylvlvndenqhslwpdfvevpagwrvvhgpgsrqscldhvethwtdmrprslaesmdg
Streptomyces sp. 11-1-2	WP_078638235.1	CGL27_RS02360, CGL27_02270	msnpfddadarylvlvndegqyslwpvfaevpagwtgvhgpdgrqacldhinenwtdmrpkslveamdaaas
Streptomyces sp. 11-1-2	WP_046086618.1	CGL27_RS04255, CGL27_04160	mddnaryqvlrndedqyslwpadlevpegwrpvgkegtkdecsayvdevwtdmrprslrermdnvas
Streptomyces sp. 11-1-2	WP_119986043.1	CGL27_RS10110,	msaspalrpagdryvvvvnheeqysvwfadralpagwtatgaqgtrqecldhieqvwtdltprsvrasr

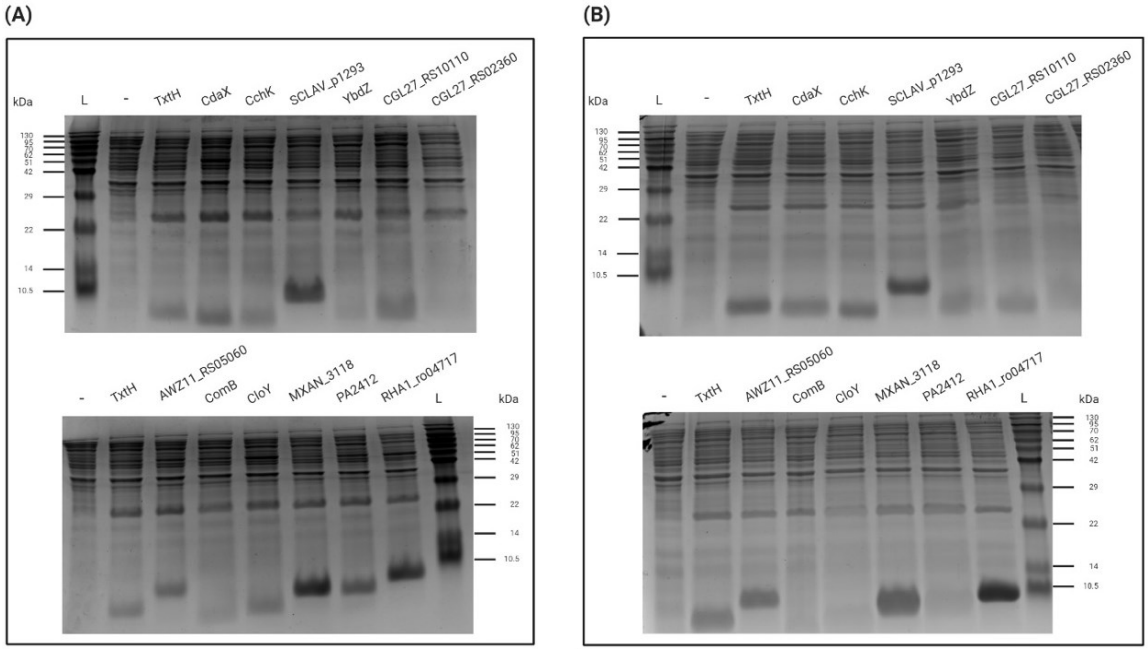
		CGL27_09910	
Streptomyces sp. 11-1-2	WP_086709734.1	CGL27_RS10655, CGL27_10440	msveqnddntvyrvvlndeeqysiwwahrdlpagwhaegtegtrdeclarigdiwtdmrplslrrrmegqhtav
Streptomyces sp. 11-1-2	WP_078641478.1	CGL27_RS18335, CGL27_18035	msnpfenpdgryfvlvneegqhslwpafaevpagwtvahgeddrqacldyvnehwtdmrpkslitamdgtsa
Streptomyces sp. 11-1-2	WP_078644297.1	CGL27_RS34600,CGL27_34265	msnpfenaegrffvlvneerqyslwpafaevpagwtvvhgedtrdaclehinqnwtdmrpkslvdamsaa
Streptomyces sp. CB02056	WP_074005862.1	AMK13_RS36500	mtnpfddqdgtflvlvneenqhslwpqfadvpdgwtvvhgpdtnaacleyvekswtdmrprsladamdtrk
Streptomyces sp. cf386	SDP55221.1	SAMN04487981_12661	manpfddnsgvfrvlvndegqhslwpdfapvpdgwssvhgpddraacleyiernwtdmrprslaeamrkaeg
Streptomyces sp. NRRL B-1140	WP_053672917.1	ADK65_RS25300	mstnpfddedgrfhvvvndeeqhslwpafaevpagwrvvfgeaaraecleyveqnwtdlrpkslreamaag
Streptomyces sp. NRRL S-920	WP_030794579.1	IG54_RS0135380	mstnpfddengtfhvlendegqhslwpvfvdvpdgwrvvlgdaardecleyveenwtdlrprslreamaad
Streptomyces sp. SNA15896	BAI63290.1	Swb18	mstnpfddedgrfhvlvndedqhslwpafaevpagwravfgpagraeslayveenwtdmrprslqqamda
Streptomyces tsukubensis VKM Ac-2618D c11	WP_006350627.1	EWI31_RS28810	mtnstnpfddpdgifhvvvndegqhalwpvfadipagwegvwgpgpragaleyveanwtdirpkslllaqya
Streptomyces turgidiscabies T45	GAQ77365.1	T45_09183, TxtH	mpspfddhhgqfyvlrneegqfslwpdfadipsgwhsvsgpvsrdsalgyiegqwtdmrptsara
Streptomyces vinaceus ATCC 11861	AAP92504	VioN	mndtpadtayqvvlndeeqysvwpvgrplpagwraegtvggrqacldhietvwtdlrplsara
Streptomyces viridifaciens DSM 40239	WP_046386617.1	BOQ63_RS18895	mtanpfdndagefhvlvneegqhslwpafaavpagwqsvfgpgsrgsaleyvetswtdirprsvgspersgtvtaegaanr
Streptomyces yeochonensis CN732	WP_037906290.1	BS72_RS03480	mtnpfddqdgtflvlvneedqhslwprfadvpdgwttvhgpdthaacleyieenwtdmrprsladamaaqr
Streptomyces viridochromogenes strain NRRL 3414	WP_048580916.1	ACM01_RS10820	mpspfddhdgqfhvlrneegqfslwpnfadipsgwrsvngpsprdsaleyiekewtdmrpasvre
Streptomyces zhaozhouensis CGMCC 4.7095	WP_097229363.1	CRP51_RS03180	msnpfddvdgvfrvlrneagqyslwpdfaevpagwtsvhgpaaraacleyvesdwtdqrpaglvrapgagd

Thermobifida fusca strain NBRC 14071	WP_011292287.1	TF1_RS05620	mtnpfdddegvflvlvndedqyslwpefaevpqgwrtvfgptsraaaldyinthwtdlrprslreameahstag
Thermocrispum agreste DSM 44070	WP_028847742.1	YWY_RS0111615	mstnpfddpdgrfhvlvndenqhslwpsfadipagwrsvfgpdtkdaclayveknwtdmrpasliaadd
Variovorax paradoxus 110B G369	WP_019658284.1	G369_RS0132670	mstscfdredetfivlvneedqysiwphwkavpsgwkavdgvkgdkkaalefveknwtdmrprslrdwmaaqdrapsaeaaas
Variovorax sp. YR266	WP_093178213.1	G369_RS0132670	mtnpfdngdasfvvltndenqhsiwpdfieipkgwkkvygpsakgdclefveknwtdmrpkslaqamgaqg
Xenorhabdus eapokensis strain DL20	WP_074023416.1	Xedl_RS08665	mnleqknpfdddeatfyvlinnhqqyslwpafaahptgwelvigpnsraaciayieehwvdmrpaslrepqinqidnadgltyr

Supplementary Table 3.3 Quality parameters of structural models built for $TxtA^A$, $TxtB^A$ and TxtH using SWISS-Model.

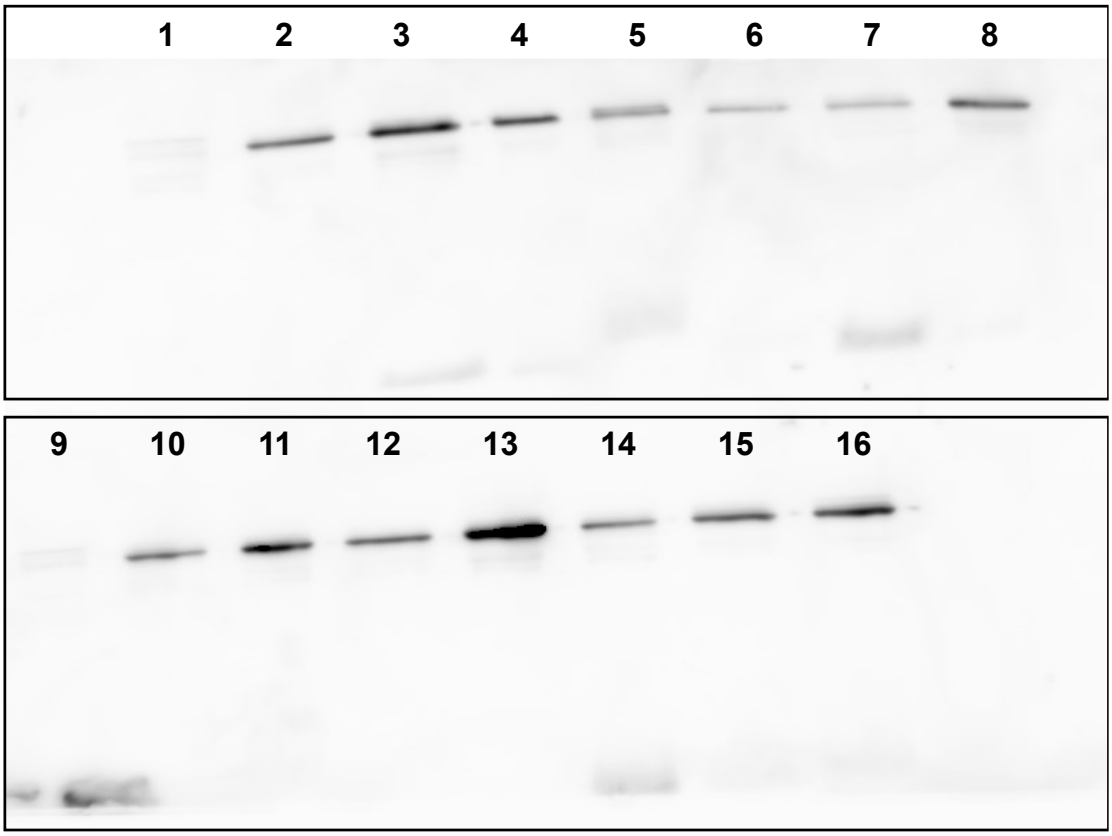
Protein	**$TxtA^A$**	**$TxtB^A$**	**TxtH**
Template PDB	5wmm_1	5wmm_1	6ea3_1
Description	TioS NRPS from *Micromonospora* sp. ML1	TioS NRPS from *Micromonospora* sp. ML1	FscK MLP from *Thermobifida fusca*
Reference	Mori et al., 2018b	Mori et al., 2018b	Bruner and Zagulyaeva, to be published
Method	X-ray, 2.9Å	X-ray, 2.9Å	X-ray, 1.65Å
Identity	42.28%	48.66%	56.92%
GMQE	0.72	0.7	0.82
QMEAN	-2.21	-3.05	0.31
Cβ	-2.48	-3.18	-0.66
All Atom	-2.03	-2.12	-0.38
Solvation	-1.65	-1.91	-0.37
Torsion	-1.26	-1.95	0.80

GMQE: Global Model Quality Estimation
QMEAN: Qualitative Model Energy ANalysis

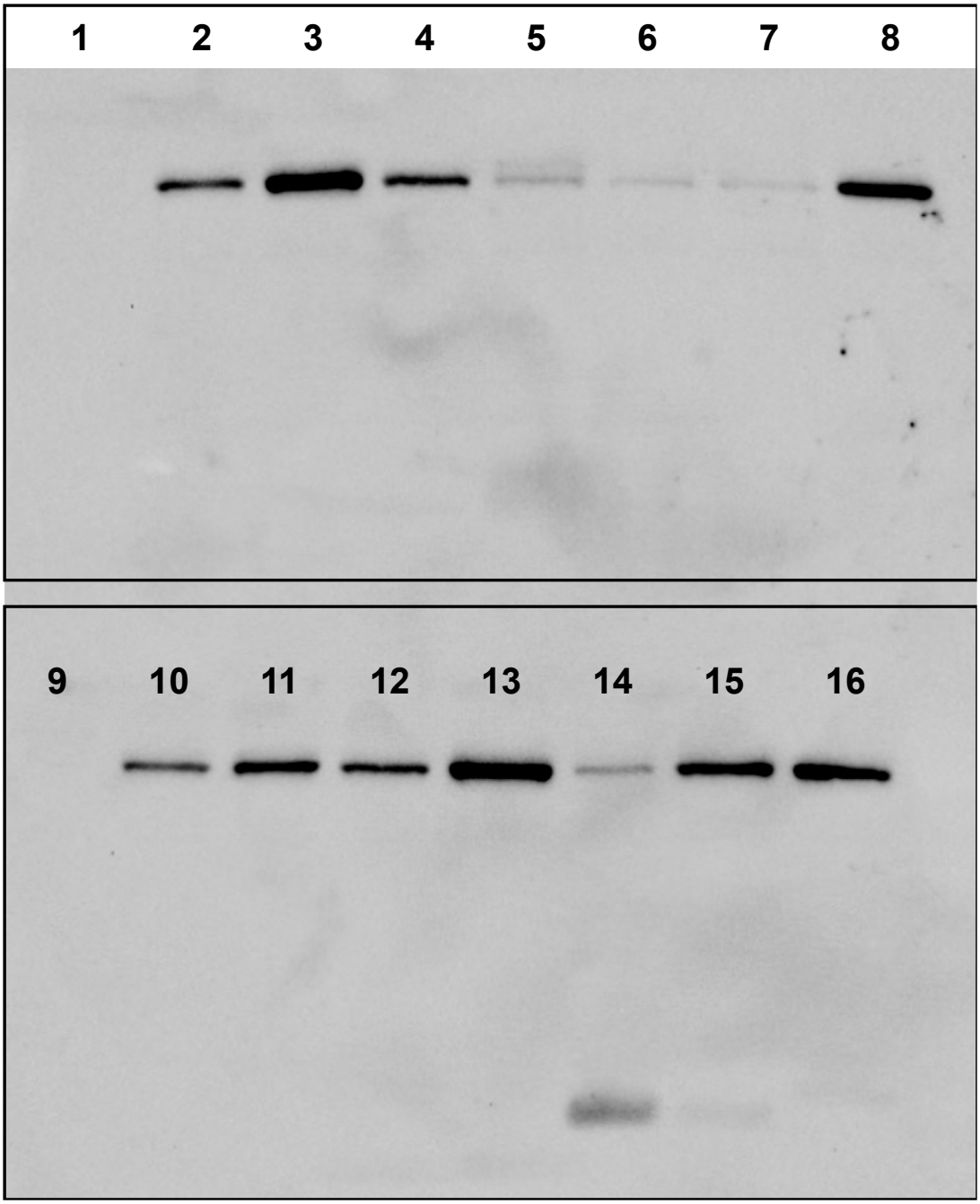


Supplementary Figure 3.1 SDS-PAGE analysis of total soluble protein extracts from *E. coli* BL21(DE3)*ybdZ:aac(3)IV* expressing HIS_6-$TxtA^A$ (A) or HIS_6-$TxtB^A$ (B) in the presence and absence (-) of different HIS_6-tagged MLPs. Lane L: PiNK Plus Prestained Protein Ladder (FroggaBio Inc). The MLPs are visible as prominent bands below the 10.5 kDa marker band.

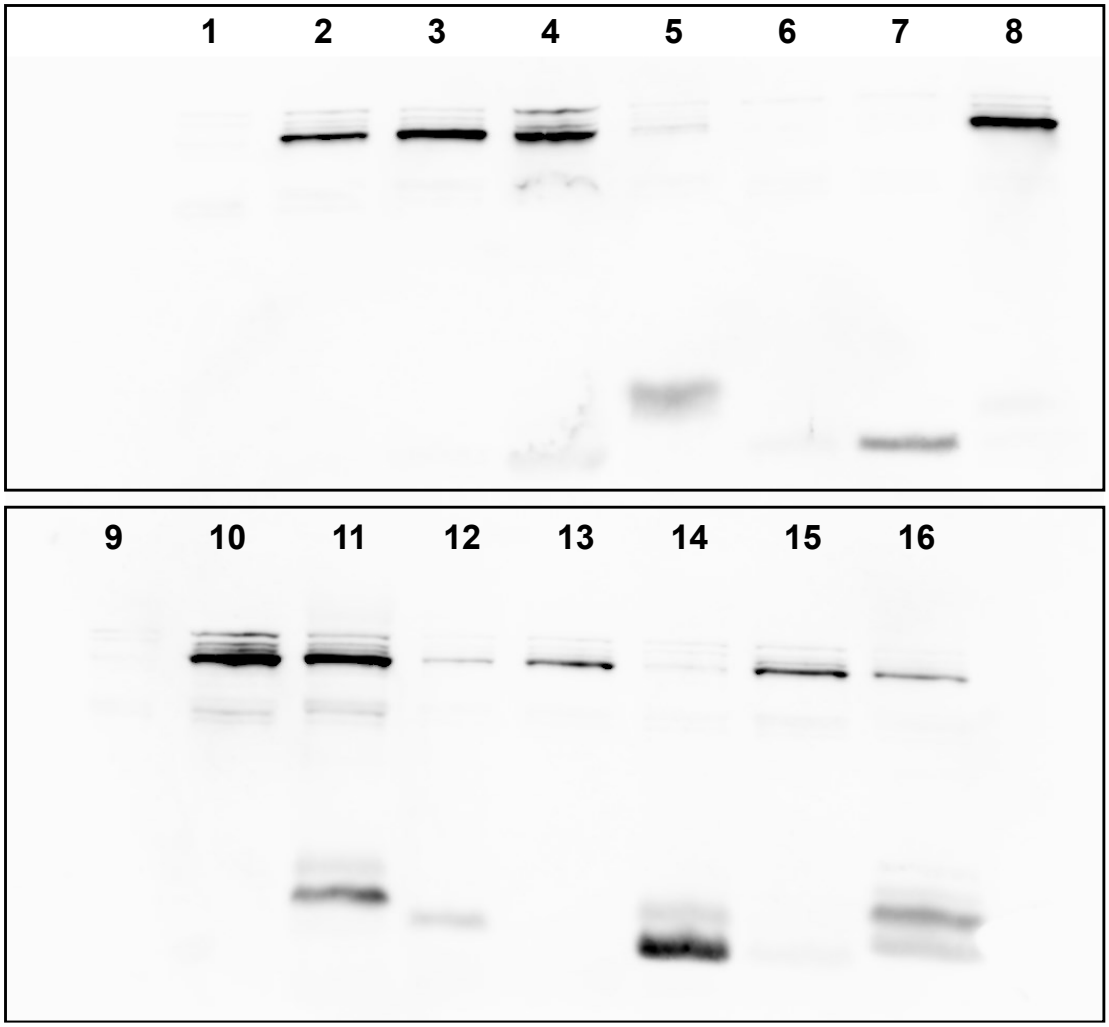
(A)



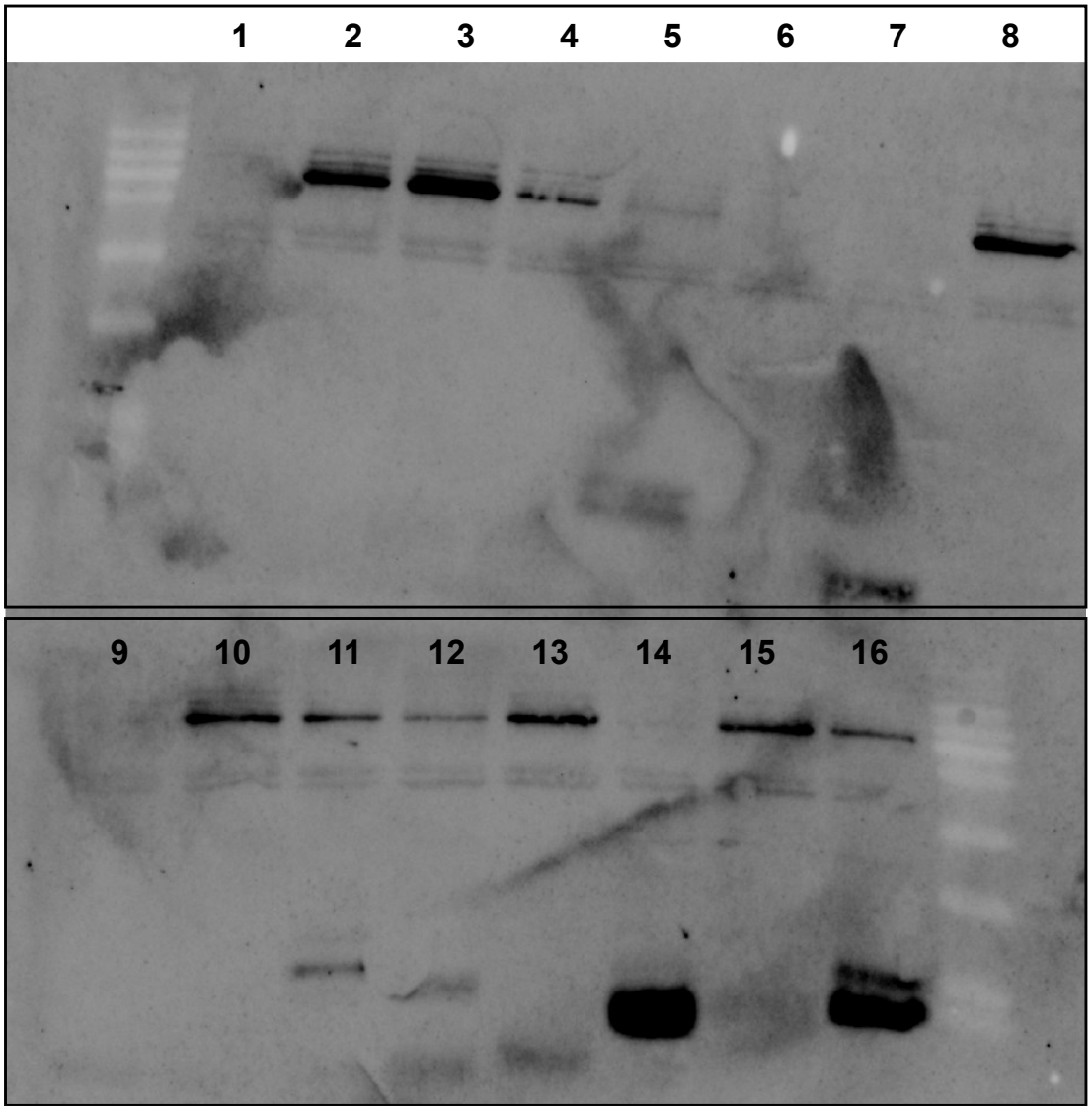
(B)



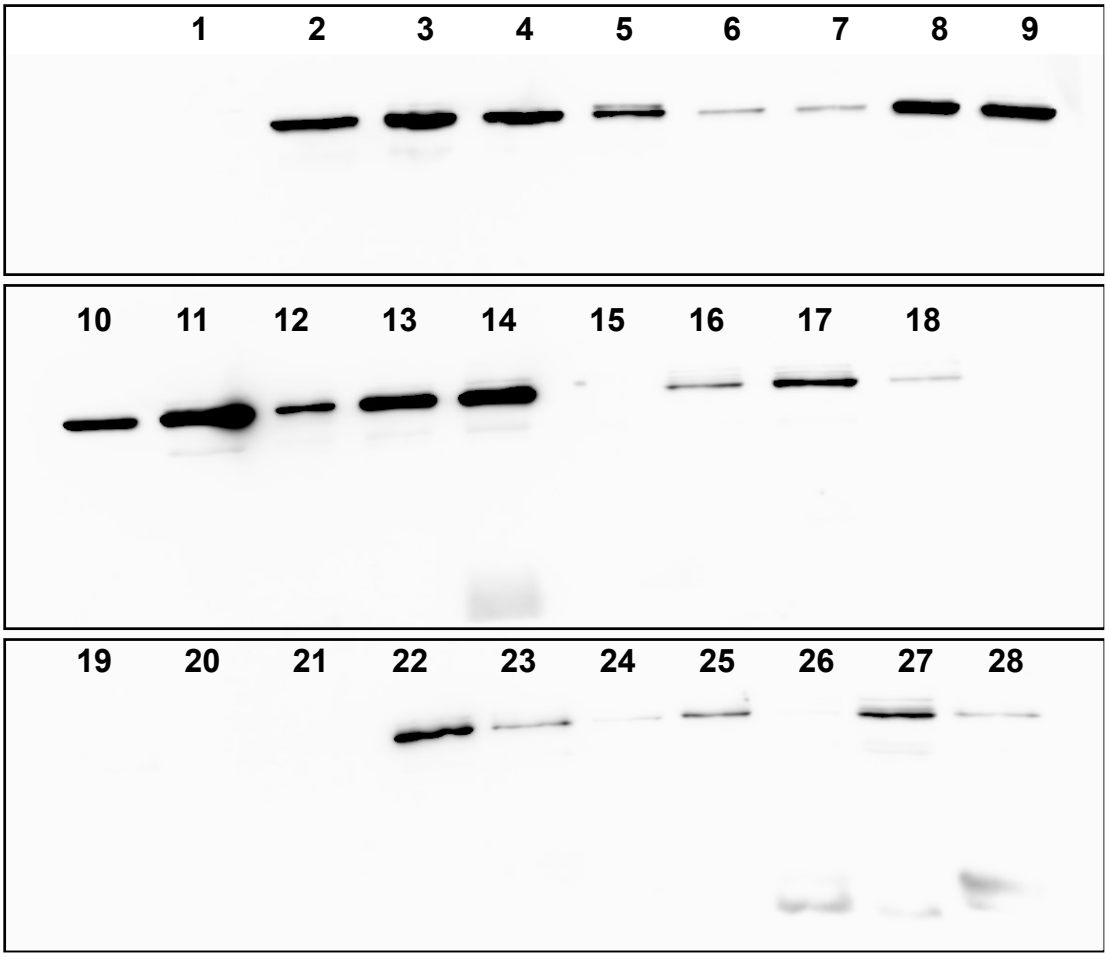
(C)



(D)

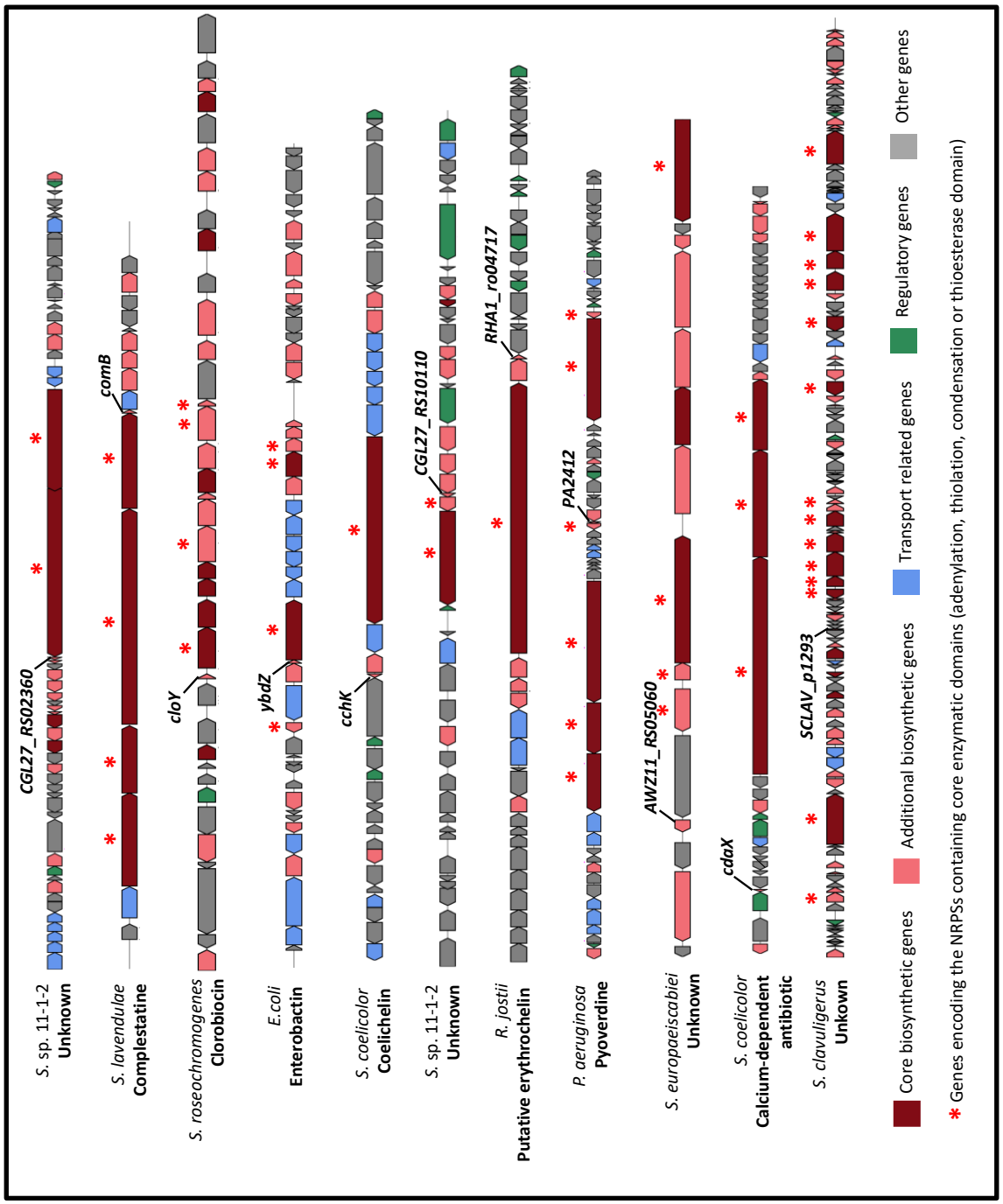


(E)

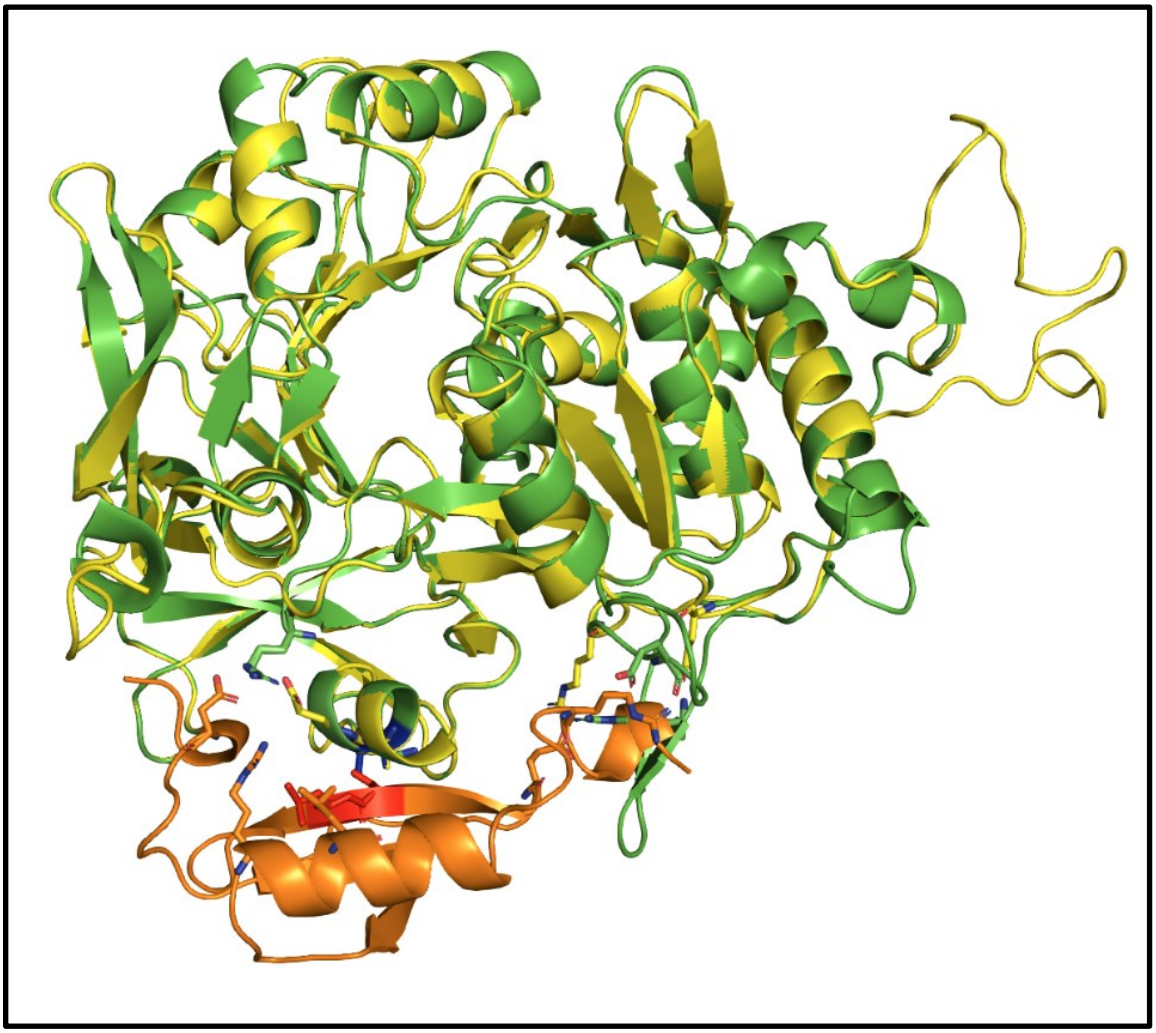


Supplementary Figure 3.2 Western blot analysis of soluble HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$ proteins expressed in the presence and absence of different HIS_6-tagged MLPs. (A, B) Duplicate membranes of soluble HIS_6-$TxtA^A$ expressed in the absence of an MLP (lanes 1 and 9) or co-expressed with HIS_6-TxtH (lanes 2 and 10), HIS_6-CdaX (lane 3), HIS_6-CchK (lane 4), HIS_6-SCLAV_p1293 (lane 5), HIS_6-YbdZ (lane 6), HIS_6-CGL27_RS10110 (lane 7), HIS_6-CGL27_RS02360 (lane 8), HIS_6-AWZ11_RS05060 (lane 11), HIS_6-ComB (lane 12), HIS_6-CloY (lane 13), HIS_6-MXAN_3118 (lane 14), HIS_6-PA2412 (lane 15), and HIS_6-RHA1_ro04717 (lane 16). (C, D) Duplicate membranes of soluble HIS_6-$TxtB^A$ expressed in the absence of an MLP (lanes 1 and 9) or co-expressed with HIS_6-TxtH (lanes 2 and 10), HIS_6-CdaX (lane 3), HIS_6-CchK (lane 4), HIS_6-SCLAV_p1293 (lane 5), HIS_6-YbdZ (lane 6), HIS_6-CGL27_RS10110 (lane 7), HIS_6-CGL27_RS02360 (lane 8), HIS_6-AWZ11_RS05060 (lane 11), HIS_6-ComB (lane 12), HIS_6-CloY (lane 13), HIS_6-MXAN_3118 (lane 14), HIS_6-PA2412 (lane 15), and HIS_6-RHA1_ro04717 (lane 16). (E) Third replicate set of membranes of soluble HIS_6-$TxtA^A$ and HIS_6-$TxtB^A$ expressed in the presence and absence of different MLPs. HIS_6-$TxtA^A$ expressed without an MLP (lane 1) or co-expressed with HIS_6-TxtH (lane 2), HIS_6-CdaX (lane 3), HIS_6-CchK (lane 4), HIS_6-

SCLAV_p1293 (lane 5), HIS_6-YbdZ (lane 6), HIS_6-CGL27_RS10110 (lane 7), HIS_6-CGL27_RS02360 (lane 8), HIS_6-AWZ11_RS05060 (lane 9), HIS_6-ComB (lane 10), HIS_6-CloY (lane 11), HIS_6- MXAN_3118 (lane 12), HIS_6-PA2412 (lane 13), and HIS_6-RHA1_ro04717 (lane 14). HIS_6-$TxtB^A$ expressed without an MLP (lane 15) or co-expressed with HIS_6-TxtH (lane 16), HIS_6-CdaX (lane 17), HIS_6-CchK (lane 18), HIS_6-SCLAV_p1293 (lane 19), HIS_6-YbdZ (lane 20), HIS_6-CGL27_RS10110 (lane 21), HIS_6-CGL27_RS02360 (lane 22), HIS_6-AWZ11_RS05060 (lane 23), HIS_6-ComB (lane 24), HIS_6-CloY (lane 25), HIS_6-MXAN_3118 (lane 26), HIS_6-PA2412 (lane 27), and HIS_6-RHA1_ro04717 (lane 28).



Supplementary Figure 3.3 Known or predicted non-ribosomal peptide biosynthetic gene clusters harbouring the MbtH-like protein (MLP)-coding genes used in this study.



Supplementary Figure 3.4 Predicted 3-dimentional structures of the *S. scabiei* TxtA^{A} (green), TxtB^{A} (yellow) and TxtH (orange). The A-domain structures were predicted using the crystal structure of TioS from *Micromonospora* sp. ML1 (PDB: 5wmm_1) as the template, and the structure of the TxtH MLP was predicted using the crystal structure of FscK from *Thermobifida fusca* (PDB: 6ea3_1) as the template. The generated model of TxtH positioned next to the A-domains is based on the location of the TioT MLP that is bound to the A-domain in TioS. A more detailed image of the predicted interaction interface is shown in Figure 3.6A.

Supplementary Data 3.1 Nucleotide sequences of the DNA fragments synthesized by TWIST BIOSCIENCE. The MLP coding sequence within each fragment is indicated in blue. Coding sequences that were codon-optimized for expression in *Streptomyces* are indicated with *.

Myxococcus xanthus DK1622 *MXAN_3118**
TGTCTTGACCCCTTTTCCTGGATTGTGGCCAGAATACGCCCCGTTCCAGGAGA
GAAATCA**ATGACCGACGAACGGGAGGACACCACCGTCTACAAGGTGGTG**
GTGAACCACGAAGAACAGTACTCCATCTGGCCGGCGGACCGCGAAAACG
CGCTCGGGTGGAAGGACGCGGGCAAGCAGGGCCTCAAGGCGGAGTGCC
TGGAATACATCAAGGAAGTCTGGACCGACATGCGCCCGCTCAGCCTGCG
CAAGAAGATGGAAGAACTCAAGAGCTAGTCACGGCACCGTCCGTGCGCAC
GCCCCTCG

Rhodococcus jostii RHA1 *RHA1_ro04717**
CCGCAACGGCACGCTTGCTAGTGTGAACGCACAGGACACGTATGCGACAAGC
GAGGCGAG**ATGTCCACCAACCCCTTCGACGACGAAGAGGGGCGGTTCTA**
CGTCCTGGTCAACGACGAGGACCAGCACTCCCTGTGGCCCACGTTCTCC
GAGGTCCCCGCGGGGTGGCGGGTCGTGTTCGGCGAAGAGTCGCGCGCC
GCGTGCCTCGAATACGTCGAAAAGAACTGGACCGACATGCGCCCCAAGA
GCCTCCGCGAAGCGATGGAAGCCGACGAAAAGTCCGGCGGGCGGCACA
GCGTCGACAAGAGCTGACCTCCCCGGCGCCGCGTTTCGGCGACGATG

Streptomyces roseochromogenes subsp. *oscitans* DS12.976 *cloY*
CACTCCGGGCGGTCAATGGTGGAACAAGACGCGTGAAATAACATCTGGGAGG
TATTCGTC**ATGGCGACGAACCCGTTCGAGGACGAGAACGGCTCCTATCTG**
GTCCTGATCAATGGCGAGGGGCAGCATTCTCTGTGGCCTTCGTTCGCTG
ATGTTCCCAACGGGTGGACTGTCATTTTCAACGAGGCGTCGCGGCAAGA
CTGCCTCGATTACGTCAATGAGCATTGGACAGACATGCGGCCGTTGAGC
CTGCAGCGGGCGATGGGTGGCGAGTAGCATCTGCTCATTCAAAAGTG
GATTTGGCAA

Streptomyces lavendulae comB
CCGCCCCTCGACAGACCGAAGAATCCGCGAACCACTCCGGAAAAGAAGGTG
AATGACACC**ATGACTAACCCTTTCGACAACGAGAACGGCACTTTCCTGGT**
GCTCGTCAACGACGAGGGTCAGCACTCGCTCTGGCCGGTTTTCGCGGAG
ATCCCGCAGGGCTGGACGACCGCGTTCGGTGAGGCGAGCCGGGCCGAA
TGCCTGGAATTCGTCGAGCAGAACTGGACCGACATGCGGCCCAAGAGCC
TCGTCGCCCGTATGGAGGGCACCGCCACGGCCTGAGAGATTGAGCACTCA
CGCATTGAGAAGAAG

Pseudomonas aeruginosa PA01 *PA2412**
TCGTTATCCAACGCAAGGGCCGTGCTGCGCGCGGCCGTTCGATCCCATCAGG
AGCAAGCA**ATGACTTCAGTGTTCGACCGTGACGACATCCAGTTCCAGGTA**
GTGGTCAACCATGAGGAGCAGTATTCCATCTGGCCGGAATACAAGGAGA

TTCCCCAGGGCTGGCGGGCGGCCGGCAAGAGCGGCCTGAAGAAGGACT
GCCTGGCCTACATCGAGGAAGTCTGGACCGACATGCGCCCGCTGAGCCT
GCGCCAGCACATGGACAAGGCGGCCGGCTGAGTGGGCGGTACGCCCGTAC
GGCTGTTCTGC

Escherichia coli BL21(DE3) *ybdZ**
GGGGCTAATCGACCTCTGGCAACCACTTTTCCATGACAGGAGTTGAAT**ATGG**
CCTTCTCCAACCCCTTCGACGACCCCCAGGGGGCCTTCTACATCCTCCGG
AACGCCCAGGGCCAGTTCTCCCTCTGGCCCCAGCAGTGCGTGCTCCCGG
CGGGCTGGGACATCGTGTGCCAGCCGCAGAGCCAGGCCTCGTGCCAGCA
GTGGCTCGAAGCGCACTGGCGGACGCTGACCCCGACGAACTTCACCCAG
CTCCAGGAAGCGCAGTGAGCCAGCATTTACCTTTGGTCGCCGCACAGCCCG

Supplementary Data 3.2 Predicted 3-dimentional structures of the *S. scabiei* TxtA^{A}.

Supplementary Data 3.3 Predicted 3-dimentional structures of the *S. scabiei* TxtB^{A}.

Supplementary Data 3.4 Predicted 3-dimentional structures of the *S. scabiei* TxtH.

The above supplementary data files can be accessed using the link below:

https://drive.google.com/file/d/1hywM8kkPlxKcXpjjyPL79OhTO1ecl4EG/view?usp=sharing

CHAPTER 4

Assessing the impact of MbtH-like proteins on the adenylation activity of the thaxtomin non-ribosomal peptide synthetases in *Streptomyces scabiei*

Yuting Li, Kapil Tahlan and Dawn R. D. Bignell

4.1 Abstract

The biosynthesis of nonribosomal peptide metabolites involves large, multidomain enzymes called nonribosomal peptide synthetases (NRPSs). A typical NRPS module incorporates one amino acid into the growing peptide backbone and is composed of three core domains: adenylation (A-), thiolation (T-) and condensation (C-). Optional domains such as a methylation (M-) domain can also be present that incorporate additional modifications to the building blocks. In addition, an auxiliary protein belonging to the MbtH-like protein (MLP) family is considered an integral component for the proper function of many NRPSs. The production of thaxtomin A in *Streptomyces scabiei* involves two NRPS modules encoded by the *txtA* and *txtB* genes, and an MLP encoded by the *txtH* gene. Our previous work revealed the importance of TxtH for thaxtomin A production and showed that TxtH is essential for promoting the soluble production of the TxtA and TxtB A-domains in *Escherichia coli*. Additionally, we demonstrated that MLPs from other biosynthetic pathways can functionally replace TxtH to promote the soluble expression of the Txt A-domains in *E. coli* and restore the production of thaxtomins in *S. scabiei* to varying degrees. In this study, we sought to investigate the impact of TxtH and other MLPs

on the enzymology of the Txt A-domains. The A-domain of TxtA was overexpressed with TxtH in *E. coli* as HIS-tagged proteins, and *in vivo* chemical cross-linking revealed that the proteins form a complex with a 2:2 molar ratio. The A-domain was subsequently purified and analyzed for adenylation activity using a colorimetric assay; however, no activity was detected in the presence of its preferred substrate L-phenylalanine despite testing different assay conditions. Further efforts were made to co-express the Txt AMT-domains with TxtH, but the proteins were mainly detected as truncated forms. The possible reasons for our inability to detect the NRPS enzyme activity, and future directions of this research, are discussed.

4.2 Introduction

Thaxtomins are a family of phytotoxic cyclic dipeptides synthesized from 4-nitro-L-tryptophan and L-phenylalanine. Eleven thaxtomin analogues have been isolated and characterized, and they differ in the presence or absence of *N*-methyl and/or hydroxyl groups at specific locations on the thaxtomin (Txt) backbone (King and Calhoun, 2009). Thaxtomin A is the major analogue produced by different plant pathogenic *Streptomyces* species that are responsible for scab disease of potato, including *Streptomyces scabiei*, *Streptomyces acidiscabies, Streptomyces turgidiscabies, Streptomyces europaeiscabiei, Streptomyces stelliscabiei* and *Streptomyces niveiscabiei* (King and Calhoun, 2009). A positive correlation between the production of thaxtomin A and the pathogenicity of *Streptomyces* spp. has been documented (Goyer et al., 1998; Healy et al., 2000; King et al., 1991; Kinkel et al., 1998; Loria et al., 1995), and a thaxtomin biosynthetic mutant of *S. acidiscabies* was shown to be avirulent on potato tubers (Healy et al., 2000), suggesting

that thaxtomin A is a key pathogenicity determinant of scab-causing *Streptomyces* spp. The primary mode of action of this phytotoxin is still poorly understood, but several lines of evidence suggest that it functions as a cellulose synthesis inhibitor in the plant host (Scheible et al., 2003; Bischoff et al. 2009; Duval and Beaudoin, 2009).

The biosynthesis of thaxtomin A requires two nonribosomal peptide synthetases (NRPSs), TxtA and TxtB, which generate the cyclic dipeptide backbone (Healy et al., 2000), and a cytochrome P450 monooxygenase, TxtC, which carries out hydroxylation of the phenylalanine moiety at two positions (Alkhalaf et al., 2019; Healy et al., 2002). TxtA and TxtB each contain a single NRPS module consisting of four distinct enzymatic domains: an adenylation (A-) domain, a methylation (M-) domain, a thiolation (T-) domain [also known as a peptidyl carrier protein (PCP-) domain] and a condensation (C-) domain (Healy et al. 2000). The A-domain is considered the "gate keeper" as it is responsible for recruitment and activation of the amino acid substrate that is incorporated into the growing peptide chain. This is achieved by the specific binding of an amino acid followed by an adenylation reaction using Mg·ATP to generate an aminoacyl-AMP intermediate. Then, the A-domain catalyzes the loading of the activated aminoacyl-AMP residue onto the Ppant (4′-phosphopantetheine) arm of the T-domain (Gulick, 2009; Süssmuth and Mainz, 2017; Sieber and Marahiel, 2005). Previous work suggested that the A-domain of TxtA recruits the L-phenylalanine substrate, while the TxtB A-domain recruits the 4-nitro-L-tryptophan substrate (Johnson et al., 2009), and this was recently confirmed by Jiang and colleagues (Jiang et al., 2018). The Txt M-domains are responsible for *N*-methylation of the peptide backbone (Jiang et al., 2018), and C-domains normally contribute to the peptide bond

formation between the substrates that are tethered to the T-domains of adjacent modules (Finking and Marahiel, 2004; Sieber and Marahiel, 2005).

Some bacterial NRPSs also require a small auxiliary protein belonging to the MbtH-like protein (MLP) family for full function. Several biochemical studies suggested that the soluble expression of NRPSs depends on the presence of an MLP, suggesting a chaperone-like role to assist the proper folding of the NRPSs (Boll et al., 2011; Imker et al., 2010; Kaniusaite et al., 2020; McMahon et al., 2012; Zolova and Garneau-Tsodikova, 2012, 2014). In addition, some MLPs can influence the adenylation and aminoacylation activities catalyzed by NRPSs (Al-Mestarihi et al., 2014; Boll et al., 2011; Davidsen et al., 2013; Felnagle et al., 2010; Heemstra et al., 2009; Miller et al., 2016; Schomer et al., 2018; Zhang et al., 2010). Remarkably, a recent study demonstrated that the TioK NRPS A-domain substrate specificity can be affected by co-expressing the NRPS with non-cognate MLPs from diverse NRP biosynthetic pathways (Mori et al. 2018a). In the thaxtomin A biosynthetic gene cluster, a small gene *txtH* (198 bp) encoding an MLP is situated downstream of the *txtB* gene (Bignell et al., 2010). Our previous study revealed the importance of TxtH for thaxtomin A production and showed that TxtH is essential for promoting the soluble production of the A-domains from both TxtA and TxtB (Li et al., 2019). Additionally, we showed that phylogenetically distinct non-cognate MLPs from different organisms can functionally replace TxtH to promote the soluble expression of the Txt A-domains in *Escherichia coli* and restore the production of thaxtomins in *S. scabiei* at different degrees (Li et al., 2020). However, the impact of MLPs on the enzymology of the Txt NRPSs is currently unknown.

In this study, the goal was to investigate whether TxtH and non-cognate MLPs can influence the enzymatic activity of the Txt NRPS A-domains. To accomplish this, we co-expressed the TxtA A-domain (referred to herein as $TxtA^{A}$) with TxtH in *E. coli*. Chemical crosslinking experiments suggested that TxtH can associate with $TxtA^{A}$ in a 2:2 molar ratio. We purified $TxtA^{A}$ in the presence of TxtH from *E. coli* cells using affinity chromatography. Using a nonradioactive colorimetric assay, we attempted to detect the adenylation activity of the purified $TxtA^{A}$ towards its substrate L-phenylalanine; however, we were unsuccessful despite testing a variety of different assay conditions. Attempts were also made to express the AMT-domains of TxtA and TxtB (referred to herein as $TxtA^{AMT}$ and $TxtB^{AMT}$), but we were only able to recover truncated forms of the protein. The possible reasons for the unsuccessful detection of A-domain activity and future directions for overcoming these challenges are discussed.

4.3 Materials and Methods

4.3.1 Bacterial strains, culture conditions and maintenance

Escherichia coli BL21(DE3) *ybdZ:aac(3)IV* (Herbst et al., 2013) was used as the host for protein expression and purification, and strains NEB5α (New England Biolabs, Whitby, ON, Canada) and DH5α (Gibco-BRL) were used as general cloning hosts. *E. coli* strains were routinely cultivated at 37°C unless otherwise indicated. Liquid cultures were grown with shaking (200-250 rpm) in Luria-Bertani (LB) Lennox medium (Fisher Scientific, Ottawa, ON, Canada), SOB (Sambrook and Russell, 2001) or SOC (New England Biolabs), and solid cultures were grown on LB Lennox medium containing 1.5%

w/v agar. When required, the solid or liquid growth media were supplemented with antibiotics as described before (Li et al., 2019). *E. coli* strains were maintained at 4°C for short-term storage or at -80°C in 20% v/v glycerol for long-term storage (Sambrook and Russell, 2001).

Streptomyces scabiei 87.22 (Loria et al., 1995) was cultivated at 28°C with shaking (200 rpm) in stainless steel spring flasks containing trypticase soy broth (TSB; BD Biosciences, Mississauga, ON, Canada), or on potato mash agar (PMA; Fyans et al., 2015). The strain was maintained on agar plates at 4°C for short-term storage or at -80°C as spore suspensions in 20% v/v glycerol for long-term storage (Kieser et al. 2000).

4.3.2 DNA manipulation and primers

S. scabiei genomic DNA was isolated from TSB cultures as previously described (Li et al., 2019). Standard molecular biology procedures were implemented for all DNA manipulations performed (Sambrook and Russell, 2001). Restriction enzymes were purchased from New England Biolabs unless otherwise indicated. PCR was routinely performed using Phusion DNA polymerase (New England Biolabs) according to the manufacturer's instructions, except that 5% v/v DMSO was included in the reactions. All oligonucleotide primers used for PCR, cloning and sequencing were purchased from Integrated DNA Technologies (Coralville, IA, USA) and are listed in Supplementary Table 4.1. PCR products were purified using the Wizard SV Gel and PCR Clean-Up System (Promega Corporation, Canada). DNA sequencing was performed by The Centre for Applied Genomics (Toronto, Canada).

4.3.3 Construction of protein expression plasmids

Construction of the expression plasmids pACYCDuet-1/HIS$_6$-*txtA*^{A}, pET28b/HIS$_6$-*txtH* and pET28b/*txtH,* which enable overexpression of HIS$_6$-TxtA^A, HIS$_6$-TxtH and untagged TxtH in *E. coli,* respectively, was described in Li et al. (2019). To construct the plasmids for overexpression of the TxtA and TxtB amino acid sequences containing the AMT domains (TxtA^{AMT}, TxtB^{AMT}), the respective nucleotide sequences were PCR amplified using *S. scabiei* genomic DNA as template and using the primer pairs PL37/PL39 and PL40/PL42, respectively. The PCR products were cloned into the pGEM-T vector (Promega Corporation) as per the manufacturer's instructions, after which the inserts were released by digestion with *Eco*RI and *Hind*III and were cloned into similarly digested pACYCDuet-1 to give pACYCDuet-1/HIS$_6$-*txtA*^{AMT} and pACYCDuet-1/HIS$_6$-*txtB*^{AMT}. The cloned inserts in all constructed expression vectors were verified by DNA sequencing.

4.3.4 *In vivo* protein crosslinking

The co-expression of HIS$_6$-TxtA^A with HIS$_6$-TxtH was conducted as described in Li et al., (2019). The induced cultures (3 mL) were harvested by centrifugation (3500 rpm, for 10 minutes) and were washed twice with 25 mL phosphate-buffered saline solution (PBS; Vasilescu et al. 2004). Formaldehyde (9 mL, Sigma Aldrich, Oakville, ON, Canada) at different concentrations (0, 1, 2, 3 and 5% w/v) was added to the washed cells, and the cells were then incubated for 15 minutes at room temperature with gentle shaking. The crosslinking reaction was terminated by adding 1 mL of 1.25 M glycine in PBS and

incubating for 5 minutes at room temperature. The supernatant was removed by centrifugation (3000 rpm, for 5 minutes), and the cells were washed twice with 10 mL of ice-cold PBS before resuspending in 400 µL of ice-cold PBS with the addition of 1 × cOmplete EDTA-free protease inhibitor. The cells were then lysed by sonication as described in Li et al., (2019). Protein concentration was quantified by Bradford protein assay kit (Fisher Scientific) and soluble protein extracts (20 µg) were subjected to western blot analysis as previously described (Li et al., 2019).

4.3.5 Large-scale protein purification

Large-scale protein purification of HIS_6-$TxtA^A$ (co-expressed with HIS_6-TxtH or untagged TxtH) was conducted by culturing the *E. coli* strain harboring the corresponding expression plasmids in 50 mL of LB medium supplemented with 1% w/v glucose and the appropriate antibiotics overnight with shaking (200 rpm). The cells were then subcultured (1% v/v) into 1 L of LB containing antibiotics, after which the pET28b/HIS_6-*txtH*-containing strain was incubated at 37°C while the pET28b/*txtH*-containing strain was incubated at 30°C. Once the OD_{600} of the cultures reached 0.4-0.6, the cells were cooled to 16°C and were then induced with 1 mM isopropyl β-D thiogalactopyranoside (IPTG; Fisher Scientific, Canada). The cultures were further incubated at 16°C and 200 rpm for 48 hours (for pET28b/HIS_6-*txtH*) or 24 hours (for pET28b/*txtH*). Cells from the 1 L cultures were harvested by centrifugation (4000 rpm for 15 minutes at 4°C) and were then kept on ice. The cell pellets were resuspended in binding buffer (40 mL for HIS_6-$TxtA^A$ co-expressed with HIS_6-TxtH and 25 ml for HIS_6-$TxtA^A$ co-expressed with untagged TxtH) containing

10 mM Tris-HCl (Fisher Scientific, Canada), pH 7.8, 300 mM NaCl (Fisher Scientific, Canada), 10 mM imidazole (Sigma Aldrich, Canada), 5% v/v glycerol (Sigma Aldrich Canada) and 1× cOmplete EDTA-free protease inhibitor (Roche Diagnostics, Laval, QC, Canada). The cells expressing HIS_6-$TxtA^A$ with HIS_6-TxtH were lysed by sonication using a 3 mm diameter probe (70% intensity, 5 seconds on, 15 seconds off, until a total time of 2 minutes was reached, repeated 3–4 times), while the cells expressing HIS_6-$TxtA^A$ with untagged TxtH were lysed using a French Press (SLM Instruments Inc., USA) at 4°C. The cell debris was removed by centrifugation (10,000 rpm, for 15 minutes, at 4°C). Yielded supernatants (4 mL) were subsequently mixed with 1 mL of a 50% Ni-NTA slurry (QIAgen Inc, Toronto, ON, Canada) and were incubating on ice with gentle shaking (200 rpm) for 1hour. The lysate-Ni-NTA mixture was loaded into an Econo-Column® Chromatography Column [1.5 × 10 cm (diameter × length)] (Bio-Rad Laboratories, USA), and the unbound proteins were washed four times with 5 mL of binding buffer and three times with 10 mL of washing buffer at a flow rate of ~0.5 mL/min. Then, the HIS_6-tagged proteins were eluted using 0.5 mL of elution buffer per fraction for a total of 15 fractions. The washing and elution buffers were of the same composition as the binding buffer except that they contained imidazole at a final concentration of 20 mM and 250 mM, respectively. The collected fractions (10 µL) were analyzed for the presence of protein by pipetting onto a piece of filter paper and staining with Coomassie Blue stain (50% v/v methanol, 10% v/v glacial acetic acid, 0.1% w/v Coomassie Brilliant Blue G-250). The eluted fractions containing proteins were combined and desalted by dialysis (MWCO 3500 Da, Fisher Scientific, Canada) at 4°C overnight in a total of 3 L of desalting buffer (40 mM Tris-HCl

pH 7.8, 200 mM NaCl and 10% v/v glycerol). The buffer was changed three times over a total period of 17 hours. Different fractions were analysed by standard sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) on a 15% w/v gel and were visualized by staining the gel with Coomassie Blue stain. The desalted samples were then aliquoted, and flash frozen in 95% ethanol and dry ice, prior to storage at -80°C. Protein concentration was determined using a Bradford protein assay kit (Fisher Scientific) as per the manufacturer's instructions and using bovine serum albumin (BSA) as the standard.

4.3.6 Small-scale protein purification

The co-expressions of HIS_6-$TxtA^A$ with HIS_6-TxtH or untagged TxtH, and HIS_6-$TxtA^{AMT}$ with HIS_6-TxtH, were conducted as previously described (Li et al., 2019) with some modifications. The cells were induced with 1 mM IPTG, after which they were further incubated at 16°C at 200 rpm for 24 hours (HIS_6-$TxtA^A$ with HIS_6-TxtH) or 48 hours (HIS_6-$TxtA^A$ with untagged TxtH and HIS_6-$TxtA^{AMT}$ with HIS_6-TxtH). The cells were harvested by centrifugation (4000 rpm, for 15 minutes, at 4°C), and the pellets were resuspended in BugBuster™ protein extraction reagent (5 mL per gram of wet cell paste, Novagen, Canada) containing 1 × cOmplete EDTA-free protease inhibitor. 50% Ni-NTA slurry (800 µL) was subsequently added to 3 mL of cell lysate and incubated on ice with gentle shaking (200 rpm) for 1 hour. The unbound proteins were washed twice with 500 µL of binding buffer and three times with 500 µL of washing buffer, and the HIS_6-tagged proteins were eluted three times using 200 µL of elution buffer. The supernatant was removed by centrifugation

(3000 rpm, for 10 seconds). Different fractions were analyzed by SDS-PAGE on a 15% (w/v) gel as described above.

4.3.7 Detection of HIS_6-$TxtA^{AMT}$ and HIS_6-$TxtB^{AMT}$ by western blot analysis

The co-expression of HIS_6-$TxtA^{AMT}$ or HIS_6-$TxtB^{AMT}$ with HIS_6-TxtH was conducted as previously described in Li et al., (2019) with some modifications. The cells were induced with 1 mM IPTG and were further incubated at 16°C and 200 rpm for 20-48 hours (HIS_6-$TxtA^{AMT}$ with HIS_6-TxtH) or 16-48 hours (HIS_6-$TxtB^{AMT}$ with HIS_6-TxtH). Cells from 1 mL of culture were lysed with 200 µL of BugBuster reagent with 1× cOmplete EDTA-free protease inhibitor. Soluble fractions (20-60 µg) of HIS_6-$TxtA^{AMT}$ and HIS_6-$TxtB^{AMT}$ in the presence or absence of HIS_6-TxtH were assessed by western blot analysis as previously described (Li et al., 2019).

4.3.8 Colorimetric assay for detection of adenylation activity

The non-radioactive colorimetric enzyme assay described by McQuade *et al.* (2009) was used to assess the adenylation activity of the purified HIS_6-$TxtA^{A}$. The reaction (100 µL) contained 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ (Fisher Scientific, Canada), 1 mM dithiothreitol (DTT; Fisher Scientific, Canada), 5% v/v glycerol, 0.1-0.5 mM ATP (Sigma Aldrich Canada), 0.4 U/mL inorganic pyrophosphatase (Sigma Aldrich Canada), 0.25-2 µM of purified protein and 0.5-6 mM of amino acid substrate (L-alanine or L-phenylalanine). Enzyme reactions were performed in 96-well plates and were initiated with the addition of ATP. After incubation at 25°C for 10-30 minutes, the reactions were stopped

by adding 75 µL of molybdate/malachite green reagent (Cell Signaling Technology, Inc.) and were incubated at 37°C for 20-60 minutes to allow green color development. The absorbance at 630 nm was measured using a Synergy H1 microplate reader (BioTek Instruments Canada). Heat-inactivated proteins and the dialysis buffer were used as negative controls. To test sources of phosphate contamination, 100 µL of molybdate/malachite green reagent was added to 5 µL of protein, and the mixture was incubated at 37°C for 20 minutes. The color development in each reaction was visually assessed.

4.3.9 Bioinformatics analysis

Identification of the AMTC domains within the TxtA and TxtB amino acid sequences was performed using the Pfam database version 33.1 (http://pfam.xfam.org/; El-Gebali et al. 2019). Amino acid sequence alignments of TxtA and TxtB were generated using ClustalW through the EMBL-EBI sequence alignment tool (https://www.ebi.ac.uk/Tools/msa/clustalo/). The *in silico* structures of the *S. scabiei* $TxtA^{AM}$ and TxtH were constructed using SWISS-MODEL as described in Li et al. (2020). The crystal structure of the TioS NRPS (PDB ID: 5wmm_1; Mori et al., 2018b) from *Micromonospora* sp. ML1 was used as the template for the $TxtA^{AM}$ model. The generated models were evaluated by different parameters using the SWISS-MODEL webserver (Supplementary Table 4.2; https://swissmodel.expasy.org/) and visualized using PyMOL (DeLano, 2002).

4.4 Results and Discussion

4.4.1 Purification and adenylation activity of the TxtA A-domain

Previously, we demonstrated that TxtH and non-cognate MLPs have the ability to promote the soluble production of the TxtA and TxtB A-domains in *E. coli* with varying efficiencies (Li et al., 2020), but the impact of these MLPs on the enzymology of the Txt NRPSs has not been investigated. To address this, we expressed $TxtA^A$ with an N-terminal HIS_6 tag (HIS_6-$TxtA^A$) in the presence of HIS-tagged TxtH (HIS_6-TxtH) in the *E. coli* BL21(DE3)*ybdZ:aac(3)IV* expression host. To confirm that the two proteins are able to interact, we conducted an *in vivo* chemical crosslinking experiment using formaldehyde, which allows the crosslinking of proteins within 2 Å in the cell (Herzberg et al., 2007). As shown in Figure 4.1, a protein band was detected with increasing concentration of formaldehyde and was consistent with the formation of a heterotetrametic $(TxtH)_2$-$(TxtA^A)_2$ complex. MLP-NRPS complexes with a 2:2 molar ratio have been described before (Boll et al., 2011; Mori et al., 2018a), though other studies also report a 1:1 molar stoichiometry after purification of such complexes (Davidsen et al., 2013; Felnagle et al., 2010; Mori et al., 2018a; Tarry et al., 2017). Particularly, Mori and colleagues performed sedimentation equilibrium experiments with twelve different MLP-TioK complexes. Their result shows that eight complexes including the one with its native MLP TioT display a 2:2 molar stoichiometry while the remaining four form 1:1 complexes (Mori et al., 2018a). The reason why some MLP-NRPS complexes dimerize whereas others do not is still obscure.

To purify the proteins, a large-scale (1 L) culture was induced with IPTG, after which the cells were harvested, lysed by sonication, and the soluble protein extract was

subjected to Ni^{2+}-affinity purification. SDS-PAGE analysis revealed two major bands with the expected molecular weight for HIS_6-$TxtA^A$ (~62 kDa) and HIS_6-TxtH (~9.5 kDa) in the eluted fractions (Figure 4.2). These fractions were then combined and desalted prior to activity analysis.

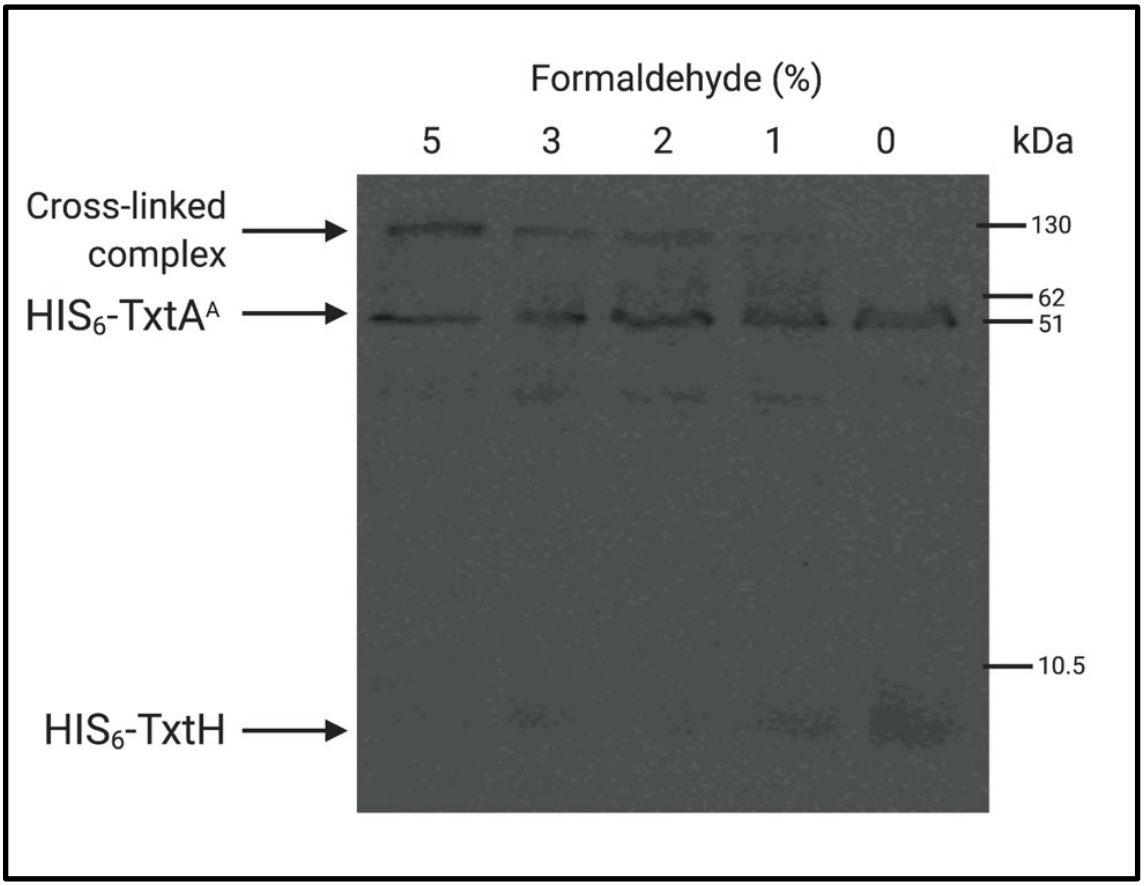


Figure 4.1 Detection of TxtH-$TxtA^A$ complex formation by *in vivo* chemical crosslinking. Cells expressing HIS_6-TxtH and HIS_6-$TxtA^A$ were treated with different concentrations (0, 1, 2, 3, 5% w/v) of formaldehyde, after which the cells were lysed, and the soluble extracts were subjected to western blot analysis using an anti-HIS_6 antibody. The bands corresponding to HIS_6-TxtH (expected MW = 9.5 kDa), HIS_6-$TxtA^A$ (expected MW = 62 kDa) and the cross-linked heterotetrametic $(TxtH)_2$-$(TxtA^A)_2$ protein complex (expected Mw = 143 kDa) are shown.

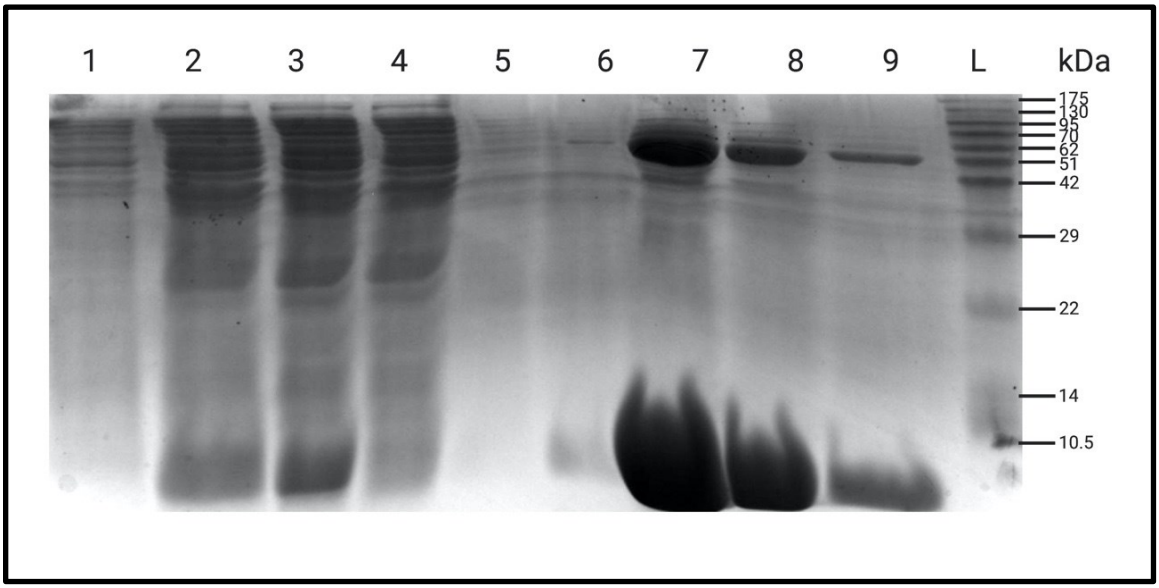


Figure 4.2 Purification of HIS_6-$TxtA^A$ and HIS_6-TxtH. Protein samples collected at different stages during the purification process were analyzed by SDS-PAGE on a 15% w/v polyacrylamide gel and were visualized using Coomassie Blue stain. Lane 1: total proteins before induction with IPTG; lane 2: total proteins after induction with IPTG; lane 3: total soluble proteins following cell lysis; lane 4: column flow through; lane 5: column wash with binding buffer; lane 6: column wash with wash buffer; lanes 7-9: different fractions of eluted proteins. The bands at ~62 kDa and at <10.5 kDa correspond to the expected molecular weight for HIS_6-$TxtA^A$ (62 kDa) and HIS_6-TxtH (9.5 kDa), respectively. Lane L: PiNK Plus Prestained Protein Ladder (FroggaBio Inc.).

It has been demonstrated that TxtA utilizes L-phenylalanine as its preferred amino acid substrate (Jiang et al., 2018). To detect the adenylation activity of $TxtA^A$ and confirm its substrate specificity, we used a previously described colorimetric molybdate/malachite green phosphate assay (McQuade et al., 2009; Xia et al., 2012). During the adenylating reaction (Figure 4.3), the amino acid substrate reacts with ATP to form the aminoacyl-AMP intermediate and pyrophosphate (PP_i) in the presence of Mg^{2+}. Addition of inorganic pyrophosphatase allows for conversion of the released PP_i into orthophosphate (P_i), which then reacts with the molybdate in the molybdate/malachite green reagent to form a negatively charged phosphomolybdate complex. The added reagent also serves to stop the

enzyme reaction by reducing the pH of the solution. The phosphomolybdate complex then binds to the cationic malachite green dye, leading to a change of pH in the solution and a subsequently a change of the dye from pale yellow to green/blue. The color development is monitored and quantified by measuring the absorbance at 630 nm (Cogan et al. 1999; Mok and Edwards, 2005). This absorbance is proportional to the amount of P_i in the solution and reflects the adenylation activity of the A-domain (McQuade et al., 2009).

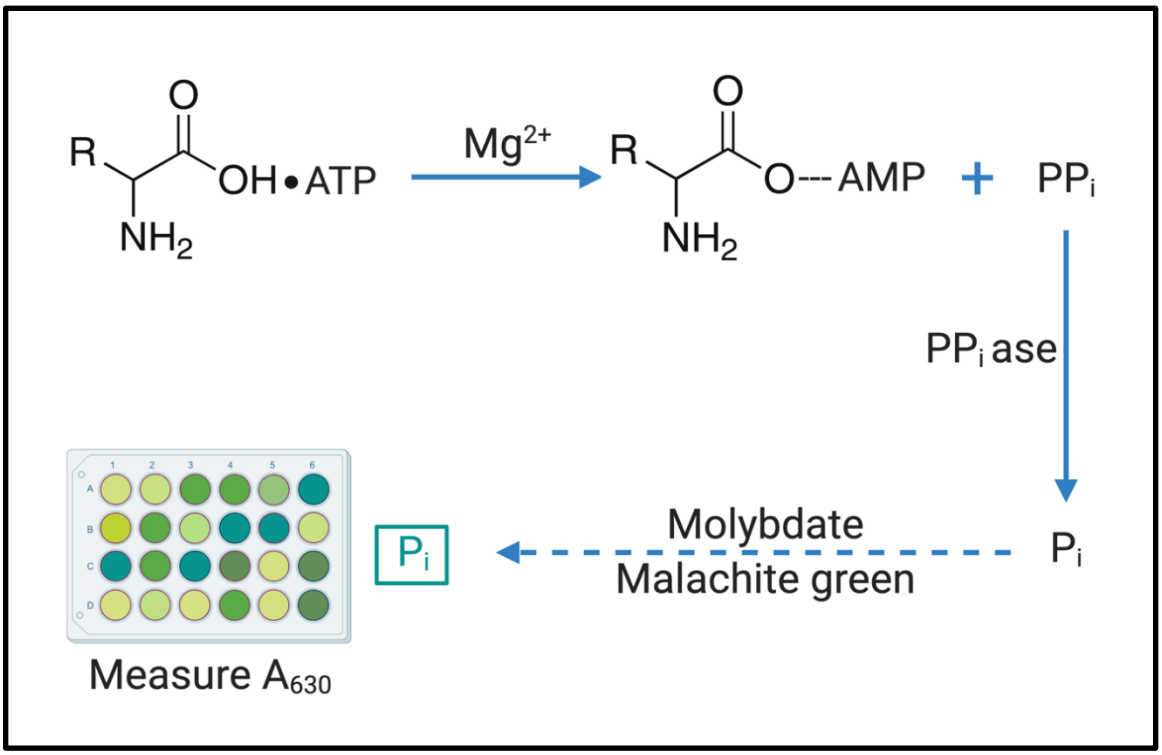


Figure 4.3 The enzymatic mechanism for detecting the adenylation activity of A-domains by the molybdate/malachite green phosphate assay. The adenylating reaction initiates with the binding of the amino acid substrate and ATP to the A-domain in the presence of Mg^{2+}. The A-domain catalyzes the formation of the aminoacyl-AMP intermediate, and pyrophosphate (PP_i) is released. Inorganic pyrophosphatase (PP_iase) then converts the PP_i into orthophosphate (P_i) molecules, which react with molybdate to form a phosphomolybdate complex. The complex binds to the malachite green stain (indicated by P_i in green box), leading to a colour change that can be measured by absorbance at 630 nm using a microplate reader.

To characterize the activity of $TxtA^A$, we first prepared reactions containing 0.25-1 μM of co-purified HIS_6-$TxtA^A$/HIS_6-TxtH and 0.5 mM of the L-phenylalanine substrate. The reactions were initiated by adding 0.25 mM ATP, and after incubating at 25°C for 10 minutes, the molybdate/malachite green reagent was added to stop the reactions, and the reactions were further incubated at 37°C for 10 minutes to allow colour development. Heat-inactivated proteins and dialysis buffer were used in place of the purified proteins as negative controls. When observing the reactions, those containing the purified proteins displayed a similar light-yellow color as the negative controls, implying that no adenylation activity was detected. We therefore made the following changes to the reaction conditions: ATP and L-phenylalanine concentrations were increased to 0.5 mM and 5 mM, respectively; incubation time of the reactions was increased to 30 minutes prior to the addition of molybdate/malachite green reagent; incubation time was increased to 1 hour following molybdate/malachite green reagent addition. Although the longer incubation times allowed for better color development, no significant color difference was observed between the reactions with the purified proteins and the negative controls.

During the protein purification, the *E. coli* cells were sonicated multiple times to allow for complete cell lysis, and it is possible that the proteins were damaged during this process (Brown and Audet, 2008). In addition, the long induction time (48 hours) for expressing $TxtA^A$ may have had an impact on its enzymatic activity, even though more soluble $TxtA^A$ is produced when the cells are incubated for 48 hours following induction versus 24 hours (data not shown). Mori and collogues co-expressed the NRPS TioK with its MLP partner TioT by inducing the cells with 1 mM IPTG at 25°C for 16-18 hours (Mori et al., 2018a). Therefore, we subsequently co-expressed HIS_6-$TxtA^A$ with HIS_6-TxtH using

a small volume (50 mL) of culture and inducing the cells at 16°C for 24 hours. The cells were lysed using the Bugbuster reagent, which is a mixture of non-ionic detergents formulated for gentle disruption of *E. coli* cells without denaturing soluble proteins (Grabski et al. 1999). The purified proteins (Supplementary Figure 4.1) were analyzed immediately in the assay to avoid any potential effects of freezing/thawing on the enzyme activity of $TxtA^A$. In addition, the expression strain [BL21(DE3)*ybdZ:aac(3)IV*] lacking the expression plasmids was also included in the purification process and was used as a negative control in the enzyme assay to rule out any potential adenylation activities detected from non-specific bound proteins. Despite these changes, no adenylation activity was observed for $TxtA^A$ in the enzyme assay under the conditions used.

Previous studies suggested that the A-domain requires an equimolar molar ratio with its MLP partner to stimulate its adenylation activity (Boll et al., 2011; Davidsen et al., 2013). We speculated that the co-purification of HIS-tagged $TxtA^A$ and TxtH would result in an excess of the MLP, and this might affect the precise concentration of the protein complex used in the activity assay (Mori et al., 2018a). Additionally, it is possible that the enzyme activity of $TxtA^A$ may have been affected by the presence of the N-terminal HIS_6 tag on TxtH, though we previously showed that the HIS_6 tag does not impact the ability of TxtH to promote soluble production of the TxtA and TxtB A-domains in *E. coli*. (Li et al., 2019). Therefore, we co-expressed HIS_6-$TxtA^A$ with untagged TxtH. The cells were incubated for 24 hours following IPTG induction, after which the harvested cells were lysed using a French press in order to minimize damage to the soluble proteins. SDS-PAGE analysis of the purified proteins (Figure 4.4) indicated the expected band for HIS_6-$TxtA^A$ at approximately 62 kDa, while the untagged TxtH (~7.4 kDa) was not observed in the

eluted fractions. The reason why TxtH was not detected requires further investigation, but it is possible that the amount of TxtH that was co-purified with HIS_6-$TxtA^A$ may not have been sufficient for detection by SDS-PAGE. Similar findings have been reported in other studies that co-expressed an untagged MLP with a HIS-tagged NRPS partner (Felnagle et al., 2010, Mori et al., 2018a).

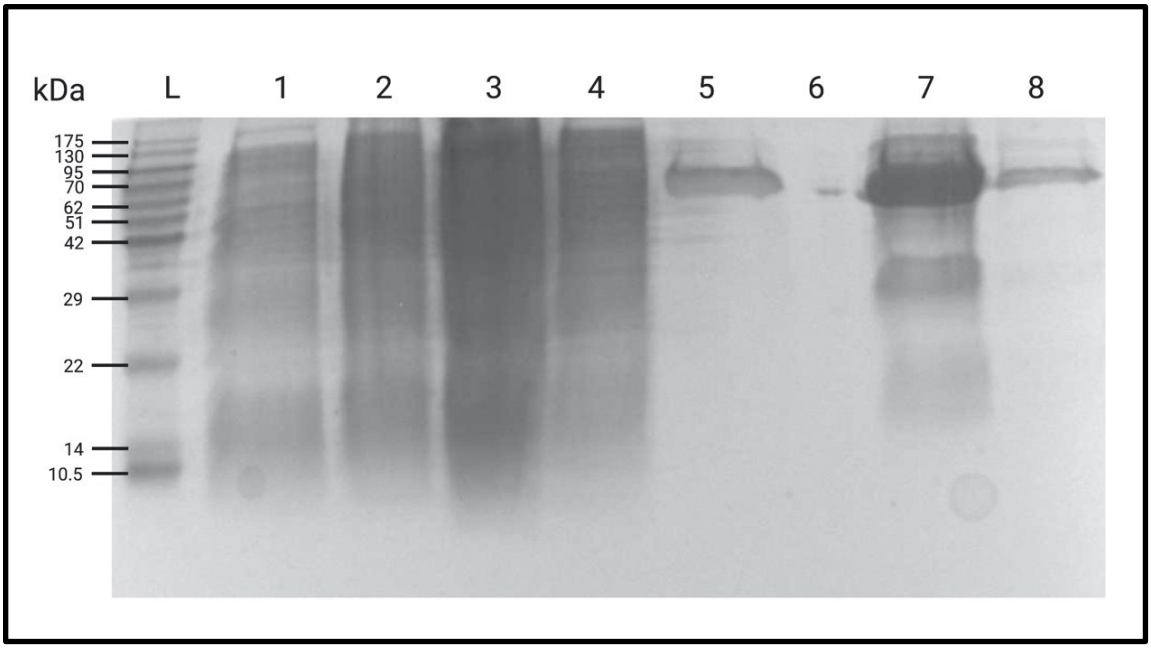


Figure 4.4 Purification of HIS_6-$TxtA^A$ co-expressed with TxtH in *E. coli*. Protein samples collected at different stages during the purification process were analyzed by SDS-PAGE on a 15% w/v polyacrylamide gel and were visualized using Coomassie Blue stain. Lane 1: total proteins before induction with IPTG; lane 2: total proteins after induction with IPTG; lane 3: total soluble proteins following cell lysis; lane 4: column flow-through; lane 5: column wash with binding buffer; lane 6: column wash with wash buffer; lanes 7 and 8: different fractions of eluted proteins. The band at ~62 kDa corresponds to the expected molecular weight of HIS_6-$TxtA^A$. Lane L: PiNK Plus Prestained Protein Ladder (FroggaBio Inc.).

To assess the adenylation activity of HIS_6-$TxtA^A$ co-expressed with untagged TxtH, we prepared a reaction containing 1 μM of purified protein and 0.5 mM of L-phenylalanine

or L-alanine, the latter of which was included as a negative control for the assay. The addition of 0.25 mM ATP initiated the reaction, and after 30 minutes incubation, the reactions were stopped with the molybdate/malachite green reagent and were further incubated for 20 minutes. The absorbance at 630 nm was measured for each reaction, followed by subtraction of the reading from the negative control reaction that contains dialysis buffer in place of the purified protein. As shown in Figure 4.5, there was no obvious difference in the absorbance readings from the reactions containing L-phenylalanine and the negative control reactions containing L-alanine, indicating that no adenylation activity was observed for HIS_6-$TxtA^A$ co-expressed with untagged TxtH under the assay conditions tested in the current study.

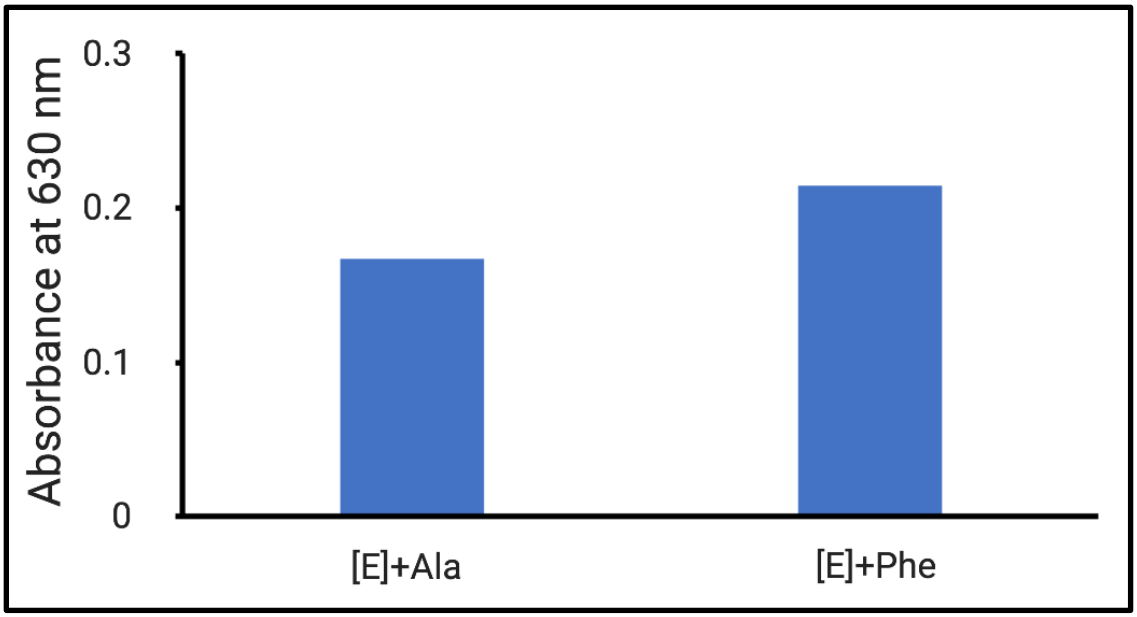


Figure 4.5 Quantification of adenylation activity of HIS_6-$TxtA^A$ co-expressed with TxtH. Reactions were prepared using 1 µM of purified protein ([E]) and 0.5 mM of L-alanine (Ala) or L-phenylalanine (Phe) as substrate. The reactions were initiated by addition of 0.25 mM ATP. Each reaction was performed once, and the columns represent the absorbance reading at 630 nm following subtraction of the reading from the negative control (reaction containing dialysis buffer in place of protein).

As discussed by Bisswanger (2014), substrates should be saturating in the enzyme reaction, and the concentration of the substrate should be at least 10-fold higher than the enzyme K_m (the concentration of substrate at half maximum velocity of the enzyme). Additionally, the concentration of the enzyme should be as low as possible but enough to observe the reaction (Bisswanger, 2014). The work conducted by McQuade and colleagues demonstrated that the K_m of the NRPS NcpB A-domain against its preferred substrate L-isoleucine is 0.6 mM, and they measured the adenylation activity of the purified protein (0.5 μM) against different amino acid substrates at a concentration of 6 mM (McQuade et al., 2009). Therefore, we modified our assay by decreasing the concentration of purified protein to 0.5 μM and increasing the concentration of the L-phenylalanine to 6 mM, while the other components of the reaction remained the same as described above. The results showed no obvious difference in the absorbance readings between the reaction containing dialysis buffer and that containing purified protein (Supplementary Figure 4.2). In addition, a control reaction in which the pyrophosphatase was not added was also found to produce a green color following incubation, suggesting that there was contaminating P_i in the reactions. This is considered a common cause of high background signal in the molybdate/malachite green phosphate assay (Kittilä et al., 2015). We tested each component of the reaction for P_i contamination and found that the ATP (0.5 mM) reagent enabled a dark green color development when mixed with the molybdate/malachite green reagent, suggesting that the ATP was the source of the background P_i. We therefore reduced the amount of ATP (0.1 mM) in the reaction; however, we were still unable to detect any adenylation activity for HIS_6-$TxtA^A$ when co-expressed with untagged TxtH. It has been

previously demonstrated that ATP can be hydrolyzed under the acidic conditions generated when the molybdate/malachite green reagent is added, and this effect may be overcome by the addition of sodium citrate after addition of the reagent (Lanzetta et al., 1979; Biswas et al., 2013). This is modification could be included in future studies to minimize P_i release by the nonenzymic hydrolysis of ATP.

Overall, we were unable to detect any enzymatic activity of $TxtA^A$ towards its phenylalanine substrate using the molybdate/malachite green assay. As no positive control was included in our tests, we cannot rule out that the assay was not working properly in our experiments. It is noteworthy, though, that this assay has been successfully used by another researcher in our laboratory to characterize the adenylation activity of a different enzyme (Bown, 2018). Possibly, the assay could be further optimized using the *E. coli* EntF NRPS, which is known to select L-serine as its substrate, and whose adenylation activity has been previously characterized (Felnagle et al., 2010).

4.4.2 Bioinformatics and structural analysis of TxtA and TxtB

Since we could not detect any adenylation activity for the purified $TxtA^A$ protein, we re-examined the amino acid sequences of TxtA and TxtB in an effort to identify whether any potential sequence motifs critical to adenylation function may be missing in our fusion construct. The NRPS domains within TxtA and TxtB were detected using the most updated Pfam version 33.1 (El-Gebali et al. 2019; Supplementary Table 4.3). For TxtA, four enzymatic domains were identified as significant Pfam-A matches (E-values < 0.001): AMP-binding enzyme (A-domain, PF00501), methyltransferase domain (M-domain,

PF08241), phosphopantetheine attachment site (T-domain, PF00550) and condensation domain (C-domain, PF00668). In addition, two small AMP-binding enzyme C-terminal domains (PF13193) flanking the M-domain, were identified as insignificant Pfam-A matches (Figure 4.6). For TxtB, the AMTC-domains as well as one AMP-binding enzyme C-terminal domain located at the N-terminus of the M-domain were identified as significant Pfam-A matches, and one other AMP-binding enzyme C-terminal domain downstream of the M-domain was recognized as an insignificant Pfam-A match (Supplementary Table 4.3). Typically, an A-domain is composed of approximately 550 amino acid residues and contains a large N-terminal domain (also known as the A_{core}-domain, ~50 kDa) and a small C-terminal domain (also known as the A_{sub}-domain, ~10 kDa) as well as a hinge region of about five residues that connect the two subunits. A conformational orientation occurs between the two subdomains when the A-domain switches its state between the adenylation and thiolation reactions. In the adenylation state, the amino acid substrate together with Mg·ATP binds to the A_{core}-domain at the region closest to the interface between these two subdomains (Süssmuth and Mainz, 2017). In the case of TxtA and TxtB, the M-domain is predicted to be embedded in between the two small AMP-binding enzyme C-terminal domains (Figure 4.6). Besides a few stand-alone methyltransferases that modify the NRP during and after the assembly, most M-domains are found to be embedded at different positions within the A-domain (Labby et al. 2014; Lundy et al., 2020; Süssmuth and Mainz, 2017). Recently, the first crystal structure of the A-domain TioS from *Micromonospora* sp. ML1 has be recently unravelled, and this demonstrates that the embedding of the M-domain within the C-terminal region of the A-domain does not affect the normal folding of the A-

domain. Furthermore, biochemical studies indicated that TioS is able to perform both the adenylation and methylation activities (Mori et al. 2018b). Johnson and colleagues previously reported the accumulation of *N*-methyl-4-nitrotryptophan in a *txtA* mutant of *S. scabiei* (Johnson et al., 2009). Therefore, we postulate that the Txt NRPSs, which are both predicted to harbour a M-domain embedded at the C-terminal end of the A-domain, are also able to catalyze both adenylation and methylation of their amino acid substrates.

```
                    ┌──►
TxtA     1  ----------MSHLTGEDLPEGALATTWPSLLEARVADTPDAIALVAGDTALTYAQFNA    49
TxtB     1  MSMLPPGRSRTTASPAGAQAGPEFTPGLWGRLFEARVDAAPESTAINSASERLSYAELNR    60
                     :  :* :          *  *:****  :*:: *: :..  *:**::*
                                                                 A1
TxtA    50  RANRLARWLKYLGAGPERSVGLVLGRSADFFLCATAVLKCGAAYLPLDPNYPVERLSFMA   109
TxtB    61  RANRLARLLIARGAGPESLVGLALPRSTDFVVAVAAVLKSGAGYFPMDPDYPPQRLAFML   120
            ******* *   *****  ***.* **:**.:..:****.**.*:*:**:** :**:**
                                                     A2
TxtA   110  RDAAPVVLVTTSDVRGDLLGQLPTGSLVVLDDEATEDVLRRLPDHDMEDGERLEPLRPAS   169
TxtB   121  ADAAPMLVLTRSDIEPELPAEAAS-RTVVLDDPAVVRTLADCSAADVADDERGAPLRTRH   179
             ****::::* **:. :* .:  :   ***** *.  .*      *: *.**  ***
                                                                 ** *
TxtA   170  PAYIIYTSGSTGIPKGVVVTHQGVASLIATQRRRLAVTGASRVLAFSSPSFDASFWEMSM   229
TxtB   180  PAYVIYTSGSTGTPKGVVLTHHGIASLVGSHARDLGIGPSSRLLLFSSPSFDGAFWDVSM   239
            ***:******** *****:**:*:***:.:: * *.:  :**:* *******.:**::**
                A3                                                A4  * *
TxtA   230  ALLAGAALVVGRPGRLLPDAELAALIADHGVTHVTLPPSVAGALGPDMLPPSVTLVVAGE   289
TxtB   240  ALLTGATLVVAPRERLLPGPEFSALAAEEGITHFTLPASTLAALPDGALPAGATVVNVGE   299
            ***:**:***.   ****. *::** *:.*:**.*** *. .**   .  ** ..*:* .**
                     *         **                                   ●●●●
TxtA   290  ACPAALVQRWRPHRTMVNAYGPTESTVCATMSDPLADDVAPPVGRAVDGTRIHVLDDRLA   349
TxtB   300  ACNSELVRRWSPGRLLVNAYGPTESTVSATMSGPLAGAGIPPIGRPLSDTRIHVLDERLR   359
            ** : **:** * * :*********.****.***.   **:** :..*******:**
                                A5
            ●  ●     ●●●●●●●●  ●●●●●●●●●   ●●●●●  ●
TxtA   350  PVVPGAVGEIYIAGHSLARGYLERPGLTAQRFVADPFGPAGSRMYRSGDLGRWTRSGDLE   409
TxtB   360  PVPPGAVGEIHIAGAGLARGYLGRPALTAERFVADPFGTPGERMYRTGDRVRVRDDGQLE   419
            ** *******:*** .****** **.***:********  *.*****:**  *   .*:**
                  ●          A6                          A7
TxtA   410  FVGRADDQVKVRGFRIEPGEIESVIAGCRGVRQAAVVLREDRPGEPYLAAYVIPENAAAD   469
TxtB   420  FVGRVDDQAKIRGFRVEPGEVEAVLRDHPEVAQAAVVVREDTPGDQRLVAYVVPDHPAVR   479
            ****.***.*:****:****:*:*: .   * *****:*** **:  *.***:*:: *.
                     A8
TxtA   470  EAAGEEPDGQLDAWRRLYDDLYGRADTADFGEDFSGWVSSYGGRPI--EGMREWREQTVR   527
TxtB   480  QA-DDTTSEHVEEWQRLYDEVYSAVGALPLGEDFSGWNSTYDGEPIPVPQMQAWRDATVD   538
            :* .:  . ::: *:****::*. ..:  :******* *:*.*.**    *: **: **
                                   TxtA^A ◄┐
TxtA   528  QIRELAPRRVLEIGCGSGLLLSQLAGDCESYWGTDISGALIERLRGQVAERPGLADRVVL   587
TxtB   539  SIRALRPRRVLEIGVGTGLLLSRLAGDCEAYWATDFSAEVIETLGKKVDVDPVLREKVHL   598
            .** * ******** *:*****:******:**.**:*. :** *  :*   * * ::* *

TxtA   588  HQLSAHELGSLPSGGFDTVVLNSVIQYFPSGDYLFDLLREVSRLLVPGGAVFLGDVRNLR   647
TxtB   599  LHGPAHDLPGLPEGYFDTVVLNSVIQYFPSADYLVSVLREAARLLAPGGRVFVGDIRHLR   658
             :  **:* .**.* ***************.***..:***.:***.*** **:**:*:**

TxtA   648  LLRTFHAGGLLAAATHTD-TPQTVCAAIDRAMAQEKELLVDPEFFTTAVGALPGMTLESC   706
TxtB   659  LLRPLRSAVRLRSATRREASASAVRAAVEQDLVDEKELLLDPAFFAAVPRWIPQLRGVRT   718
            *** :::.  * :**: : : .:* **::: :.:*****:** **::.   :* :

TxtA   707  TLKRGGYDNELSRYRYEVVLRKHAGPADDTGPTDDAGPVVRLRWDGEMASLADVADRLRR   766
TxtB   719  AVQRGTHHNELTRYRYDAVLIKEPVETGTAAP-----DAQTLTWGTDVSGLQELSGLLAR   773
            :::** :.***:****:.** *.   :. :.*      .  * *. :::.* :::. * *

TxtA   767  GKPERLCVTGIPNGRVAGEHAATLALFDRRPLHEVLSLGQ-APAGVAPEDLRRLGAELGY   825
TxtB   774  T-RTSLLLRGVPNSRILGEASAATALTTARSLDEPLRLLQEPAAGIDPEELHALGGGAGC   832
                 * : *:**.*: ** :*: **   * *.* * * *   **: **:*: **.  *

TxtA   826  RVDCTWSSEDDALIDASFTRAGALVPRPAPRTDAEPDGFSPARFTNRPAFARPDSQTMAS   885
TxtB   833  EVHLTWSAQDPTRLDACFTPVGGEPGAVPLAESADSGRTSPGDHANQPTTHRTGNALMGK   892
            .*. ***::* : :**.** .*.        .*: .  **. .:*:*:  * ..  *..
                                                  *
TxtA   886  LPGQVAAKLPAFMVPEVFVPLDRLPVTVNGKLDRGALPRPRRAAHASGRPPRTAREEVLA   945
TxtB   893  LPGYLAARLPAYLRPSAVVRIASLPLTVNGKLDRTALPRPALFPRADGQAPRTPREEILA   952
            *** :**:***:: *...* :  **:******** *****    :*.*: *** ***:**
                    A9                    A10
TxtA   946  AIFADVLATADVTADSDFFAVGGNSLLATRLAAEVRRRLNTEMPLSWLFESPTVGALAAR  1005
TxtB   953  NLFADVLGLPGVPRDADFFALGGNSLLATRLVGRIAKHLEVDVPIAWIFETPTVEGLAGR  1012
             :*****.  .*  *:****:**********...: ::*:.::*:::*:**:*** .**.*
                   TxtA^AMT ◄┐
TxtA  1006  FDAGDEARPLPVPSEYASGSTAPLSAQQMQMWHEYR-RSLCRDMFNVPLSQRLTGAVDAE  1064
TxtB  1013  TAPASRLRPLLLCRDE-NHAAVPLSHSQYGMWFINQLGGPASRIYNVPYCLRITGRVDTG  1071
               ... *** :  :  . ::.*** .*  **.  :  .  .  ::*** . *:** **:

TxtA  1065  ALRAALADVVTRHVPLRTLVQDDGSGPCAVITEATADDIPWTETRTTPERLSEDLAHAAR  1124
TxtB  1072  ALRTALDDVVARHEPLRTVFPDDGDGPRQRVLAPEDAAVVLHETDAAEDRLAGHLARAAA  1131
            ***:** ***:** *****:. ***.**   :       :   ** :: :**: .**:**

TxtA  1125  RHFDLETEIPLRAVLFTLGPDESVLLLVMHHIAADGWSFGPLLEDLVRAYRARTEGRAPQ  1184
TxtB  1132  EPFELRTDLPLRARLFRHGQDRYTLLLLMHHITVDAWSLAPLTADLAHAYRARLGQRAPQ  1191
            . *:*.*::**** **  * *. .***:****:.*.**:.**  **.:*****   ****

TxtA  1185  WEPLSFGYLDYVAWQRRLLGATDDPSDVALRQAEYWRKTLHGADDRPVLETDSPAPAQQD  1244
TxtB  1192  WQPLPVHYRDYAVWHNEQAAEAQDRGSGFGRQLAFWERTLRGLPVETRLPADRSRPARPT  1251
            *:** . * **..*:..  . ::* ..   **  :*.:**:*   .  * :*   **:

TxtA  1245  FAGRSLDLPLEVGGHRVLTAAAREHGVTVFMILHAALVALLARRGAGGDVTVVTAVAGRT  1304
TxtB  1252  YRGGTVHTHVEASLHQELLNCARETGATLFMVLHAALAALLTRLGGGTDIVVGTAAAART  1311
            : * ::.  :*.. *: *  .*** *.*:**:*****.***:* *.* *:.* **.*.**

TxtA  1305  DTQFEPLVGLFANTLALRTDTSGNPTFRELLDRVRVTDLGAYAHQDLLFERLADVPPPQV  1364
TxtB  1312  DPALDDLVGLFANSVVLRVDTSGDPTFRTLLARTRAVDLDAFTHQEVPFDQVVDRVNPAR  1371
            *  :: *******::.**.****:**** ** *.*..**.*::**:: *:::.*   *

TxtA  1365  ----------SLVLRTVAAPPADLPGLTISPGPRPASESARYPVLWTVEHLA-SAADGGT  1413
TxtB  1372  HPARHPLYQTALVLHAPPGDGHRADSVTLTPEPPPNTGTARFDLMFNWDESRDSAGLAQG  1431
                      :***::  .       .:*::* * * : :**: :::. :.   **. .

TxtA  1414  LRSHIQYQSGLLRDDTVVRLAQQYEVVLSLLLKDPDLRVQDLPLQ------------  1458
TxtB  1432  LTGRTEYSSDLFSQETVELLLERYLLLLSAAVRDPDARLHTLDILTEPERRAFSPRP  1488
            * .: :*.*.*: ::**  * ::* ::**  ::*** *:: * :
```

Figure 4.6 The predicted domain organization of TxtA and TxtB from *S. scabiei*. The NRPS domains were identified using the Pfam 33.1 database and highlighted by different colors. Red: AMP-binding enzyme (A-domain, PF00501); pink: AMP-binding enzyme C-terminal domain; green: methyltransferase domain (M-domain, PF08241); yellow: phosphopantetheine attachment site (T-domain, PF00550); purple: condensation domain (C-domain, PF00668). The conserved motifs (A1-A10) of the Txt A-domains are highlighted by lines below the sequences, and the motifs involved in substrate binding are indicated by green lines. The ten residues predicted to line the substrate binding pocket of the A-domains are highlighted by the asterisks. Orange circles indicates the predicted residues at the interface between TxtA/TxtB and TxtH (Li et al., 2020). The start and end sequences included in the HIS_6-$TxtA^A$ and HIS_6-$TxtA^{AMT}$ expression constructs are indicated by arrows.

We also examined the presence of conserved motifs in the Txt A-domains. Typically, an A-domain contains ten highly conserved motifs (A1–A10): motifs A1, A2 and A6 are distal from the active site and are designated with structural roles, motifs A3-A5, A7-A10 are involved in substrate binding and/or catalysis (reviewed by Gulick, 2009; Labby et al., 2015). For both Txt NRPSs, the M-domain is embedded between motifs A8 and A9, which is similar to the domain organization of TioS (Mori et al., 2018b). It is notable that the conserved motifs A9 and A10 are not present in the HIS_6-$TxtA^A$ protein that was tested for adenylation activity (Figure 4.6). It has been proposed that the motif A9 is required to stabilize the protein conformation for the thiolation reaction and/or to properly position the substrate interacting with the T-domain. Mutations in the A9 motif of the tyrocidine synthetase 1 A-domain were shown to have no effect on the adenylation activity but to adversely affect the transfer of activated amino acid to the acceptor substrate (Bučević-Popović et al., 2012). Therefore, the absence of the A9 motif in the purified HIS_6-$TxtA^A$ protein may not explain why no adenylation activity was detected in our assays. Then, the Txt NRPS amino acid sequences were examined for the presence of the ten

residues that line the A-domain substrate binding pocket and determine the substrate specificity of the NRPS (Stachelhaus et al., 1999). We noticed that nine of the ten residues are located between A4 and A6, and the absolutely conserved lysine residue (K916 of TxtA and K940 of TxtB) is located within A10 (Figure 4.6). Structural and biochemical data have implicated the importance of the lysine residue within the A10 motif of A-domains. The crystal structure of the phenylalanine adenylating domain PheA from *Bacillus brevis* revealed that the lysine residue (K517 of PheA) of the A10 motif interacts with the carboxylate group of the substrate as well as the phosphate moiety of AMP (Conti et al., 1997; Stachelhaus et al., 1999). The importance of this residue has been confirmed by a site-directed mutagenesis study of the AT-didomain PA1221 from *Pseudomonas aeruginosa,* where substitution of the lysine residue with a leucine abolished the adenylation activity of the enzyme (Mitchell et al., 2012).

To investigate the importance of the lysine residue in the Txt NRPSs, we conducted *in silico* structural analysis of TxtA using SWISS-MODEL (Biasini et al., 2014). Since TioS (PDB ID: 5wmm_1; Mori et al., 2018b) is currently the only M-domain-interrupted A-domain with the crystal structure solved, it serves as the best template for TxtA structural modeling (Supplementary Table 4.2). The predicted 3-dimentional structure of TxtA shows that K916 is situated away from the predicted substrate binding pocket where the other residues of the signature sequence are located (Figure 4.7A). However, it is notable that the crystal structure of TioS is solved when the aminoacyl-AMP intermediate is bound to the A-domain, indicating that the enzyme is in the thiolation state when the activated amino acid is ready to be loaded onto the Ppant arm of the downstream T-domain (Mori et al., 2018b; Süssmuth and Mainz, 2017). Thus, it is possible that the lysine residue is situated

close to the $TxtA^A$ active site when the enzyme is in the adenylation state, and following substrate adenylation, the NRPS then changes its orientation so that the T-domain can approach the activated substrate for the thiolation reaction (Figure 4.7B). The importance of K916 in TxtA and K940 in TxtB warrants further investigation; however, it is reasonable to propose that the absence of the lysine residue from the HIS_6-$TxtA^A$ protein construct may explain why no adenylation activity was observed in our enzyme assays.

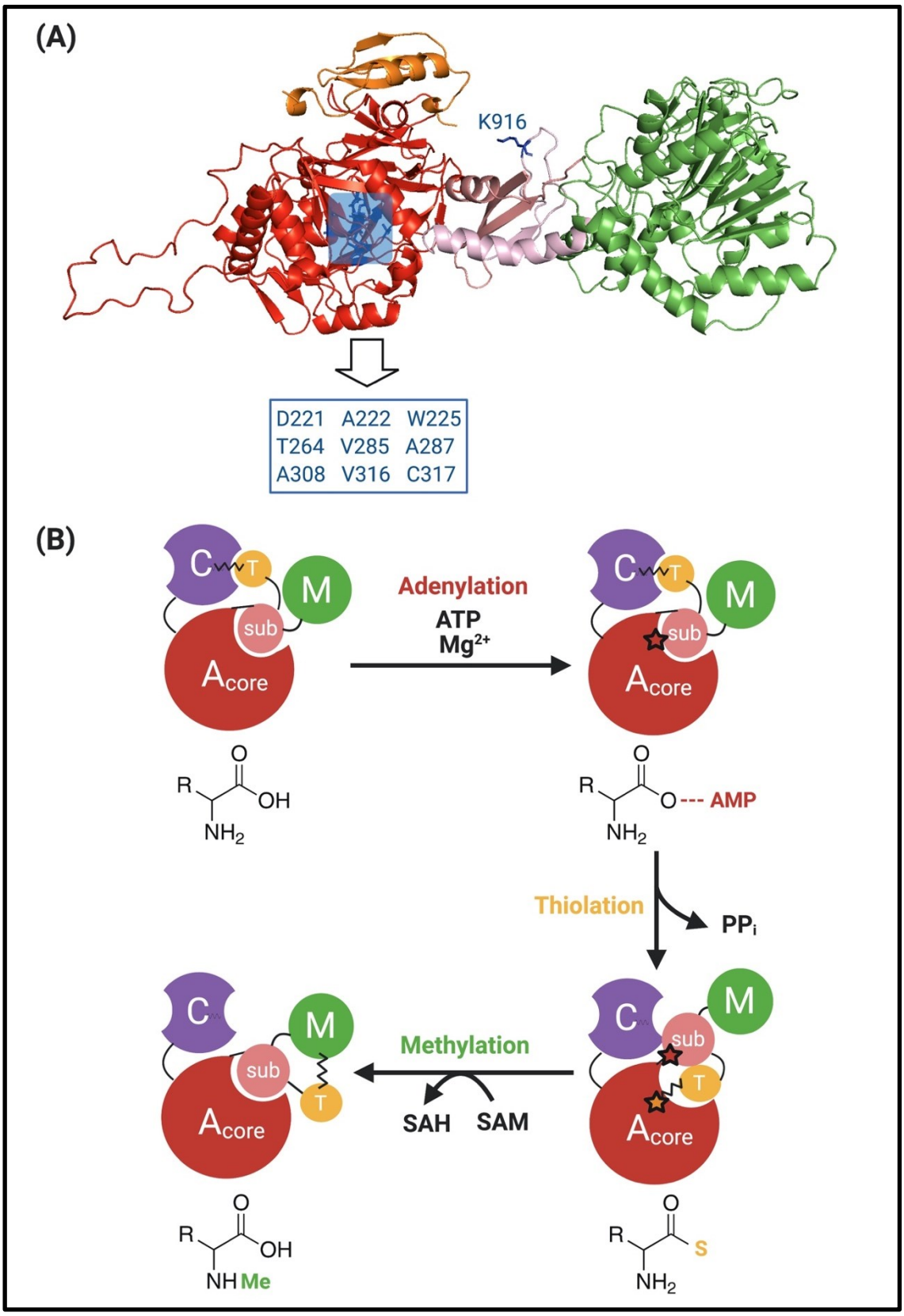

(A)
K916
D221 A222 W225
T264 V285 A287
A308 V316 C317
(B)
C
T
M
sub
Acore
Adenylation
ATP
Mg2+
R
O
OH
NH2
O
AMP
Thiolation
PPi
Methylation
SAH
SAM
S
NH Me

Figure 4.7 (A) Predicted 3-dimentional structure of the *S. scabiei* TxtA A-domain (red: A_{core}-domain, pink: A_{sub}-domain) and M-domain (green) docking with the predicted structure of TxtH (orange). The AM-domain structures were predicted using the crystal structure of TioS from *Micromonospora* sp. ML1 (PDB: 5wmm_1; Mori et al., 2018b) as the template. The two AMP-binding enzyme C-terminal domains (also known as A_{sub}-domain) located at the N-terminus and C-terminus of the M-domain is indicated by dark and light pink color, respectively. The predicted model of TxtH is positioned next to the A-domain based on the location of the TioT MLP that is bound to the A-domain in TioS (PDB: 5wmm; Li et al., 2020). The nine conserved residues of $TxtA^{A}$ and the lysine residues (K916) in the A_{sub}-domain that line the A-domain substrate binding pocket are highlighted as blue sticks. (B) Predicted model of the *S. scabiei* TxtA AMT-domain dynamics. In the adenylation reaction, the A-domain (red and pink) activates the amino acid substrate to form an aminoacyl adenylate intermediate. Then, rotation of the A_{sub}-domain (pink) facilitates the Ppant attachment site (yellow star) to penetrate into the A_{core}-domain (red) substrate pocket for thiolation. Once the aminoacyl-thioester intermediate has formed, it is subject to methylation by the M-domain (green). The red star denotes the locations of the conserved lysine residue in the A_{sub}-domain, which is considered essential for the adenylation activity.

4.4.3 Co-expression of HIS_6-$TxtA^{AMT}$ and HIS_6-$TxtB^{AMT}$ with HIS_6-TxtH

Given our *in silico* analysis of TxtA and the potential importance of the conserved A10 motif for the adenylation activity of the protein, we next sought to express the AMT domains of both TxtA and TxtB in *E. coli.* The T-domain was included for each protein since the co-expression of A- and T-domains from the same module has been done in other studies that performed *in vitro* characterization of amino acid activation by NRPSs (Felnagle et al., 2010; McMahon et al., 2012; Mori et al., 2018a; Zhang et al., 2010). The AMT-domains of both TxtA and TxtB were expressed as N-terminal HIS_6-tagged proteins together with HIS_6-TxtH. As shown in Figure 4.8, both HIS_6-$TxtA^{AMT}$ and HIS_6-$TxtB^{AMT}$ required the presence of HIS_6-TxtH for soluble expression, as expected; however, only truncated forms of both proteins were detected by western blot analysis. A band consistent with the full-length protein for both HIS_6-$TxtA^{AMT}$ and HIS_6-$TxtB^{AMT}$ could be detected

when the cultures were incubated for longer periods of time following induction with IPTG, but a significant amount of the truncated form for each was still present in the extracts (Figure 4.8). When we attempted a small-scale purification of HIS_6-$TxtA^{AMT}$, only the truncated form was recovered (Supplementary Figure 4.3).

It is known that large heterologous proteins (more than 100 kDa) are prone to degradation and premature termination when expressed in *E. coli*. Moreover, large proteins expressed in the cytoplasm can be physically difficult to translocate (Kaur et al., 2018). The size of HIS_6-$TxtA^{AMT}$ and HIS_6-$TxtB^{AMT}$ may account for the protein truncation during expression in *E. coli*. Furthermore, the codon adaption index for expression of $TxtA^{AMT}$ and $TxtB^{AMT}$ in *E. coli* is 0.67 and 0.66, respectively, which suggests that codon bias could cause truncated polypeptide formation during protein synthesis (Rosano and Ceccarelli, 2014). Thus, future work on the expression of $TxtA^{AMT}$ and $TxtB^{AMT}$ could be conducted using *E. coli* strains such as Rosetta (DE3) (Novagen, US), which expresses tRNAs that recognize the rare codons AGG, AGA, AUA, CUA, CCC, and GGA in *E. coli* (Kaur et al., 2018).

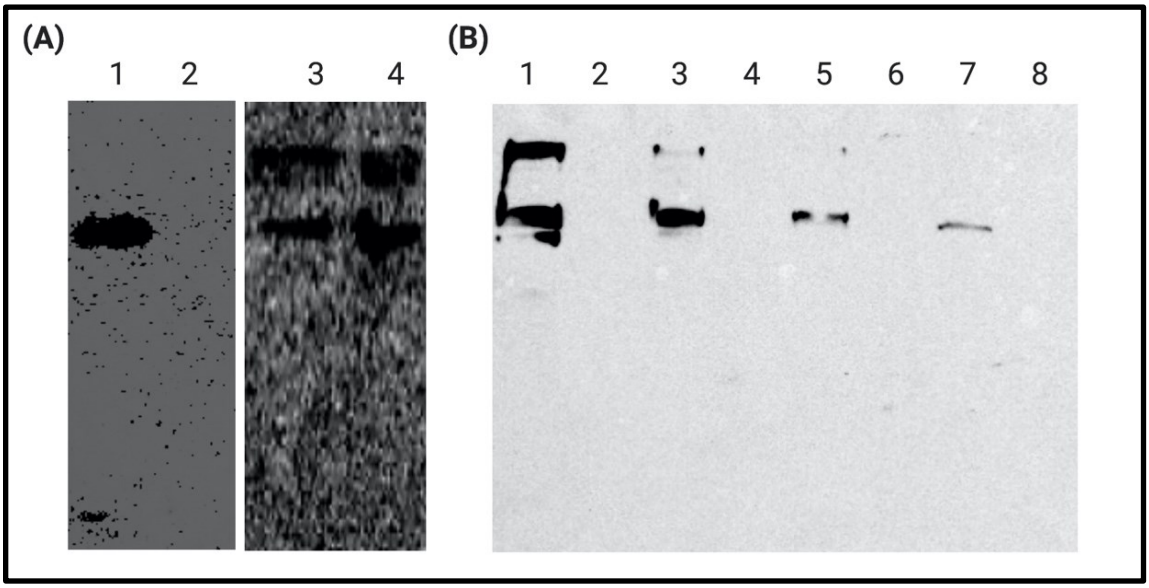


Figure 4.8 (A) Western blot analysis of soluble HIS_6-$TxtA^{AMT}$ expressed in the presence and absence of HIS_6-TxtH (lane 2). The production of the HIS_6-tagged proteins was induced for 20 (lane 1), 24 (lane 3) and 48 hours (lane 4). (B) Western blot analysis of soluble HIS_6-$TxtB^{AMT}$ expressed in the presence and absence of HIS_6-TxtH (lane 2, 4, 6, 8). The production of the HIS_6-tagged proteins was induced for 48 (lanes 1 and 3), 24 (lane 5) and 16 hours (lane 7). Different amount of total soluble proteins was analyzed in lanes 1, 2 (60 µg) and lanes 3-8 (30 µg). The upper band (~112 kDa) corresponds to the expected molecular weight of HIS_6-$TxtA^{AMT}$ and HIS_6-$TxtB^{AMT}$, and the lower band suggests the truncation of the target protein.

It is noteworthy that Jiang and colleagues recently reported the successful purification of the full-length TxtA and TxtB proteins when co-expressed with TxtH (Jiang et al., 2018). In this study, the authors used a C-terminal HIS_6 tag on both the NRPS and MLP rather than an N-terminal HIS_6 tag, and they cultured the *E. coli* expression strain in terrific broth, which is known to produce more recombinant protein than Luria-Bertani broth (Losen et al., 2004; Tripathi et al., 2009). In addition, the proteins were induced using a lower concentration of IPTG (0.1 mM) and the cultures were incubated at 16°C for 16 hours (Zuo et al., 2016). Thus, future studies on the expression of $TxtA^{AMT}$ and $TxtB^{AMT}$, or even the full-length Txt NRPSs, could be conducted following this approach.

4.5 Conclusion

In this chapter, we attempted to investigate the impact of TxtH and non-cognate MLPs on the enzymatic activities of the Txt NRPSs A-domains. To achieve this, we co-expressed HIS_6-$TxtA^A$ with HIS_6-TxtH in *E. coli*, and we showed that the two proteins can form a complex *in vivo* with a 2:2 molar ratio. The A-domain was then purified, and its activity was tested *in vitro* using a molybdate/malachite green phosphate assay; however, no adenylation activity towards the L-phenylalanine substrate could be detected despite testing different reaction conditions. Re-examination of the Txt NRPSs amino acid sequences indicated that a conserved lysine residue that may play a catalytic role in the adenylating reaction is absent in the purified HIS_6-$TxtA^A$ protein used in the enzyme assays. Additional efforts were made to express the AMT-domains of both TxtA and TxtB in *E. coli*; however, the proteins were unstable and were mainly detected as truncated forms. Future work will focus on the expression and purification of the full-length (or the AMT domains) of the Txt NRPSs with TxtH as described by Jiang and colleagues (Jiang et al., 2018). Characterization of the adenylation activity of the Txt NRPSs will benefit from the optimization of the colorimetric assay by introducing positive controls and reducing nonenzymic hydrolysis of ATP. The importance of the lysine residue in A10 motif of the TxtA A-domain should be examined by mutating this residue and testing the adenylation activity of the mutant. Additionally, whether non-cognate MLPs also have the ability to bind to the Txt A-domains and the stoichiometries of such complexes requires further examination. The influence of TxtH and other MLPs on the enzyme kinetics and substrate

specificity of the Txt NRPSs in *S. scabiei* still remains obscure but is of great interest for future studies.

4.6 Acknowledgement

This work was supported by a Natural Sciences Engineering Research Council of Canada Discovery Grant (RGPIN-2018-05155) to D.R.D.B. and by a Natural Sciences Engineering 547 Research Council of Canada Discovery Grant (RGPIN-2018–05949) to K.T. Y.L. was supported in part by the President's Doctoral Student Investment Fund at Memorial University of Newfoundland.

4.7 References

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., et al. (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. 42, W252-W258. doi:10.1093/nar/gku340.

Bignell, D.R.D., Huguet-Tapia, J.C., Joshi, M.V., Pettis, G.S., and Loria, R. (2010) What does it take to be a plant pathogen: Genomic insights from *Streptomyces* species. *Antonie van Leeuwenhoek*. 98, 179–194. doi:10.1007/s10482-010-9429-1.

Bischoff, V., Cookson, S.J., Wu, S., and Scheible, W.R. (2009) Thaxtomin A affects CESA-complex density, expression of cell wall genes, cell wall composition, and causes ectopic lignification in *Arabidopsis thaliana* seedlings. *J Exp Bot.* 60, 955–965. doi:10.1093/jxb/ern344.

Bisswanger, H. (2014) Enzyme assays. *Perspect Sci.* 1, 41–55. doi:10.1016/j.pisc.2014.02.005.

Biswas, T., Resto-Roldá N, E., Sawyer, S.K., Artsimovitch, I., and Tsodikov, O.V. (2013) A novel non-radioactive primase-pyrophosphatase activity assay and its application to the discovery of inhibitors of *Mycobacterium tuberculosis* primase DnaG. *Nucleic Acids Res.* 41, e56-e56. doi:10.1093/nar/gks1292.

Boll, B., Taubitz, T., and Heide, L. (2011) Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. *J Biol Chem*. 286, 36281–36290. doi:10.1074/jbc.M111.288092.

Bown, L. (2018) Coronafacoyl phytotoxin biosynthesis in *Streptomyces scabies* and identification of common scab-associated *Streptomyces* species in Newfoundland (Doctoral dissertation, Memorial University of Newfoundland).

Brown, R.B., and Audet, J. (2008) Current techniques for single-cell lysis. *J R Soc Interface.* 5, S131-S138. doi: 10.1098/rsif.2008.0009.focus.

Bučević-Popović, V., Šprung, M., Soldo, B., and Pavela-Vrančič, M. (2012) The A9 core sequence from NRPS adenylation domain is relevant for thioester formation. *ChemBioChem*. 13, 1913-1920. doi: 10.1002/cbic.201200309.

Cogan, E.B., Birrell, G.B., and Griffith, O.H. (1999) A robotics-based automated assay for inorganic and organic phosphates. *Anal Biochem*. 271, 29–35. doi:10.1006/abio.1999.4100.

Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* 16, 4174-4183. doi:10.1093/emboj/16.14.4174.

Davidsen, J.M., Bartley, D.M., and Townsend, C.A. (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. doi:10.1021/ja307710d.

DeLano, W.L. (2002) Pymol: An open-source molecular graphics tool. *Protein Crystallogr.* 40, 82–92.

Duval, I., and Beaudoin, N. (2009) Transcriptional profiling in response to inhibition of cellulose synthesis by thaxtomin A and isoxaben in *Arabidopsis thaliana* suspension cells. *Plant Cell Rep.* 28, 811–830. doi:10.1007/s00299-009-0670-x.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., et al. (2019) The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi:10.1093/nar/gky995.

Felnagle, E.A., Barkei, J.J., Park, H., Podevels, A.M., McMahon, M.D., Drott, D.W., et al. (2010) MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry.* 49, 8815–8817. doi:10.1021/bi1012854.

Finking, R., and Marahiel, M.A. (2004) Biosynthesis of nonribosomal peptides. *Annu Rev Microbiol.* 58, 453–488. doi: 10.1146/annurev.micro.58.030603.123615.

Fyans, J.K., Bown, L., and Bignell, D.R.D. (2015) Isolation and Characterization of Plant-Pathogenic *Streptomyces* Species Associated with Common Scab-Infected Potato Tubers in Newfoundland. *Phytopathology.* 106, 123–131. doi:10.1094/phyto-05-15-0125-r.

Goyer, C., Vachon, J., and Beaulieu, C. (1998) Pathogenicity of *Streptomyces scabies* mutants altered in thaxtomin a production. *Phytopathology*. 88, 442-445. doi:10.1094/PHYTO.1998.88.5.442.

Grabski, A., McCormick, M., and Mierendorf, R. (1999) BugBuster™ and Benzonase®: the clear solutions to simple, efficient extraction of *E. coli* proteins. *Innovations.* 10, 17-19.

Gulick, A.M. (2009) Conformational dynamics in the acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem Biol*. 4, 811-827. doi:10.1021/cb900156h.

Healy, F.G., Wach, M., Krasnoff, S.B., Gibson, D.M., and Loria, R. (2000) The *txtAB* genes of the plant pathogen *Streptomyces acidiscabies* encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity. *Mol Microbiol*. 38, 794–804. doi:10.1046/j.1365-2958.2000.02170.x.

Heemstra, J.R., Walsh, C.T., and Sattely, E.S. (2009) Enzymatic tailoring of ornithine in the biosynthesis of the rhizobium cyclic trihydroxamate siderophore vicibactin. *J Am Chem Soc.* 131, 15317–15329. doi:10.1021/ja9056008.

Herbst, D.A., Boll, B., Zocher, G., Stehle, T., and 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. doi:10.1074/jbc.M112.420182.

Herzberg, C., Weidinger, L.A.F., Dörrbecker, B., Hübner, S., Stülke, J., and Commichau, F.M. (2007) SPINE: A method for the rapid detection and analysis of protein-

protein interactions *in vivo*. *Proteomics.* 7, 4032-4035. doi:10.1002/pmic.200700491.

Imker, H.J., Krahn, D., Clerc, J., Kaiser, M., and Walsh, C.T. (2010) *N*-Acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF. *Biol Chem.* 17, 1077-1083. doi:10.1016/j.chembiol.2010.08.007.

Jiang, G., Zuo, R., Zhang, Y., Powell, M.M., Zhang, P., Hylton, S.M., et al. (2018) One-Pot Biocombinatorial Synthesis of Herbicidal Thaxtomins. *ACS Catal.* 8, 10761–10768. doi:10.1021/acscatal.8b03317.

Johnson, E.G., Krasnoff, S.B., Bignell, D.R.D., Chung, W.C., Tao, T., Parry, R.J., et al. (2009) 4-Nitrotryptophan is a substrate for the non-ribosomal peptide synthetase TxtB in the thaxtomin A biosynthetic pathway. *Mol Microbiol.* 73, 409–418. doi:10.1111/j.1365-2958.2009.06780.x.

Kaur, J., Kumar, A., and Kaur, J. (2018) Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. *J Biol Macromol*. 106, 803-822. doi:10.1016/j.ijbiomac.2017.08.080.

Kieser, T., Bibb, M., Buttner, M., Chater, K. and Hopwood, D. (2000) Practical *Streptomyces* genetics. Norwich, UK: The John Innes Foundation.

King, R.R., and Calhoun, L.A. (2009) The thaxtomin phytotoxins: Sources, synthesis, biosynthesis, biotransformation and biological activity. *Phytochemistry*. 70, 833–841. doi: 10.1016/j.phytochem.2009.04.013.

King, R.R., Lawrence, C.H., and Clark, M.C. (1991) Correlation of phytotoxin production with pathogenicity of *Streptomyces scabies* isolates from scab infected potato tubers. *Am J Potato Res.* 68, 675-680. doi:10.1007/BF02853743.

Kinkel, L.L., Bowers, J.H., Shimizu, K., Neeno-Eckwall, E.C., and Schottel, J.L. (1998) Quantitative relationships among thaxtomin A production, potato scab severity, and fatty acid composition in *Streptomyces*. *Can J Microbiol*. 44, 768-776. doi:10.1139/w98-061.

Kittilä, T., Schoppet, M., and Cryle, M.J. (2016) Online Pyrophosphate Assay for Analyzing Adenylation Domains of Nonribosomal Peptide Synthetases. *ChemBioChem.* 17, 576–584. doi:10.1002/cbic.201500555.

Labby, K.J., Watsula, S.G., and Garneau-Tsodikova, S. (2015) Interrupted adenylation domains: Unique bifunctional enzymes involved in nonribosomal peptide biosynthesis. *Nat Prod Rep*. 32, 641-653. doi:10.1039/c4np00120f.

Lanzetta, P.A., Alvarez, L.J., Reinach, P.S., and Candia, O.A. (1979) An improved assay for nanomole amounts of inorganic phosphate. *Anal Biochem.* 100, 95-97. doi:10.1016/0003-2697(79)90115-5.

Li, Y., Liu, J., Adekunle, D., Bown, L., Tahlan, K., and Bignell, D.R.D. (2019) TxtH is a key component of the thaxtomin biosynthetic machinery in the potato common scab pathogen *Streptomyces scabies*. *Mol Plant Pathol.* 20, 1379–1393. doi:10.1111/mpp.12843.

Li, Y., Tahlan, K., and Bignell, D.R. (2020). Functional Cross-Talk of MbtH-Like Proteins During Thaxtomin Biosynthesis in the Potato Common Scab Pathogen *Streptomyces scabiei*. *Front Microbiol*. 11, 585456. doi: 10.3389/fmicb.2020.585456.

Loria, R., Bukhalid, R.A., Creath, R.A., Leiner, R.H., Olivier, M., and Steffens, J.C. (1995) Differential production of thaxtomins by pathogenic *Streptomyces* species *in vitro*. *Phytopathology.* 85, 537–541. doi:10.1094/Phyto-85-537.

Losen, M., Frölich, B., Pohl, M., and Büchs, J. (2004) Effect of oxygen limitation and medium composition on *Escherichia coli* fermentation in shake-flask cultures. *Biotechnol Prog.* 20, 1062-1068. doi:10.1021/bp034282t.

Lundy, T.A., Mori, S., and Garneau-Tsodikova, S. (2020) A thorough analysis and categorization of bacterial interrupted adenylation domains, including previously unidentified families. *RSC chem biol.* doi:10.1039/d0cb00092b.

McMahon, M.D., Rush, J.S., and Thomas, M.G. (2012) Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *J Bacteriol.* 194, 2809–2818. doi:10.1128/JB.00088-12.

McQuade, T.J., Shallop, A.D., Sheoran, A., DelProposto, J.E., Tsodikov, O.V., and Garneau-Tsodikova, S. (2009) A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. *Anal Biochem*. 386, 244–250. doi:10.1016/j.ab.2008.12.014.

Miller, B.R., Drake, E.J., Shi, C., Aldrich, C.C., and Gulick, A.M. (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. doi:10.1074/jbc.M116.746297.

Mitchell, C.A., Shi, C., Aldrich, C.C., and Gulick, A.M. (2012) Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry.* 51, 3252–3263. doi:10.1021/bi300112e.

Mok, M.T.S., and Edwards, M.R. (2005) Critical sources of error in colorimetric assay for UDP-*N*-acetylglucosamine pyrophosphorylase. *Anal Biochem*. 343, 341–343. doi:10.1016/j.ab.2005.05.023.

Mori, S., Green, K.D., Choi, R., Buchko, G.W., Fried, M.G., and Garneau-Tsodikova, S. (2018a) Using MbtH-Like Proteins to Alter the Substrate Profile of a Nonribosomal Peptide Adenylation Enzyme. *ChemBioChem.* 19, 2186–2194. doi:10.1002/cbic.201800240.

Mori, S., Pang, A.H., Lundy, T.A., Garzan, A., Tsodikov, O.V., and Garneau-Tsodikova, S. (2018b) Structural basis for backbone *N*-methylation by an interrupted adenylation domain. *Nat Chem Biol*. 14, 428–430. doi:10.1038/s41589-018-0014-7.

Rosano, G.L., and Ceccarelli, E.A. (2014) Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front Microbiol.* 5, 1–17. doi:10.3389/fmicb.2014.00172.

Sambrook, J., and Russel, D.W. (2001) Molecular cloning: a laboratory manual (3-volume set). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Scheible, W.R., Fry, B., Kochevenko, A., Schindelasch, D., Zimmerli, L., Somerville, S., Loria, R. and Somerville, C.R. (2003) An Arabidopsis mutant resistant to thaxtomin A, a cellulose synthesis inhibitor from *Streptomyces* species. *Plant Cell*. 15, 1781-1794. doi: 10.1105/tpc.013342.

Schomer, R.A., Park, H., Barkei, J.J., and Thomas, M.G. (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. doi:10.1021/acs.biochem.8b00552.

Schomer, R.A., and Thomas, M.G. (2017) Characterization of the functional variance in MbtH-like protein interactions with a nonribosomal peptide synthetase. *Biochemistry*. 56, 5380–5390. doi: 10.1021/acs.biochem.7b00517.

Sieber, S.A., and Marahiel, M.A. (2005) Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics. *Chem Rev.* 105, 715–738. doi:10.1021/cr0301191.

Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol.* 6, 493–505. doi:10.1016/S1074-5521(99)80082-9.

Süssmuth, R.D., and Mainz, A. (2017) Nonribosomal Peptide Synthesis—Principles and Prospects. *Angew Chem.* 56, 3770-3821. doi:10.1002/anie.201609079.

Tarry, M.J., Haque, A.S., Bui, K.H., and Schmeing, T.M. (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. doi: 10.1016/j.str.2017.03.014.

Tripathi, N.K., Shrivastva, A., Biswal, K.C., and Rao, P.V.L. (2009) Optimization of culture medium for production of recombinant dengue protein in *Escherichia coli. Biotechnol*. 5, 179-183. doi:10.1089/ind.2009.3.179.

Vasilescu, J., Guo, X., and Kast, J. (2004) Identification of protein-protein interactions using *in vivo* cross-linking and mass spectrometry. *Proteomics.* 4, 3845–3854. doi:10.1002/pmic.200400856.

Xia, S., Ma, Y., Zhang, W., Yang, Y., Wu, S., Zhu, M., et al. (2012) Identification of sare0718 as an Alanine-Activating adenylation domain in marine Actinomycete *Salinispora arenicola* CNS-205. *PLoS ONE*. 7, 1–7. doi: 10.1371/journal.pone.0037487.

Zhang, W., Heemstra, J.R., Walsh, C.T., and Imker, H.J. (2010) Activation of the pacidamycin pacl adenylation domain by MbtH-like proteins. *Biochemistry.* 49, 9946–9947. doi:10.1021/bi101539b.

Zolova, O.E., and Garneau-Tsodikova, S. (2012) Importance of the MbtH-like protein TioT for production and activation of the thiocoraline adenylation domain of TioK. *MedChemComm.* 3, 950–955. doi:10.1039/c2md20131c.

Zolova, O.E., and Garneau-Tsodikova, S. (2014) KtzJ-dependent serine activation and *O*-methylation by KtzH for kutznerides biosynthesis. *J Antibiot.* 67, 59-64. doi:10.1038/ja.2013.98.

Zuo, R., Zhang, Y., Huguet-Tapia, J.C., Mehta, M., Dedic, E., Bruner, S.D., et al. (2016) An artificial self-sufficient cytochrome P450 directly nitrates fluorinated tryptophan analogs with a different regio-selectivity. *Biotechnol.* 11, 624-632. doi:10.1002/biot.201500416.

4.8 Supplementary Information

Supplementary Table 4.1 Oligonucleotide primers used in this study.

Primer	Sequence (5′ - 3′) †	Use
PL37	GCGC**GAATTC**GATGTCGCACCTG ACCGGTGAA	Forward primer for construction of pACYCDuet-1/HIS$_6$-*txtA*
PL38	GCGC**AAGCTT**CCAGTAGCTTTCG CAGTCAC	Reverse primer for construction of pACYCDuet-1/HIS$_6$-*txtA*
PL35	GCGC**CATATG**CCCTCACCCTTCG ACGAC	Forward primer for construction of pET28b/HIS$_6$-*txtH*
PL36	GCGC**GAATTC**TCATTCACGGACG GACGCCG	Reverse primer for construction of pET28b/HIS$_6$-*txtH*
PL150	GCGC**CCATGG**GCGTGCCCTCACC CTTCGAC	Forward primer for construction of pET28b/*txtH*
PL36	GCGC**GAATTC**TCATTCACGGACG GACGCCG	Reverse primer for construction of pET28b/*txtH*
PL37	GCGC**GAATTC**GATGTCGCACCTG ACCGGTGAA	Forward primer for construction of pACYCDuet-1/HIS$_6$-*txtA*^{AMT}
PL39	GCGC**AAGCTT**GGAGGCGTACTCG CTCGGCA	Reverse primer for construction of pACYCDuet-1/HIS$_6$-*txtA*^{AMT}
PL40	GCGC**GAATTC**GATGTCCATGCTG CCGCCGGG	Forward primer for construction of pACYCDuet-1/HIS$_6$-*txtB*^{AMT}
PL42	GCGC**AAGCTT**GTCGCGGCAGAGC AACAGC	Reverse primer for construction of pACYCDuet-1/HIS$_6$-*txtB*^{AMT}
PL46	GTCGTCGTCACCCACCAAG	Primer for verify the sequence of pACYCDuet-1/HIS$_6$-*txtA*^{AMT}
PL47	TTCGTGGCGGACCCCTTC	Primer for verify the sequence of pACYCDuet-1/HIS$_6$-*txtA*^{AMT}
PL48	GTCGTCCTGCATCAGCTCTC	Primer for verify the sequence of pACYCDuet-1/HIS$_6$-*txtA*^{AMT}
PL49	TCTACACCTCGGGTTCCACC	Primer for verify the sequence of pACYCDuet-1/HIS$_6$-*txtB*^{AMT}
PL50	CCATTCCCGTGCCTCAGATG	Primer for verify the sequence of pACYCDuet-1/HIS$_6$-*txtB*^{AMT}
PL59	GATCACCTCGGCGGAGAAG	Primer for verify the sequence of pACYCDuet-1/HIS$_6$-*txtB*^{AMT}

† Non-homologous extensions are underlined, while engineered restriction sites are indicated in bold.

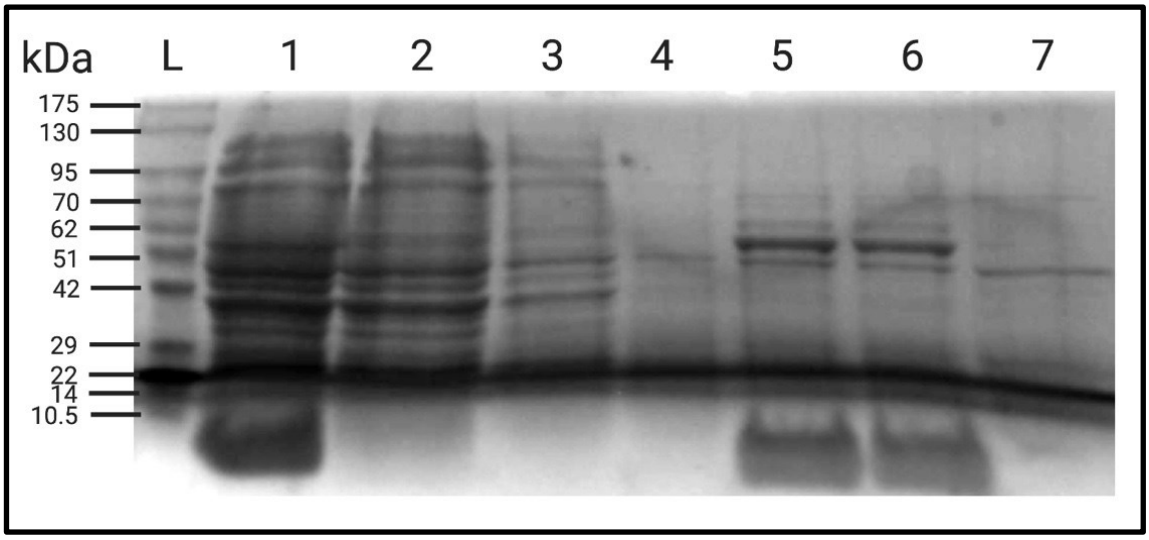
Supplementary Table 4.2 Quality parameters of structural models built for $TxtA^{AM}$, $TxtB^{AM}$ using the SWISS-MODEL.

Protein	**$TxtA^{AM}$**	**$TxtB^{AM}$**
Template PDB	5wmm.1.A	5wmm.1.A
Description	TioS NRPS from *Micromonospora* sp. ML1	TioS NRPS from *Micromonospora* sp. ML1
Reference	Mori et al., 2018b	Mori et al., 2018b
Method	X-ray, 2.9Å	X-ray, 2.9Å
Identity	40.19%	43.33%
GMQE*	0.45	0.44
QMEAN†	-2.58	-2.43
Cβ	-2.68	-2.35
All Atom	-1.61	-0.96
Solvation	-0.76	-1.30
Torsion	-1.81	-1.57

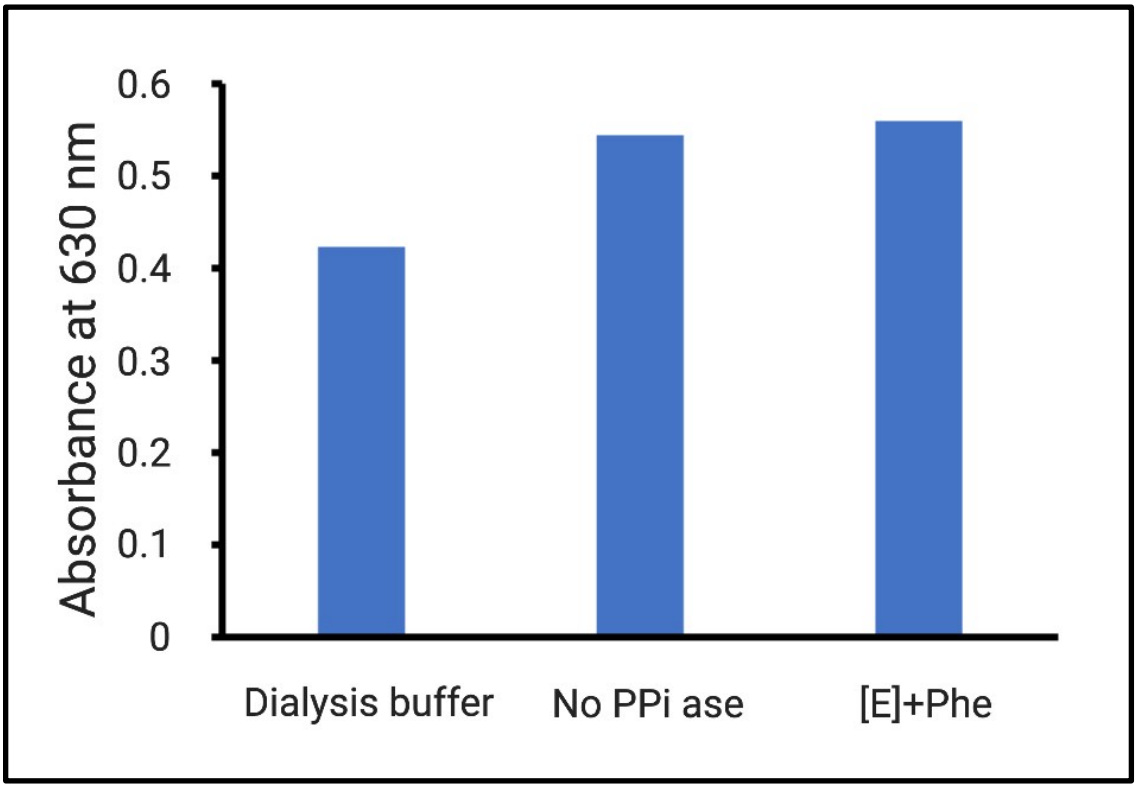
* GMQE: Global Model Quality Estimation
† QMEAN: Qualitative Model Energy ANalysis

Supplementary Table 4.3 Domain prediction of TxtA and TxtB from *S. scabiei* 87.22.

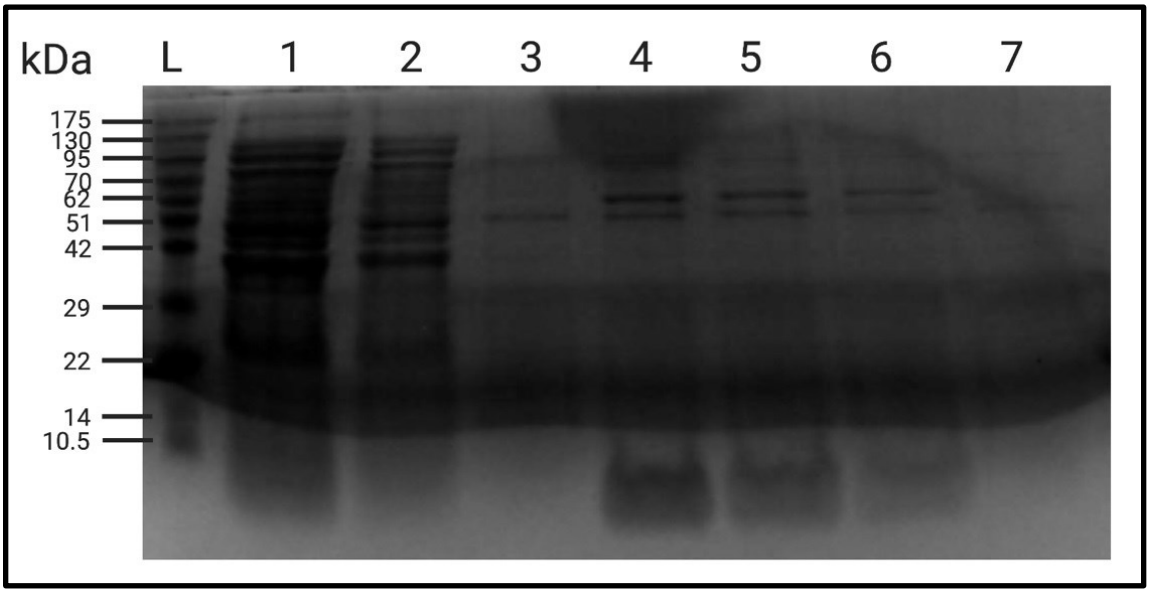
TxtA domains with significant Pfam-A Matches						
Family	Description	Clan	Start	End	Bit score	E-value
AMP-binding	AMP-binding enzyme	CL0378	23	421	319.2	3.3e-95
Condensation	Condensation domain	CL0149	1026	1455	194.5	2.7e-57
PP-binding	Phosphopantetheine attachment site	CL0314	943	1005	53.1	3.1e-14
Methyltransf_11	Methyltransferase domain	CL0063	538	639	51.2	1.5e-13
TxtA domains with insignificant Pfam-A Matches						
AMP-binding_C	AMP-binding enzyme C-terminal domain	CL0531	429	466	19.9	0.0011
AMP-binding_C	AMP-binding enzyme C-terminal domain	CL0531	874	916	14.0	0.074
TxtB domains with significant Pfam-A Matches						
AMP-binding	AMP-binding enzyme	CL0378	33	431	309.0	3.9e-92
Condensation	Condensation domain	CL0149	1032	1482	288.7	7.1e-86
Methyltransf_11	Methyltransferase domain	CL0063	549	651	58.1	1.1e-15
PP-binding	Phosphopantetheine attachment site	CL0314	951	1011	51.2	1.2e-13
AMP-binding_C	AMP-binding enzyme C-terminal domain	CL0531	439	492	34.0	4.2e-08
TxtB domains with insignificant Pfam-A Matches						
AMP-binding_C	AMP-binding enzyme C-terminal domain	CL0531	889	923	13.9	0.08



Supplementary Figure 4.1 Small-scale purification of HIS_6-$TxtA^A$ and HIS_6-TxtH. Protein samples collected at different stages during the purification process were analyzed by SDS-PAGE on a 15% w/v polyacrylamide gel and were visualized using Coomassie Blue stain. Lane 1: total soluble proteins following cell lysis; lane 2: column flow through; lane 3: column wash with binding buffer; lane 4: column wash with wash buffer; lanes 5 and 6: two different fractions of eluted proteins; lane 7: eluted proteins of the expression strain [BL21(DE3)*ybdZ:aac(3)IV*] lacking the expression plasmids. The bands (in lanes 5 and 6) at ~62 kDa and at <10.5 kDa correspond to the expected molecular weight for HIS_6-$TxtA^A$ (62 kDa) and HIS_6-TxtH (9.5 kDa), respectively. Lane L: PiNK Plus Prestained Protein Ladder (FroggaBio Inc.).



Supplementary Figure 4.2 Quantification of adenylation activity of HIS_6-$TxtA^A$ co-expressed with TxtH. Reactions were prepared using 0.5 µM of purified protein ([E]) and 6 mM of L-phenylalanine (Phe) as substrate. The reactions were initiated by addition of 0.25 mM ATP. Reaction containing dialysis buffer in place of protein and pyrophosphatase (PP_iase) were used as negative controls. Each reaction was performed once, and the columns represent the absorbance reading at 630 nm.



Supplementary Figure 4.3 Small-scale purification of HIS_6-$TxtA^{AMT}$ and HIS_6-TxtH. Protein samples collected at different stages during the purification process were analyzed by SDS-PAGE on a 15% w/v polyacrylamide gel and were visualized using Coomassie Blue stain. Lane 1: column flow through; lane 2: column wash with binding buffer; lane 3: column wash with wash buffer; lanes 4-6: different fractions of eluted proteins; lane 7: eluted proteins of the expression strain [BL21(DE3)*ybdZ:aac(3)IV*] lacking the expression plasmids. The main bands (in lanes 4-6) lower than the expected molecular weight of HIS_6-$TxtA^{AMT}$ (~112 kDa) suggest only the truncated form of the target protein was recovered. Lane L: PiNK Plus Prestained Protein Ladder (FroggaBio Inc.).

CHAPTER 5

Summary and Future Directions

5.1 Summary of Results

Thaxtomin A is a phytotoxic NRP that functions as the main pathogenicity determinant of *Streptomyces scabiei* and other CS-causing *Streptomyces* spp. The biosynthesis of thaxtomin A involves two NRPSs, TxtA and TxtB, both of which contain an A-domain that recognizes and activates the appropriate amino acid substrate, which is then incorporated into the thaxtomin backbone. In addition to the Txt NRPS machinery, a small gene, *txtH,* is located at the downstream of the *txtB* gene and encodes a protein belonging to the MLP family (Bignell et al., 2010). Prior to starting this work, the role of TxtH during the biosynthesis of thaxtomin A in *S. scabiei* and other thaxtomin-producing species was not known. It has been reported that MLPs are often encoded within BGCs that produce NRP molecules in bacteria (Baltz, 2011), and several biochemical studies demonstrated the importance of MLPs in promoting the solubility and/or activity of NRPS A-domains (Boll et al., 2011; Davidsen et al., 2013; Felnagle et al., 2010; Heemstra et al., 2009; Imker et al., 2010; McMahon et al., 2012; Zhang et al., 2010; Zolova and Garneau-Tsodikova, 2012). However, not all NRPS A-domains require an MLP for full functionality, even if the associated BGC encodes an MLP. Within a NRP biosynthetic pathway involving more than one NRPS, the same MLP can have varying effects on different NRPSs (Felnagle et al., 2010; McMahon et al., 2012). It has also been demonstrated that in organisms harbouring multiple MLP-encoding genes in different BGCs, the MLPs can sometimes

exhibit functional cross-talk with each other, thus enabling the production of a particular NRP in the absence of the corresponding MLP (Lautru et al., 2007; Wolpert et al., 2007). Interestingly, the ability to exhibit functional redundancy varies among different MLPs, and the mechanism behind this was not well understood.

The overall goal of this thesis was to decipher the role of MLPs in the thaxtomin biosynthetic pathway in *S. scabiei*. In Chapter 2, we used a combination of genetic and biochemical approaches to study the function of TxtH during the biosynthesis of thaxtomin A. The results demonstrated that TxtH is required for promoting the soluble expression of both $TxtA^A$ and $TxtB^A$ in *E. coli*, suggesting that TxtH serves as a chaperone to assist the proper folding of the Txt NRPSs in *S. scabiei*. We also identified amino acid residues within TxtH that are essential for the solubility-promoting activity of the protein. Gene deletion studies in *S. scabiei* showed that TxtH is required for the production of thaxtomin A, and that two non-cognate MLPs encoded elsewhere on the *S. scabiei* chromosome can exhibit functional cross-talk with TxtH. In contrast, two non-cognate MLPs from other *Streptomyces* spp. were unable to functionally replace TxtH in the constructed MLP mutant. Overall, the work presented here is the first to examine the impact of MLPs on thaxtomin production and plant pathogenicity in *S. scabiei,* and the results were published in the journal Molecular Plant Pathology in 2019.

In an attempt to investigate why some MLPs can exhibit functional cross-talk with TxtH while others cannot, I conducted a broad survey of MLPs from diverse phylogenetic lineages to explore the ability of these MLPs to replace TxtH in biochemical and genetic assays. The results, which are presented in Chapter 3, showed that several MLPs from distinct phylogenetical clades can promote the soluble expression of the Txt NRPS A-

domains to varying degrees in *E. coli*, though two of the MLPs tested (YbdZ and CGL27_RS10110) were unable to do so. I showed that the impact of a given MLP on protein solubility differed between the two Txt A-domains in some instances, with $TxtB^A$ solubility being more strongly affected by the MLP partner with which it was co-expressed. I also demonstrated that the non-cognate MLPs varied in their ability to restore thaxtomin production in the *S. scabiei* MLP mutant, and the same two MLPs that failed to promote soluble A-domain protein production in *E. coli* (YbdZ and CGL27_RS10110) were unable to support production of the phytotoxin. Notably, I found no relationship between the ability of an MLP to serve as a functional partner for the thaxtomin NRPSs and its overall amino acid similarity with TxtH. Rather, *in silico* structural analysis revealed that the ability of an MLP to exhibit functional redundancy with TxtH likely depends on the conservation of important residues within the MLP that lie at the interacting interface with the Txt NRPSs. Overall, this chapter provides additional insights into the mechanism of the MLP cross-talk and its impact on the biosynthesis of NRPs. The results presented here were recently published in the journal Frontiers in Microbiology.

Given that some MLPs can influence the enzyme activity of NRPS A-domains, I aimed to assess whether TxtH and other MLPs have an impact on the enzymology of the Txt NRPS A-domains in Chapter 4 of the thesis. First, I demonstrated that TxtH can form a complex with $TxtA^A$ *in vivo* in a 2:2 molar ratio, and this agrees with the stoichiometry reported for other MLP-NRPS complexes (Boll et al., 2011; Mori et al., 2018). Subsequently, I purified $TxtA^A$ (co-expressed with TxtH) from *E. coli* using affinity chromatography, and I tested the ability of the A-domain to adenylate its amino acid substrate, L-phenylalanine, using an *in vitro* molybdate/malachite green phosphate assay.

However, I was unable to detect any adenylation activity for the purified A-domain despite testing different assay conditions. Bioinformatics analysis of the TxtA amino acid sequences revealed that the $TxtA^A$ protein used in the enzyme assays is missing a key amino acid residue that may be required for the catalytic activity of the domain, and this may explain why I was unable to detect any activity in the assays performed. I also attempted to express the AMT-domains of both TxtA and TxtB in *E. coli* in the presence of TxtH, but the proteins were unstable and were mainly detected as truncated forms. Thus, the impact of TxtH and other MLPs on the catalysis of the Txt NRPSs in *S. scabiei* remains to be determined.

5.2 Future Directions

Collectively, my research has revealed new insights into the impact of TxtH and other MLPs on the biosynthesis of thaxtomin A in *S. scabiei*. Site-directed mutagenesis and *in silico* analyses identified amino acid residues that contribute (or are predicted to contribute) to the chaperone-like activity of TxtH, and future efforts should focus on obtaining structural data for the TxtH-TxtA/B complexes. The results suggest that TxtH and other MLPs interact differently with TxtB compared to TxtA, and the structural analysis would provide insights into the stoichiometry of the TxtH-TxtB protein complex and whether it differs from that of the TxtH-TxtA complex. Also, key residues identified within the MLP interaction interface of TxtB that differ from residues in the TxtA interface could be mutated, and this would provide further insights into the residues that are responsible for the observed differences in the interaction of the two Txt NRPSs with their

cognate MLP. As YbdZ and CGL27_RS10110 were the only non-cognate MLPs that were unable to promote thaxtomin production in the absence of TxtH in *S. scabiei,* it would be interesting to mutate the key residues identified in these MLPs that differ from TxtH to determine whether the mutant proteins can acquire the ability to promote Txt NRPS protein solubility and thaxtomin production. Such work would further enhance our understanding of the residues that are critical for promoting the interaction of MLPs with the Txt NRPSs.

Further work is also required to elucidate what influence, if any, TxtH has on the enzymology of the Txt NRPSs. Optimization of the molybdate/malachite green assay for detecting adenylation activity could be performed using the *E. coli* EntF NRPS as a positive control. Traditionally, the radiolabeled ATP/PP_i exchange assay has been used for characterizing the adenylation activity of NRPSs (Otten et al., 2007), and so this could also be used for studying the Txt NRPSs. Expression of the complete TxtA and TxtB proteins in *E. coli* as described previously (Jiang et al., 2018) would enable us to examine the influence of TxtH on the enzymatic activity of the entire NRPS machinery and whether it associates with other enzymatic domains in addition to the A-domains. One potential complication that may be encountered is the inability to express and purify the NRPSs in the absence of TxtH. In this scenario, site-directed mutagenesis of TxtH could be performed to determine if mutant proteins can be isolated that maintain the solubility-promoting activity but are deficient in stimulating the enzymatic activity of one or both Txt NRPSs. Similarly, I could test the enzymatic activity of the NRPSs when co-expressed with the different non-cognate MLPs that were as efficient as TxtH in promoting soluble A-domain production in *E. coli,* but were less efficient in promoting thaxtomin production in *S. scabiei* (Li et al., 2020).

Interestingly, Mori and colleagues demonstrated that non-cognate MLPs can change the adenylation activity of the NRPS TioK to activate different amino acid substrates (Mori et al., 2018). This finding reveals a practical application of MLPs for increasing the diversity of NRP products produced by bacteria. On-going studies in our lab are addressing whether the deletion of endogenous MLPs and the overexpression of non-cognate MLPs can affect the metabolomic profile of *S. scabiei* 87.22. Additionally, the ability of TxtH to functionally replace the other endogenous MLPs of *S. scabiei* (MLP_{lipo} and MLP_{scab}) and promote the production of the associated NRPs (lipopeptide, scabichelin) is also under investigation.

Overall, the results of this thesis further enhance our understanding of the thaxtomin biosynthetic machinery in *S. scabiei.* Thaxtomin is essential for CS disease development by several different plant pathogenic *Streptomyces* spp., and TxtH is conserved in the all Txt BGCs identified to date. Thus, our research on the biosynthesis of thaxtomin and on the function of TxtH is expected to have useful applications for the development of better control strategies for managing CS disease. In addition, the potent phytotoxic activity of thaxtomin A makes it an attractive alternative to traditional agrichemicals for the control of weed growth (Koivunen et al., 2013; Leep et al., 2010). As such, a better understanding of the thaxtomin biosynthetic pathway may facilitate the large-scale commercial production of this compound for agricultural applications. An important question that remains unanswered is what makes an NRPS MLP-independent. By studying the interaction of MLPs with NRPSs, it may be possible to engineer NRPS enzymes that do not require an MLP for functionality. This could have beneficial applications not only for the commercial

production of thaxtomin A, but also for the production of other bacterial NRPs with useful bioactivities.

5.3 References

Baltz, R.H. (2011) Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J Ind Microbiol Biotechnol*. 38, 1747-1760. doi:10.1007/s10295-011-1022-8.

Bignell, D.R.D., Huguet-Tapia, J.C., Joshi, M.V., Pettis, G.S. and Loria, R. (2010) What does it take to be a plant pathogen: genomic insights from *Streptomyces* species. *Antonie Van Leeuwenhoek*. 98, 179-194. doi:10.1007/s10482-010-9429-1.

Boll, B., Taubitz, T. and Heide, L. (2011) The role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. *J Biol Chem*. 286, 36281-36290. doi:10.1074/jbc.M111.288092.

Davidsen, J.M., Bartley, D.M., and Townsend, C.A. (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: 5. doi: 10.1021/ja307710d.

Felnagle, E.A., Barkei, J.J., Park, H., Podevels, A.M., McMahon, M.D., Drott, D.W. and Thomas, M.G. (2010) MbtH-Like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry*. 49, 8815-8817. doi:10.1021/bi1012854.

Heemstra, J.R., Walsh, C.T., and Sattely, E.S. (2009) Enzymatic tailoring of ornithine in the biosynthesis of the Rhizobium cyclic trihydroxamate siderophore vicibactin. *J Am Chem Soc.* 131:42. doi: 10.1021/ja9056008.

Imker, H.J., Krahn, D., Clerc, J., Kaiser, M. and Walsh, C.T. (2010) *N*-acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF. *Chem Biol*. 17, 1077-1083. doi: 10.1016/j.chembiol.2010.08.007.

Jiang, G., Zuo, R., Zhang, Y., Powell, M.M., Zhang, P., Hylton, S.M., et al. (2018) One-Pot Biocombinatorial Synthesis of Herbicidal Thaxtomins. *ACS Catal.* 8, 10761–10768. doi:10.1021/acscatal.8b03317.

Koivunen, M., and Marrone, P. (2013) *Uses of Thaxtomin and Thaxtomin Compositions as Herbicides.* U.S. Patent No. US8476195 B2.

Lautru, S., Oves-Costales, D., Pernodet, J.L. and Challis, G.L. (2007) MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. *Microbiology.* 153, 1405-1412. doi:10.1099/mic.0.2006/003145-0.

Leep D., Doricchi, L., Perez Baz, M.J., Millan, F.R., and Fernandez Chimeno, R.I. (2010) *Use of thaxtomin for selective control of rice and aquatic based weeds.* WO Patent No 2010/121079 A3.

Li, Y., Tahlan, K., and Bignell, D.R. (2020). Functional Cross-Talk of MbtH-Like Proteins During Thaxtomin Biosynthesis in the Potato Common Scab Pathogen *Streptomyces scabiei*. *Front Microbiol*. 11, 585456. doi: 10.3389/fmicb.2020.585456.

McMahon, M.D., Rush, J.S. and Thomas, M.G. (2012) Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *J Bacteriol*. 194, 2809-2818. doi:10.1128/JB.00088-12.

Mori, S., Green, K.D., Choi, R., Buchko, G.W., Fried, M.G. and Garneau-Tsodikova, S. (2018) Using MbtH-like proteins to alter the substrate profile of a nonribosomal peptide adenylation enzyme. *Chembiochem*. 19, 1-10. doi:10.1002/cbic.201800240.

Otten, L.G., Schaffer, M.L., Villiers, B.R., Stachelhaus, T., and Hollfelder, F. (2007) An optimized ATP/PP_i-exchange assay in 96-well format for screening of adenylation domains for applications in combinatorial biosynthesis. *Biotechnol J.* 2, 232-240. doi: 10.1002/biot.200600220.

Wolpert, M., Gust, B., Kammerer, B. and Heide, L. (2007) Effects of deletions of *mbtH*-like genes on clorobiocin biosynthesis in *Streptomyces coelicolor*. *Microbiology*. 153, 1413-1423. doi:10.1099/mic.0.2006/002998-0.

Zhang, W.J., Heemstra, J.R., Walsh, C.T. and Imker, H.J. (2010) Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins. *Biochemistry.* 49, 9946-9947. doi: 10.1094/MPMI-04-16-0068-R.

Zolova, O.E. and Garneau-Tsodikova, S. (2012) Importance of the MbtH-like protein TioT for production and activation of the thiocoraline adenylation domain of TioK. *MedChemComm*. 3, 950-955. doi:10.1039/c2md20131c.