Cell-Free *In vitro* Protein Synthesis of Polyketide Synthase Proteins for Production of Natural Products

CHRISTOPHER ANTONIO SARMALES-MURGA

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Master in Chemistry

Department of Chemistry and Biomolecular Sciences
Faculty of Science
University of Ottawa

Abstract

Heterologous expression of multigene biosynthetic pathways is an essential tool in the study of natural product biosynthesis. Due to its *in vivo* nature, this process is often limited by, for example, toxicity of the encoded natural product or its biosynthetic proteins, or competition of the biosynthetic proteins with other cellular enzymes for key small molecule building blocks. Cell-free *in vitro* transcription and translation can overcome some of these limitations. Natural product toxicity is rendered moot in cell-free systems since they are not alive and contain only the necessary proteins, rRNAs, cofactors, substrates, and energy sources for transcription and translation of proteins. As exogenous chemicals can be easily added to the system, building blocks supply issues can be readily solved. We thus investigated using cell-free protein synthesis (CFPS) to biochemically reconstitute the biosynthetic pathway for the fungal polyketide natural product monocillin II. Significant optimization enabled cell-free expression of the full-length monocillin II polyketide synthase (PKS) proteins Rdc5 and Rdc1 directly from plasmids containing their genes under control of the T7 promoter. Correct post-translational modification of the apo-acyl carrier protein domain of the PKS proteins was confirmed by SFP-mediated transfer of a fluorescently modified phosphopantetheinyl group from a chemically modified CoA analog. Unfortunately, treatment of the CFPS produced holo-PKS proteins with their native substrates, malonyl-CoA and NADPH, did not lead to the expected production of detectable levels of monocillin II. Our work suggests that while the CFPS system can generate full length PKS proteins that are sufficiently folded to be recognized, and post-translationally modified by SFP, one or more of the required catalytic domains on these large multidomain proteins is in an inactive state, preventing production of the final product. Identifying non-functional domains, and addressing the issue, may make CFPS an appealing strategy for characterizing PKS biosynthetic gene clusters and prototyping engineered PKS systems.

Acknowledgements

I would like to primary thank Dr. Christopher N. Boddy for giving me the opportunity to continue my studies, and to allow me to continue pursuing my interest and passion of research. I've learned a significant amount of new techniques and skills which make me a better researcher and student than I was when I first started my Masters.

It's been a long and arduous endeavor with lots of up and downs, mostly downs, which is why I would like to also thank the rest of The Boddy Lab members. They took me in as one of their own and I've grown fond of them all, and definitely enjoyed my time with them. They've helped me tremendously in term of answering any questions I would have had and/or helped me improve any skill or techniques to be more efficient and/or to conduct experiments better. Additionally, they've helped pass the time and keep me sane when things would go array or when nothing would seem to be working. I respect them all and hope they succeed in whatever endeavours they pursue.

Through the combined effort of Dr. Christopher Boddy, and the rest of the lab, I've been able to complete my master's study and to complete my research project. When I move onto my PhD, I will definitely take the knowledge and teachings everyone has bestowed upon me and use it to continue to grow and succeed as a researcher, while bestowing knowledge, teaching and helping those around me.

Table of Contents

Table of Figures

Table of Tables

Table of Abbreviations

ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
CE	Carboxylesterase
kDa	kilodalton
EPI	Epoxyketone proteasome inhibitors
GDH	Glucose dehydrogenase
HIS	Hexahistidine
HPLC	High Pressure Liquid Chromatography
Kan	Kanamycin
LC - MS	Liquid Chromatography - Mass Spectroscopy
mRNA	Messenger Ribonucleic Acid
NADPH	Nicotinamide Adenine Diphosphate Reduced
NMR	Nuclear Magnetic Resonance
rNTPs	Ribonucleotide triphosphate
TAGless	A protein lacking a tag
tRNA	Transfer Ribonucleic Acid
UV	Ultraviolet
PPTase	Phosphopanetheinyl Transferase
RNAP	RNA Polymerase
SDS - PAGE	Sodium dodecyl sulfate - Polyacrylamide Gel Electrophoresis
PKS	Polyketide Synthase
hrPKS	Highly Reducing PKS
nrPKS	Non-Reducing PKS
KS	Ketosynthase
M/AT	Malonyl/ Acyl Transferase
ER	Enoyl Reductase
DH	Dehydratase
KR	Keto Reductase
ACP	Acyl Carrier Protein
CoA	Coenzyme A
TE	Thioesterase
CFPS	Cell-Free Protein Synthesis
PURExpress	PURExpress® *In vitro* Protein Synthesis Kit
BGC	Biosynthetic Gene Cluster

Chapter 1: Heterologous Expression of Proteins

1.1…Importance of Heterologous Expression

Natural product biosynthesis occurs in all organisms, producing secondary metabolites for specific uses and functions within the organisms. These compounds have a wide range of biological activity, including antibiotic and anticancer activity. Their potent biological function coupled with their diverse and complex structures makes their biosynthesis particularly interesting.[1] An essential tool in understanding biosynthesis is the heterologous expression of a biosynthetic gene clusters (BGC), which enables production of the biosynthetic proteins and thus the secondary metabolite in a new (heterologous) host.

Heterologous expression is an essential tool in the arsenal of natural product biosynthesis researchers. It is the most effective experiment for confirming that a proposed BGC is responsible for production of a specific natural product. For example, heterologous production of the natural product epothilone in *Streptomyces coelicolor* was used to unambiguously confirm that the gene cluster identified from the deltaproteobacteria *Sorangium cellulosum* was in fact the epothilone BGC.[2]

Epothilone, with a derivative called Ixabepilone as an approved anticancer drug, is very therapeutically important and an example of why it's important to investigate biosynthetic pathways, as they can give rise to novel therapeutics such as epothilone.

Heterologous expression also plays a key role in deciphering how a biosynthetic pathway works and allows assigning of function to the various genes in the pathway. Heterologous expression enables genetic manipulation of the gene cluster, either through knockout or knockdown experiments, especially when the native producer lacks molecular biology tools such as the ability to be transformed, or to readily undergo homologous recombination. For example, Abe and coworkers used heterologous expression in *Aspergillus oryzae* with demethoxyviridin biosynthetic genes to characterize the enzymes that are responsible for the unusual pregnane side-chain cleavage in this biosynthetic pathway (**Figure 2**).[3]

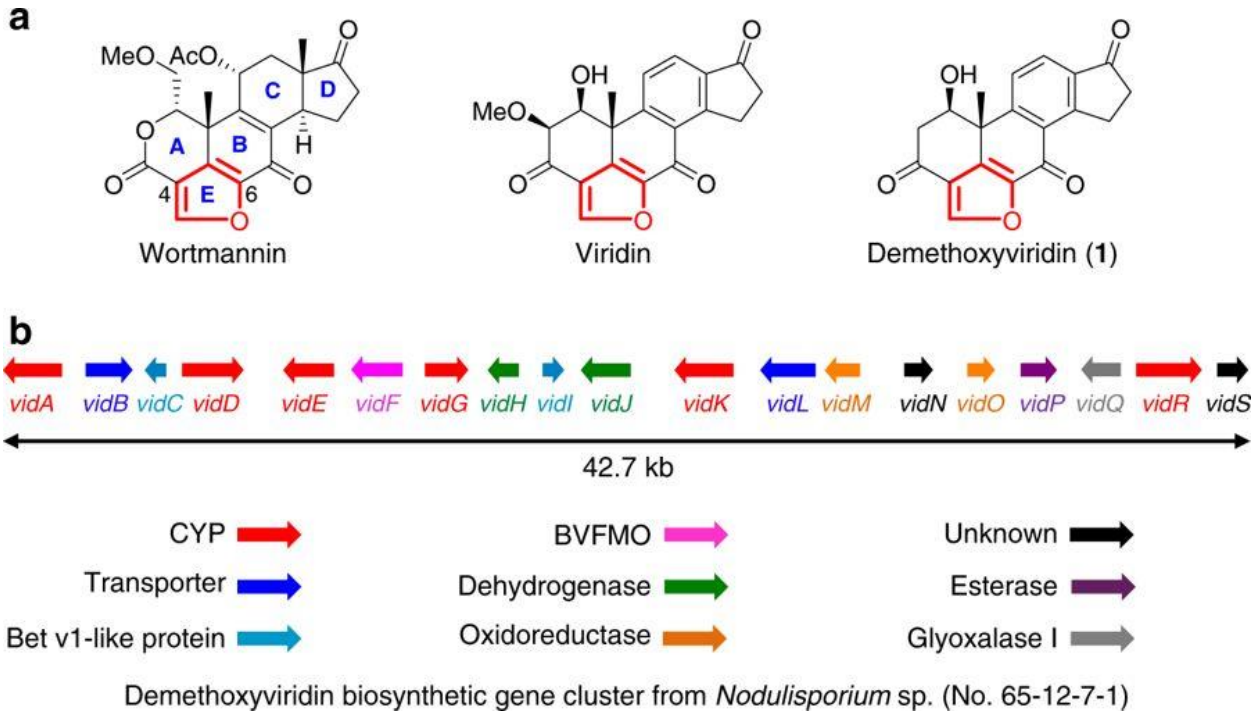


Figure 1: Genetic Overview of Demethoxyviridin[3]
Representative furanosteroids and biosynthetic gene cluster of demethoxyviridin (1). **a** Structures of wortmannin, viridin, and demethoxyviridin (1). **b** Gene map of the demethoxyviridin biosynthetic gene cluster from Nodulisporium sp. (no. 65-12-7-1), consisting of 19 genes from vidA (g3266) to vidS (g3284). The arrow indicates the direction from the start to the stop codon. Different types of genes are indicated by different colors and among them red arrows indicate six CYP genes. **Image republished from open access article "Biosynthetic pathway for furanosteroid demethoxyviridin and identification of an unusual pregnane side-chain cleavage", Wang et al., published in Nature Communications, 2018**.

It's also possible to heterologously express individual proteins of a biosynthetic pathway and combine them in an enzymatic assay to determine if they work sequentially to produce the next

precursor of the biosynthetic pathway. An example of this is the production of Radicicol through heterologous expression, isolation and purification of it's biosynthetic proteins to produce each precursor *in vitro* until Radicicol is ultimately produced.[4]

Perhaps one of the more exciting examples of the power of heterologous expression is the recent use of this tool to discover new compounds from the metagenome. By incorporating fragments of the metagenome from soil samples into heterologous hosts it is now becoming possible to produce and characterize novel natural products and/or biosynthetic pathways. An example is the discovery of epoxyketone proteasome inhibitors (EPI) natural products from the soil metagenome. EPIs inhibit the 20s proteasome through irreversible binding, which leads to a toxic accumulation of polyubiquitinated proteins in the cell.[5]

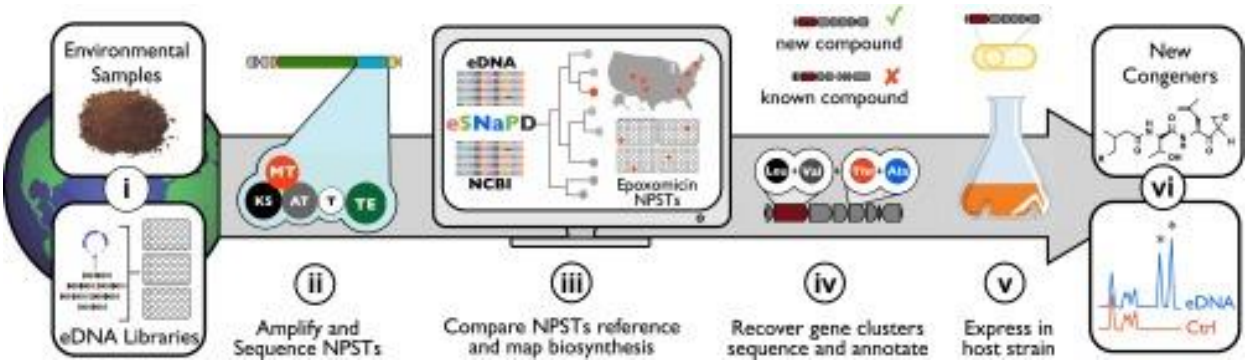


Figure 2: Congener discovery using eSNaPD[5]
(i) eDNA is extracted from samples collected around the globe; these can be archived as large insert libraries if desired. (ii) NPSTs are then generated by sequencing PCR amplicons amplified from eDNA templates with degenerate primers that target conserved biosynthetic motifs. (iii) Analysis of NPST data using eSNaPD identifies NPSTs that derive from biosynthetic gene clusters of interest; these are mapped to collection locations or positions within arrayed libraries using position information incorporated in the PCR primers. (iv) Biosynthetic gene clusters of interest are then recovered from arrayed libraries and sequenced. Bioinformatics analysis of annotated eDNA gene clusters is then used to prioritize clusters for heterologous expression studies. (v) Prioritized gene clusters are transferred to a laboratory-friendly host for heterologous expression. (vi) LCMS and/or biological activity profiles of strains harboring eDNA clusters are compared with a vector control strain to identify new metabolites for purification, structure elucidation, and bioactivity studies.

The authors amplified ~1 x 10^6 unique environment sequences acquired from soil using degenerate primers that target adenylation domains, a common biosynthetic motif of non ribosomal peptides, and ketosynthase domains that are involved in polyketide biosynthesis.

Following amplification, the EPI biosynthetic pathways underwent bioinformatics evaluation using multiple tools, such as antiSMASH, to predict the structures of the compounds encoded by these pathways. From the many analyzed sequences, two were selected for heterologous expression, with *Streptomyces* as the host, and two novel potent 20S proteasome inhibitors, AR412 and AR456, were discovered (***Figure 2***).[5]

As can be seen from these examples, heterologous expression has become a well-established technique in natural product biosynthesis. It has enabled the discovery of new natural products, biosynthetic pathways, and biosynthetic enzymes. While highly effective, heterologous expression does have significant challenges that often limit its utility.

1.2…Challenges of Heterologous Expression

Not surprisingly, many of the challenges of heterologous expression can be attributed to selection of the host organism. In many cases, the heterologous host may not be able to express one or more of the biosynthetic genes in a functional state. For example, rabbit liver carboxylesterase (CE) was expressed within *E. coli*, *S. cerevisiae* and other expression hosts to compare the abilities of the host to produce folded, active CE.[6] This study showed that although significant amounts of the proteins were detected in *E. coli*, the recombinant protein had poor solubility with very low levels of enzyme activity. Within *S. cerevisiae* no protein was detected. Functionally active protein was only detected when the CE was heterologously expressed in *Pichia pastoris*.[6] Thus, while protein may be active and functional when expressed in one host, it may not be expressed, or expressed in an active form, in another host.

A challenge unique to heterologous biosynthetic pathway expression is that the building blocks required to produce its natural product may not be present in the heterologous host. Coenzymes and substrates used in secondary metabolite biosynthesis do not necessarily have broad distribution across the kingdom of life, and as such they may or may not be present within a particular host. Furthermore, the enzymatic machinery for correct post-translational modification of biosynthetic proteins often is highly selective to the producing organism and related hosts. As such, the heterologous host may not be able to correctly post-translationally modify biosynthetic proteins. A clear example occurs with acyl carrier proteins (ACPs) from polyketide biosynthetic pathways. ACPs are important polyketide synthase enzymes and require a phosphopantetheinyl group for activity, which is added post-translationally, to hold the growing polyketide intermediate. While all bacteria possess phosphopantetheinyl transferase (PPTase) enzymes required for adding the phosphopantetheinyl group from coenzyme A (CoA) onto fatty acid synthase ACPs, these enzymes typically do not recognize polyketide synthase ACPs.[7] As such, many heterologous hosts, such as *E. coli*, are unable to post-translationally modify polyketide synthase ACPS and thus cannot produce polyketide natural products heterologously. To address this serious problem, *E. coli* has been genetically engineered to produce the *Bacillus subtilis* PPTase SFP, which has broad substrate tolerance for polyketide synthase ACP domains.[8] While the solution is effective for *E. coli* expression of ACP containing biosynthetic pathways, other heterologous hosts suffer from a lack of effective post-translational modification of biosynthetic proteins.

Toxicity is also a significant issue when dealing with heterologous expression. Many heterologously expressed proteins can be highly toxic to the producing organism. For example,

the human antimicrobial protein CAP37, encoded by the gene *AZU1*, cannot be heterologously expressed in bacteria in native form. In the context of heterologous expression of biosynthetic pathways, toxicity is further complicated by the inherent biological activity of the natural products and often of its intermediates.[9] These compounds, and intermediates, can hinder the viability of the host.[1] Thus, when the biosynthetic pathway for the antibiotic natural product kanamycin (Kan), is introduced into a bacteria lacking the resistance mechanism for Kan, the expression of the pathway will simply kill the host.[10]

In addition, the misfolding of proteins, along with toxicity of the encoded proteins and biosynthetic products, and heterologous expression of biosynthetic pathways suffers from additional challenges. A significant challenge is ensuring the heterologous host effectively transcribes all of the biosynthetic genes. Lacking effective transcription of a single gene can prevent an entire pathway from functioning, thus amplifying the challenge of a problem that can typically be readily solved for a single gene system. As such, this is a significant problem, with many of the common strategies for heterologous expression being based on specific solutions to this fundamental challenge.

1.3…Strategies for Heterologous Expression of BGCs

Perhaps the most common heterologous expression strategy employed is introduction a biosynthetic gene cluster into a very related host.[11] As many bioactive natural products are produced by actinobacteria, the model actinobacteria *Streptomyces coelicolor* and the related strain *Streptomyces lividans* are commonly used[12]. A major advantage of this strategy is that native BCGs can be used for heterologous expression, as typically the promoters used in actinobacterial BGCs are recognized and expressed by these two streptomyces. The ability to

use native biosynthetic pathways is particularly appealing since these pathways typically contain six or more genes which span, in many cases, 25-35kb in size. Thus, capturing the native BGC from a genomic library[13] or via the more modern transformation assisted recombination[14] can readily lead to the required vector. This strategy for heterologous expression has been highly successful in producing a wide number of polyketide natural products. It, however, can be challenging to implement large BGCs (>35 kb), such as those encoding very large modular polyketide synthases and non-ribosomal peptide synthetases.

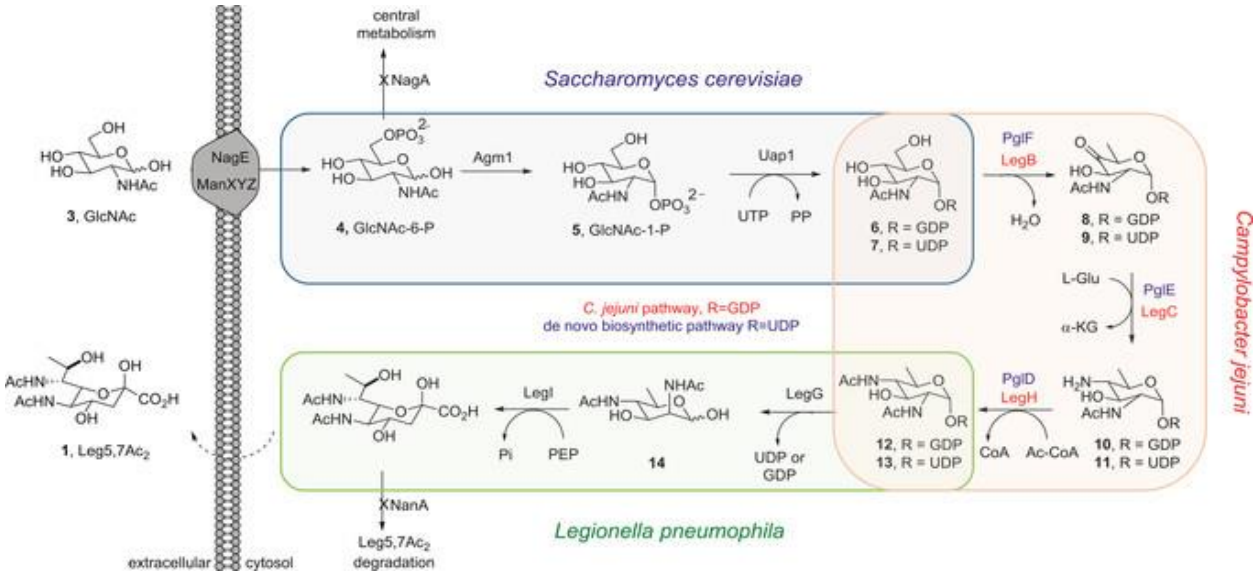


Figure 3: De novo biosynthetic pathway of Leg5,7Ac2[4]
Production in E. coli. Enzymes listed in blue are from the engineered UDP-linked pathway and those in red from the native *C. jejuni* GDP-linked biosynthetic pathway.

An alternative strategy for heterologous expression relies on re-cloning each gene of a BGC under the control of a promoter known to function in the native host. *E. coli* has proved to be a good host for this approach as strategies for ensuring transcription are well studied and documented. For example, the introduction of multiple biosynthetic genes, each under the

control of the T7 promoter, into *E. coli* enabled it to produce legionaminic acid (***Figure 3***). Legionaminic acid, a nonulosonic acid, is found on cellular surfaces of some bacteria with its presence being correlated with virulence in humans. While there have been total synthesis methods to produce legionaminic acid, they're highly demanding and very low yielding. Heterologous expression made it possible to produce significant quantities of Legionaminic acid (121 mgL^{-1} of cell culture).[4]

One of the newest strategies for heterologous expression, has been to activate transcription of an otherwise silent unexpressed BGC by expression of a transcriptional factor (***Figure 4***). An excellent example of this can be seen in the heterologous production of the polyketide oxytetracycline in *E. coli*. Direct introduction of the *Streptomyces rimosus* oxytetracycline gene cluster into *E. coli* produced no oxytetracycline. This was attributed to the lack of transcription of one of the key operons. Sigma (σ) factors control the specificity of gene transcription by binding to RNA polymerase (RNAP) and recruiting it to promoter sequences upstream of the gene to be transcribed.[15] Within the context of this example (***Figure 4b***), these sigma factors were heterologously expressed in the presence of the oxytetracycline BGC and transcription of the key *oxyB* containing operon was quantified. Expression of σ^{54} significantly increased the transcript of the *oxyB* containing transcript, and enabled production of oxytetracycline in *E. coli.*[16] Ultimately, the productivity of this system was limited by the inherent antibacterial activity of the oxytetracycline produced from the BGC.

While these examples showcase the diverse strategies for heterologous expression, it is clear that they do not address all the challenges associated heterologous production of natural products. As all are cell based, they invariably suffer from toxicity of either proteins from the

biosynthetic pathway or the encoded natural product. Moreover, they all require the generation of complex vectors to introduce the BGC into the host. Lastly, they are dependent on the production, and presence, of the natural product building blocks by the heterologous hosts. This is frequently rate limiting, as was recently shown in the heterologous production of cannabinoids in yeast.[17] A number of these challenges can potentially be addressed by the use of a cell-free *in vitro* transcription, and translation, approach to heterologous expression.

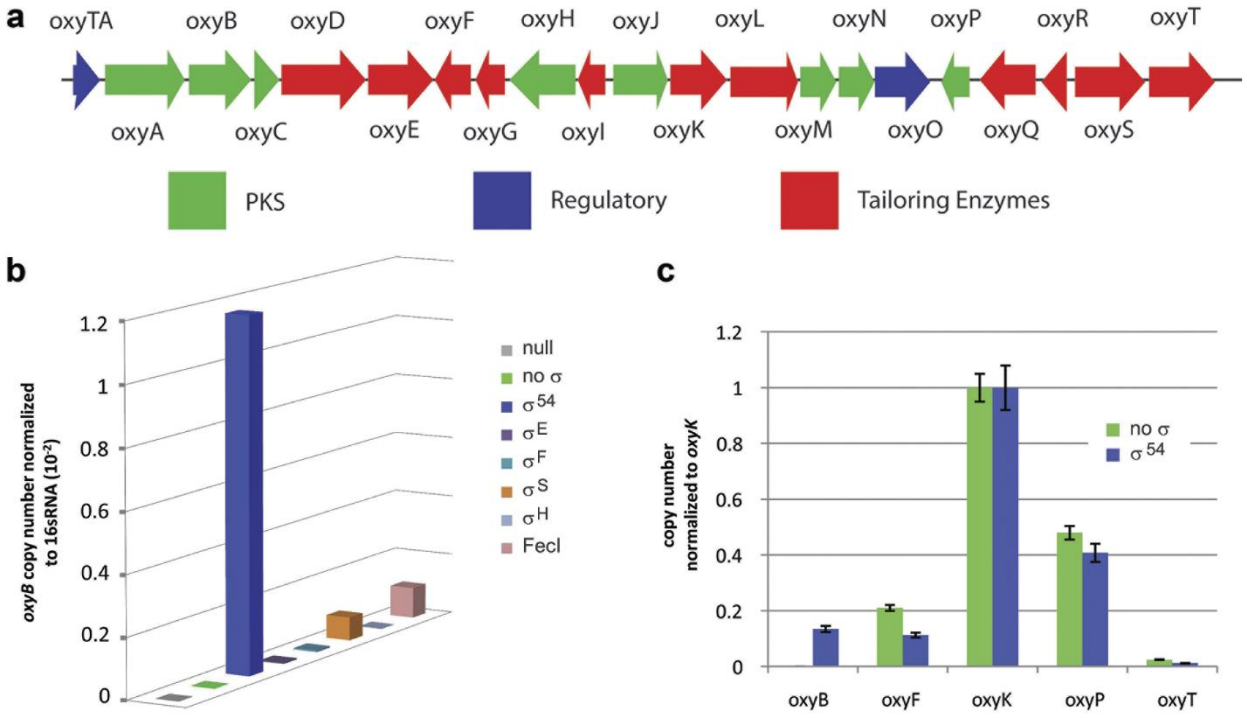


Figure 4: SYBR-green based qPCR analysis shows that transcription limits heterologous production of oxytetracycline in E. coli.[16]
(a) The 32 kb oxytetracycline biosynthetic gene cluster is shown. Five putative operons, *oxyABCDE, oxyIHGF, oxyJKLMNO, oxyRQP*, and *oxyST* are predicted for this gene cluster (b) qPCR analysis shows that over-expression of the alternative sigma factors σ54, σS and FecI enable detectable levels of the oxyB transcript to be produced. Over-expression of no sigma factor, σE, σF and σH do not lead to detectable levels of the oxyB transcript. (c) qPCR analysis shows that over-expression of the alternative sigma factor σ54 lead to detectable levels of transcripts for all five putative operons in the oxytetracycline biosynthetic pathway. In the absence of σ54 over-expression, the oxyB transcript cannot be detected. **Used under open access permission.**

1.4…Cell-Free *In vitro* Protein Synthesis

The first successful attempt to use cell-free extracts was conducted by Eduard Buchner in 1897, where he used the extracts from yeast to convert sugar to ethanol and carbon dioxide. He was awarded the Chemistry Nobel Prize in 1907. He demonstrated that it is possible to use the

extracts, absent of live cells, to conduct *in vitro* experiments [18]. This was taken a step further in 1961 with Nirenberg and Matthaei when they used cellular extracts to decipher the 64 triplet codons in the genetic code by using nucleic acid homo polymers to translate specific amino acids.[19,20] Since then, these cellular extracts have been used as bioreactors to make proteins[21], and refined to develop easy to use protocols and commercial kits for the production of high purity cell-free recombinant proteins.

There are two strategies for the preparing cell-free *in vitro* protein synthesis (CFPS) systems, the complex and the defined systems. The complex variety entails direct lyses of the cells supplying the machinery for *in vitro* transcription and translation in S30 buffer. S30 contains dithiothreitol

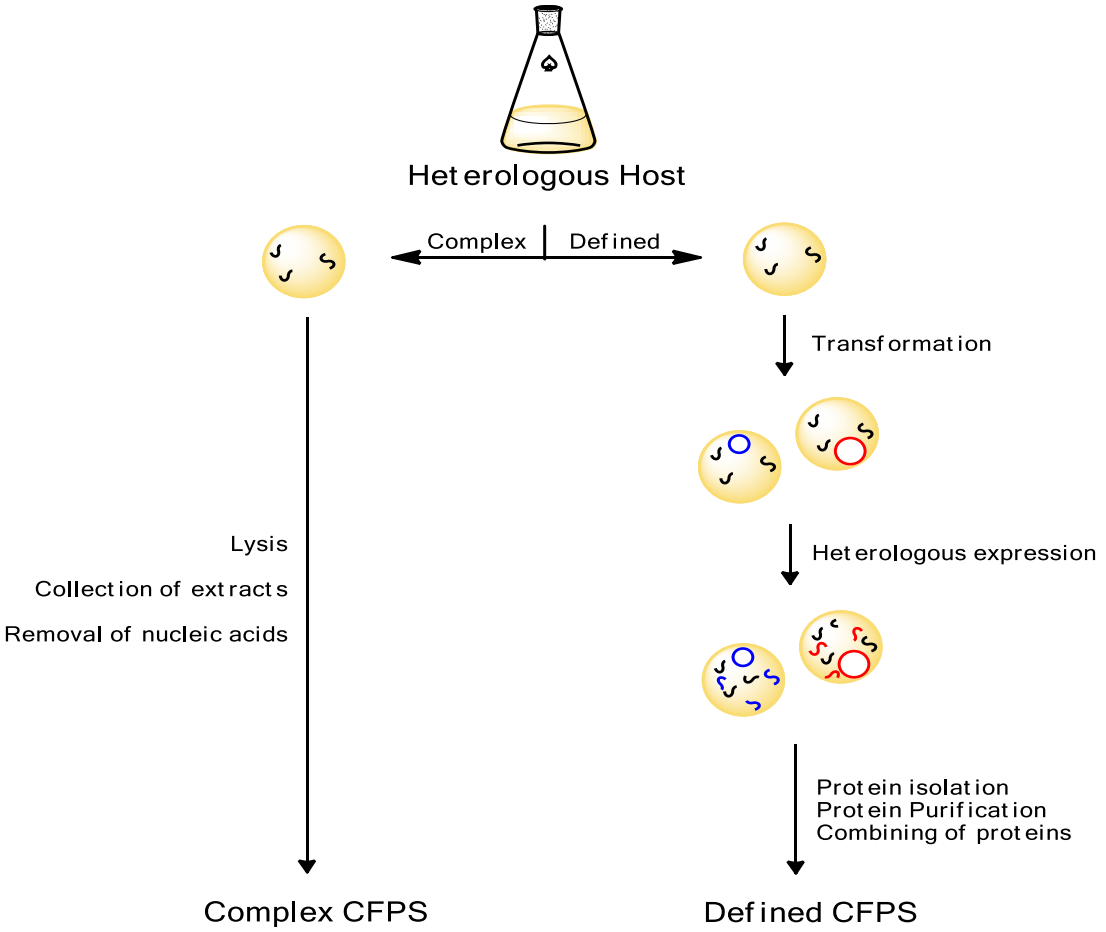


Figure 5: Complex versus Defined CFPS systems.
Complex CFPS systems are generated by direct cell lysis and contain all the soluble cellular components from the original cells. Defined systems contain only purified biochemical proteins, ribosomes, and metabolites needed to effect transcription and translation.

(DTT) and it requires addition of only a few reagents, such as nucleotides, amino acids, a source of ATP, and double stranded DNA containing a gene with the appropriate promoter to initiate *in vitro* transcription, and subsequent translation, of the protein of interest.

Complex systems have several advantages and disadvantages. The simplicity of the preparation is the major advantage to this strategy. It's main disadvantage is the fact that along with the components essential for expression of genetic material, the mixture is composed of many non-essential proteins, cofactors and substrates that can not only influence the expression of desired protein(s), but can potentially impact the ability of the protein products to function correctly. In addition, there can be a noticeable batch-to-batch variability for complex CFPS mixtures that can impact reproducibility, particularly with targets that push the limits of *in vitro* transcription and translation. These disadvantages are relevant to heterologous production of biosynthetic gene clusters. In addition, the complex mixtures generated from these CFPS systems add to the challenge of detecting and quantifying natural product produced from cell-free biosynthesis. Thus, while the complex form of cell-free *in vitro* protein synthesis is a versatile tool for protein expression, it is not ideally suited for heterologous production of biosynthetic gene clusters.[22]

The second strategy for CFPS has a defined set of components, making the system much less complex when compared to the complex strategy. The defined form, which can be purchased as a kit, requires addition of all the recombinantly expressed and purified proteins essential for transcription, translation and ATP production to power these activities.[23] In addition to these purified proteins, purified ribosomes are also added to the system, thus providing all the components needed for *in vitro* transcription and translation

A clear advantage of the defined kit is the well-defined set of concentrations of required proteins, cofactors, building blocks as well as the lack of extraneous protein components, which results in a much cleaner *in vitro* assay. In addition, to facilitate isolation of the cell-free produced proteins, typically a tag is incorporated into the proteins used for transcription and translation, enabling their removal by affinity chromatography.[22] The primary drawback of this strategy is the cost. An example of the defined system is the PURExpress® *In vitro* Protein Synthesis Kit, from New England Biolabs. In this kit, recombinant proteins for transcription, translation and ATP production were individually expressed and purified through use of a hexahistidine (HIS) tag. Thus, all protein components of the kit contain a HIS tag. Transcription is performed by the T7 polymerase, thus genes possessing an upstream T7 promoter can be readily transcribed by this system.[23,24]

Cell-free heterologous expression of biosynthetic pathways is under-explored in the literatures though it has a number of potential advantages. As the cell-free system lacks the vast majority of cellular processes, it is less likely to be impacted by the toxicity of the biosynthetic proteins or the natural products they encode. Moreover, as the cell-free systems lacks encapsulation by a membrane, the addition of target DNA, natural products, building blocks and atypical cofactors is simplified. Furthermore, as no other cellular processes are competing for resources, it is possible to express a significant amount of proteins in very short time frames.

While there are examples of producing simple natural products by cell-free heterologous expression,[25] there are very limiting examples of producing natural products encoded by complex multi-domain proteins like polyketide synthases or non-ribosomal peptide synthases. The complex biosynthetic pathways produce some of the more potent and biologically relevant

natural products known. Cell-free protein synthesis was used to investigate the expression of two non-ribosomal peptide synthase biosynthetic proteins, GrsA and GrsB1, from the gramicidin S biosynthetic pathway.[26] Together GrsA and GrsB1 are known to produce a d-Phe-l-Pro diketopiperazine product. Cyclic dipeptide products like this are produced across a wide range of organisms including bacteria, fungi and plants.[27] Using a complex cell-free production system, the authors were able to produce detectable quantities of diketopiperazine from their cell-free experiment, suggesting production of complex biosynthetic pathways was feasible.

1.5…Goals of the Project

While heterologous expression is an essential tool for biosynthetic pathway expression, it is still plagued by problems with toxicity of the biosynthetic proteins, the natural products they produce, as well as challenges related to cloning of the BGCs. Cell-free protein synthesis possesses a number of advantages for biosynthetic pathway expression, including reduced impact of protein and metabolite toxicity, simplified cloning requirements, as well as in the case of the defined cell-free protein expression systems, a background matrix of reduced complexity facilitating metabolite detection. Preliminary data with simple biosynthetic pathways, such as the violacein pathway,[28] and more complex pathways such as a subsection of the gramicidin biosynthetic gene cluster[29] shows that cell-free biosynthesis can produce the encoded natural product. In this project, I aim to determine if cell-free protein synthesis can be used to express a polyketide biosynthetic pathway. If successful, this would represent the first example of producing function polyketide synthase proteins by cell-free methods.

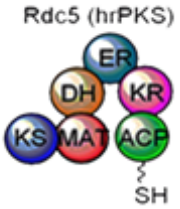


Figure 6: Example of a multi-domain PKS protein Rdc5, a large (260 kDa) highly reducing PKS protein from Radicicol Biosynthetic Pathway. Composed of keto-synthase (KS), malonyl/acyl transferase (MAT), dehydratase (DH), enoyl reductase (ER), keto reductase (KR), and acyl carrier protein (ACP). Each have individual activities that together produce a natural product.

Polyketide Synthase are the key proteins responsible for the production of polyketide natural products. These pathways can exist in many different architectures, including type I, type II, and type III polyketide synthases. The work in this thesis focuses on type I polyketide synthases. These are large, multi-domain containing proteins (***Figure 6***). The core catalytic domains common to all type I polyketide synthases are ketosynthase (KS) domain, acyltransferase (AT or MAT) domain, and an acyl carrier protein (ACP). Additionally, polyketide synthase proteins can possess up to three reductive processing domains, including a ketoreductase (KR), dehydratase (DH) and enoylreductase (ER). These catalytic domains function similar to fatty acid synthases to add an acetate unit from malonyl-CoA onto a growing polyketide chain through a decarboxylative Claisen condensation. As they work together to affect this biological chemistry, this group of catalytic domains is often referred to as a module. For bacterial polyketide biosynthetic pathways, there are typical one module worth of catalytic domains for each acetate unit added to the growing polyletide, thus the biosynthetic pathways can be extremely large. For example, the pathway that produce erythromycin possesses six unique modules with over 25 unique domains. Because of the size and complexity of bacterial polyketide pathways, I chose to tackle a more simplified fungal polyketide synthase system.[1,30]

In fungal polyketide biosynthesis, typically only one or two modules of catalytic domains are used. These catalytic domains are used iteratively, delivering multiple equivalents of acetate from malonyl-CoA to the growing polyletide. Typically, there is a highly reducing polyketide synthase (hrPKS), which contains the three reductive processing domains as well as the KS, MAT, and ACP, as well as a non-reducing polyketide synthase (nrPKS) that does not possess these reductive catalytic domains.[1,30] These two proteins work in concert to generate complex

natural products from a very streamlined set of proteins. As such, these fungal polyketide pathways appeared to be an appealing starting point for investigating cell-free polyketide biosynthesis. In the following chapter, I thus describe my work to produce the fungal polyketide monocillin II through cell-free biosynthesis.

Chapter 2: Precursor Biosynthetic Pathway of Radicicol

2.1…Introduction

The key goal of this study is to produce for the first time a PKS biosynthetic pathway via a cell-free expression system. The fungal polyketide synthase pathway encoding the natural product radicicol, and its precursor Monocillin II[31] was selected as the target pathway due to its relatively simple biosynthetic pathway. It requires only two polyketide synthase proteins to make this complex bioactive compound that competes with ATP for binding in the ADP/ATP binding pocket of Hsp90.[31] This biological activity is significant as Hsp90 plays an important role in many cellular processes with activities involved in regulation of cell cycle, cell growth, cell survival, apoptosis, angiogenesis and lastly oncogenesis.[32] By affecting these cellular processes, the compound, and by extension its biosynthetic pathway, is a significant subject for further characterization.

Radicicol is biosynthesized by four key proteins known as Rdc5, Rdc1, Rdc2 and Rdc4. Rdc5 and Rdc1 are type I, iterative PKS proteins. Rdc2 is a FADH dependant halogenase and Rdc4 is an oxidase. Of the PKS proteins, Rdc5 is a highly reducing PKS, meaning that it will is reduce a significant portion of the elongated polyketide chain prior to transferring it to Rdc1. Rdc1 is a non-reducing PKS, that further elongates this chain and ultimately cyclizes it to form monocillin II, a precursor of Radicicol. To convert monocillin II to radicicol, monocillin II is chlorinated by Rdc2 to make Pochonin D, followed by epoxide formation via Rdc4, generating radicicol.[31]

The biosynthetic gene cluster responsible for formation of radicicol and monocillin II is well characterized.[31] The cluster was unambiguously identified by heterologous expression in *S.*

cerevisiae and monocillin II has also been produced via *in vitro* enzymatic assay using recombinant Rdc5 and Rdc1, supplemented with 2 mM Malonyl-CoA and NADPH.[31] Malonyl-CoA is the source of all the carbons in monocillin II and NADPH supplies hydride required for reductions by the KR and ER domains. Thus, previous work had clearly identified the biosynthetic genes and showed they could function as recombinant proteins in an *in vitro* setting. This highly supported our proposed cell-free protein synthesis of monocillin II(***Figure 7***).

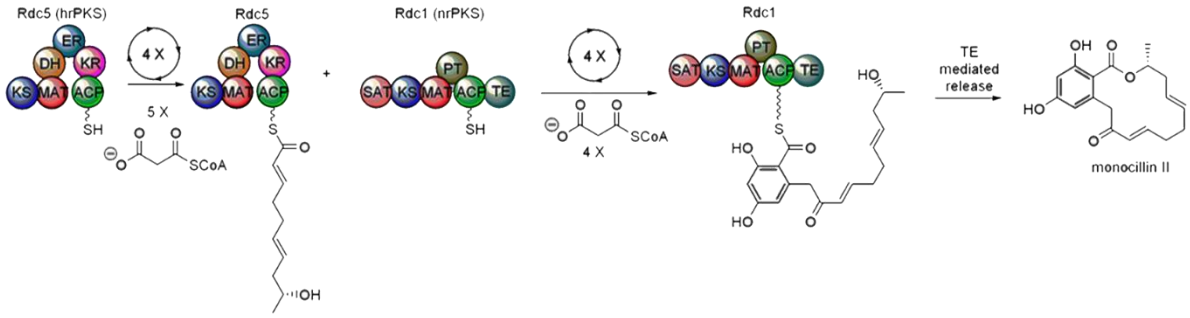


Figure 7: Biosynthetic Pathway of Monocillin II
Displayed are the proteins involved in the production of Monocillin II. Rdc5, a highly reducing PKS, and Rdc1, a non-reducing PKS, work in conjunction to produce Monocillin II using Malonyl-CoA and NADPH.

A clear challenge for this project however was the recombinant production of the monocillin II producing PKS proteins in *E. coli* did not produce soluble proteins. For example, expression of

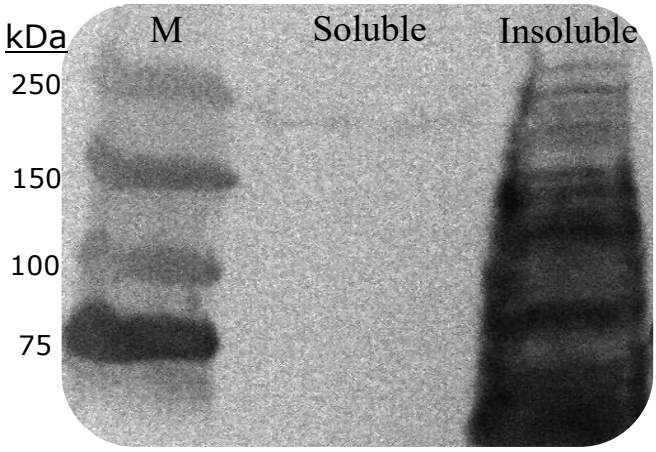


Figure 8: Anti-HIS Tag antibody Western Blot of BAP1 extracts
Anti-hexahistidine (HIS) antibody probed Western Blot of Rdc5 and Rdc1 expression in E. coli BAP1 cells. The expected full-length protein bands for Rdc5 and Rdc1 were not observed in the soluble fraction at 260 kDa and 228 kDa respectively. Denaturing lysis (insoluble fraction) shows significant His tagged protein is present suggesting Rdc5 and Rdc1 undergo aggregation and degradation.

Rdc5 and Rdc1 under the control of the T7 promoter in the *E. coli* strain optimized for polyketide synthase expression, BAP1, did not show any soluble full length Rdc5 and/or Rdc1. (**Figure 8**). Unfortunately, all the proteins are produced as an aggregated, partially proteolyzed, insoluble material.

Once soluble polyketide synthase proteins are produced, their acyl carrier proteins must be post-translationally modified. This post-translational modification is catalyzed by a phosphopantetheinyl transferase, such as the 26 kDa protein SFP from *Bacillus subtilis*.[33] SFP uses coenzyme A (CoA) as a substrate and adds the phosphopantetheinyl arm of CoA onto a seryl side chain of ACP (***Figure 9***). This phosphopantetheinyl arm is essential as it is the covalent attachment site for the growing polyketide product on the PKS enzyme.[7]

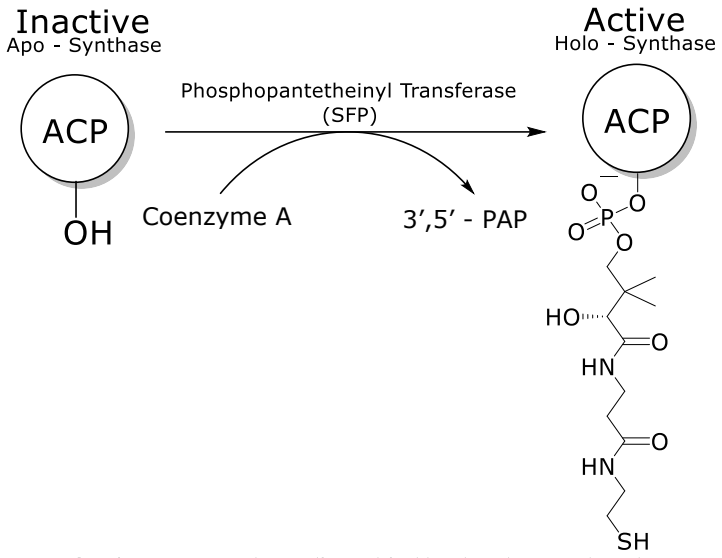


Figure 9: Acyl carrier proteins (ACP) are post-translationally modified by phosphopantetheinylation.
The *B. subtilis* enzyme SFP is a broadly substrate tolerant phosphopantetheinyl transferase which transfers the phosphopantatheinyl group from Coenzyme A on to the active site Ser of an ACP domain.

As the cellular hosts used to produce these cell-free expression systems do not naturally contain SFP, it would need to be exogenously added, along with CoA, to insure that Rdc5 and Rdc1 are correctly modified.

Once the biosynthetic Rdc5 and Rdc1 domains have been expressed in cell-free, and post-translationally modified using SFP and CoA, the final step would be to assay the enzymes for *in vitro* production of monocillin II by treatment with malonyl-CoA and NADPH. Thus, the general overview for this project is visualized in ***Figure 10.***

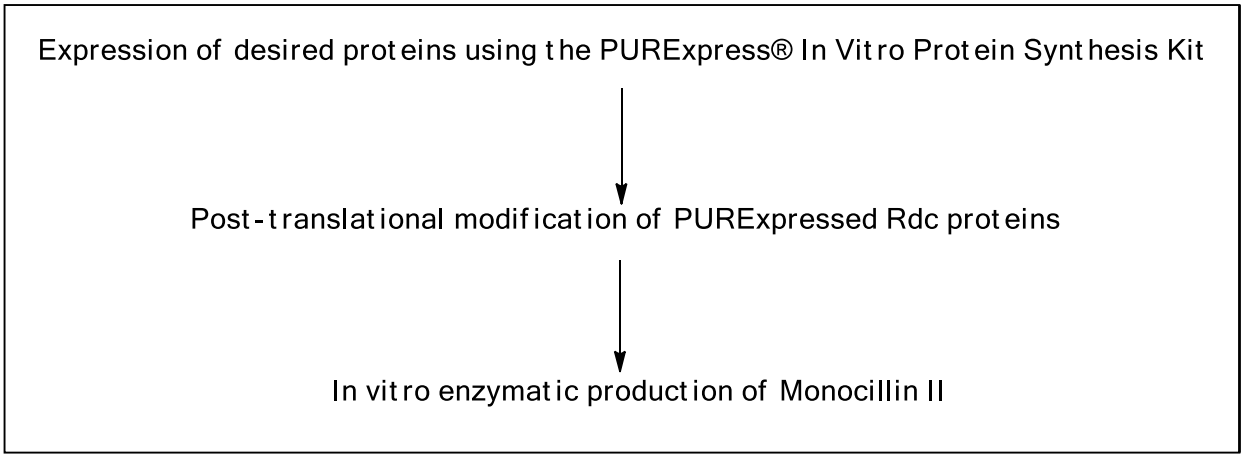


Figure 10: General overview of the study

2.2…Results

2.2.1…Expression of Rdc1 Thioesterase as CFPS proof of concept

As the defined form of CFPS is cleaner, simpler and more consistent, compared to the complex form, it was chosen for cell-free expression. The PURExpress® *In vitro* Protein Synthesis Kit, from New England Biolabs, was used due to its high quality and simplicity of use. The reaction is conducted by mixing two solutions, one containing the purified ribosomes and the other containing the recombinant proteins required for transcription, translation and energy production with the genes under T7 control encoding the desired proteins for expression.[24] (**Figure 11**).

The manufacturer notes that sufficient amounts of proteins can be synthesized, and visualized, post 2 hours of incubation at 37˚C[24] (**Figure 12**).

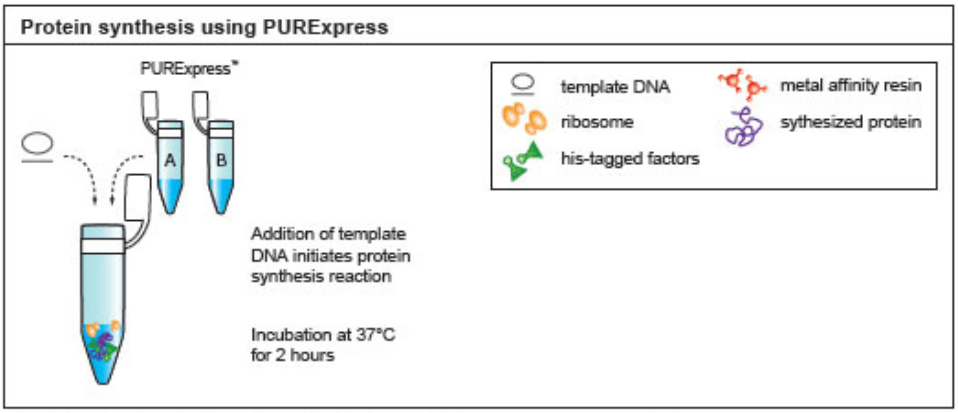


Figure 11: Schematic diagram of protein synthesis by PURExpress[24]

As can be seen in the "no DNA" lane of **Figure 12**, many prominent bands are visualized by SDS-PAGE. These bands correspond to the various components of the PURExpress kit such as T7 polymerase and creatine phosphatase, for example. The following lanes show successful expression of proteins of various sizes using the PURExpress® *In vitro* Protein Synthesis Kit. The desired proteins are indicated with a red dot. This figure shows that proteins between 20-100 kDa can be successfully expressed in only 2 hours.

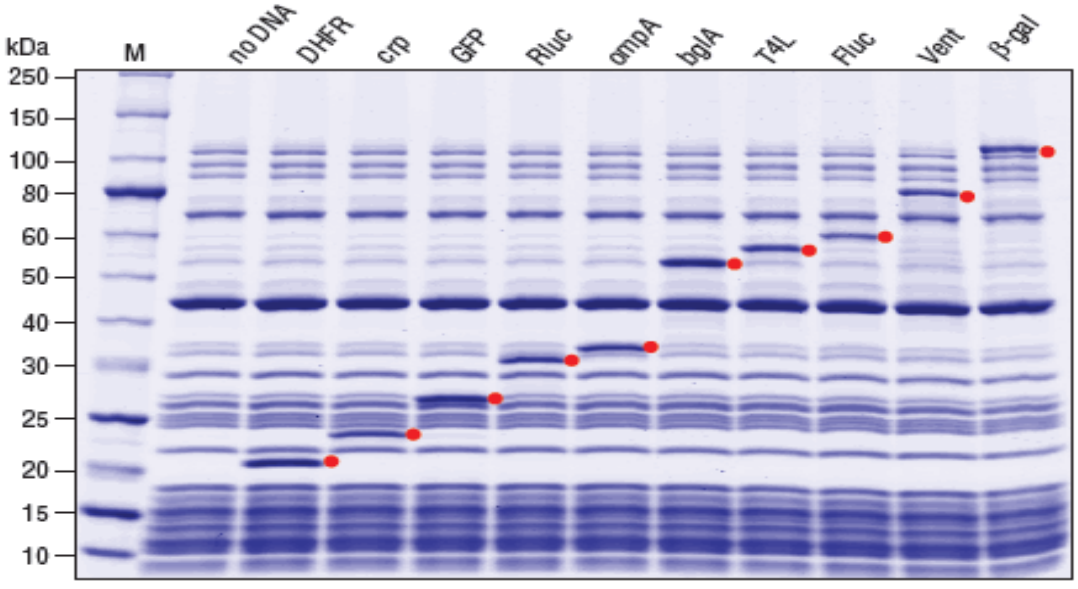


Figure 12: SDS-PAGE analysis of proteins of various sizes produced by the PURExpress® *In vitro* Protein Synthesis Kit.[24]

To evaluate our ability to monitor the enzymatic function of proteins produced via cell-free systems, we tested a single catalytic domain from the monocillin II pathway, the C-terminal thioesterase domain. Previous work had shown that the standalone C-terminal thioesterase domain, RadTE, could effect macrocyclization of a synthetic thioester activated substrate[34] (***Figure 13***).

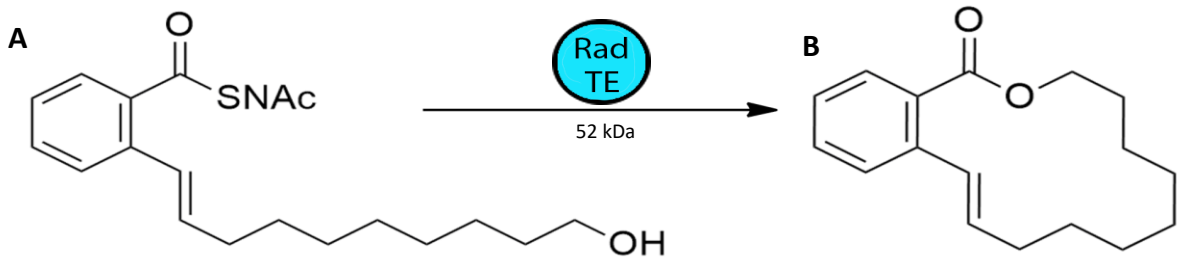


Figure 13: ***In vitro*** **Biochemical assay to evaluate recombinant purified Rdc TE enzymatic activity.**
Substrate (A) was synthesized and provided by Dr. Heberlig.

As the excised RadTE is significantly smaller than Rdc1, 52 kDa vs 228kDa, and as RadTE had been successfully expressed in *E.coli* in a functional form, we were confident this experiment would enable us to determine if cell-free expression conditions produced sufficient protein to perform and detect biosynthetic reactions.

Some of the questions that will also be answered with the proof of concept will be:

1. Can a PKS domain, be expressed using CFPS?
2. Is the recombinant protein product enzymatically active?
3. Are there unexpected challenges in preparing and conducting the enzymatic assays?

Results from this proof of concept experiment are expected to be transferable to and inform the more complex full length PKS proteins.

The components, and respective volumes, of each PURExpress® *In vitro* Protein Synthesis reaction, and some examples of scaled down reactions, can be seen in **Table 1.** ½ and ¼ volume scale of the PURExpress® *In vitro* Protein Synthesis Kit protocol would enable us to determine if scaling down the reaction would produce not only detectable protein, but enough protein to catalyze formation of a biochemical product that can then be detected by HPLC. For the remainder of this thesis, PURExpress® *In vitro* Protein Synthesis will be referred to as "**PURExpress"** and the proteins produced using this method will be referred to as "**PURExpressed**" proteins.

Table 1: Reaction conditions for PURExpress reactions along with 1/2 fraction and 1/4 fraction

	PURExpress	½ PURE	¼ PURE
Solution A	10.0	5.00	2.500
Solution B	07.5	3.75	1.875
DNA	XX.X	X.XX	X.XXX
H_2O	XX.X	X.XX	X.XXX
Total	25.00 µL	12.50 µL	6.25 µL

Using PURExpression kit, RadTE could be readily expressed. SDS-PAGE analysis of PURExpressed RadTE at ½ scale and ¼ scale ***(Figure 14B)*** showed the expected band at 52 kDa. In comparison, the negative control, where no DNA encoding RadTE was added to the PURExpress system, showed no detectable band for the protein.

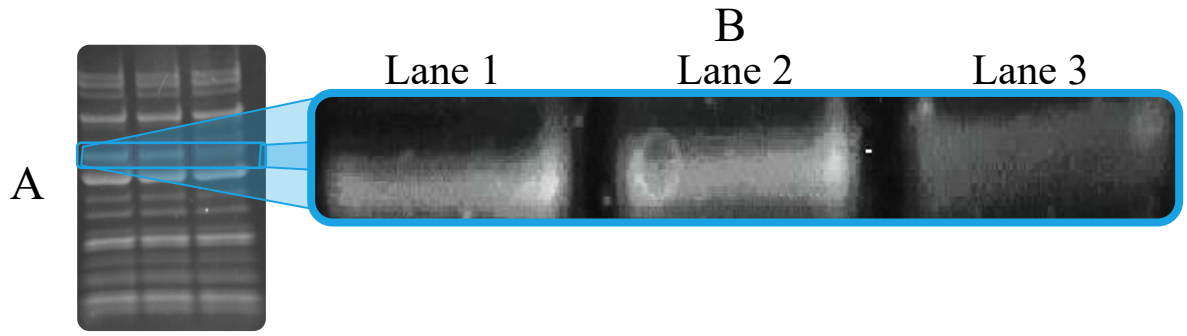


Figure 14: SDS-PAGE of cell free preparations of Rdc1TE.
(A) The PURExpressed proteins were separated by SDS-PAGE and imaged using the Biorad stain free in gel fluorescence system. **(B)** An expansion of the Rdc1TE bands. **Lane 1:** 1/2 PURE protocol (pMRH08); **Lane 2:** 1/4 Pure Protocol (pMRH08); **Lane 3:** Negative Control (Empty pET28**).**

This result confirmed that a band corresponding to full length RadTE could be generated and was dependent on the presence of the gene encoding RadTE. Importantly, we also demonstrated we could scale down the PURExpress kit protocol to increase the number of experiments per kit.

To evaluate if the protein, consistent in size with RadTE, possessed the expected enzymatic activity of RadTE expression, the PURExpressed protein from ¼ of the PURExpress protocol (***Table 1***) was incubated with the 14-membered substrate (***Figure 13a***) for 18 hours at room temperature. A negative control was also performed where ¼ PURExpress assay with no added DNA was incubated with substrate under identical conditions. The negative control , and enzymatic assays, were then analyzed by HPLC.

Analysis of the HPLC chromatograms generated by the enzymatic assay ***(Figure 15)***, showed a prominent peak at ~17 minutes. Within the negative control ***(Figure 16)***, this peak was absent and at ~12 minutes was instead observed. The peak at ~17 minutes is consistent with authentic standard of the macrocyclic product, and the peak at ~12 minutes is consistent with the substrate.

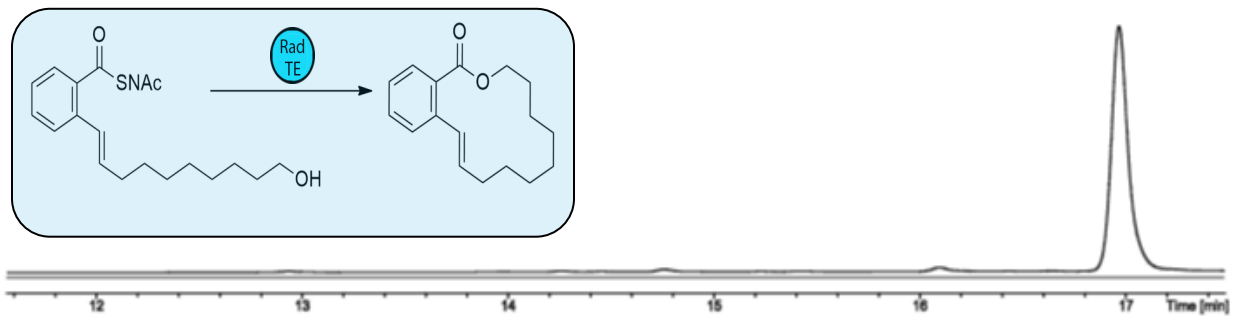


Figure 15: Enzymatic assay HPLC chromatogram of Rdc1 TE with Substrate
PURExpressed Rdc1TE *in vitro* assay with necessary components separated using HPLC and detected using UV, monitored at 280 nm. Outlined in blue is the overview of the *in vitro* enzymatic assay where TE is converting 14-membered substrate to product.

This data shows that although a faint protein band was detected by SDS-PAGE, RadTE was produced at sufficient levels to effect enzymatic conversion of the known substrate into the expected product.[34] This confirms the identity of RadTE as well as validating the scale and protocols for cell-free biosynthetic pathway production.

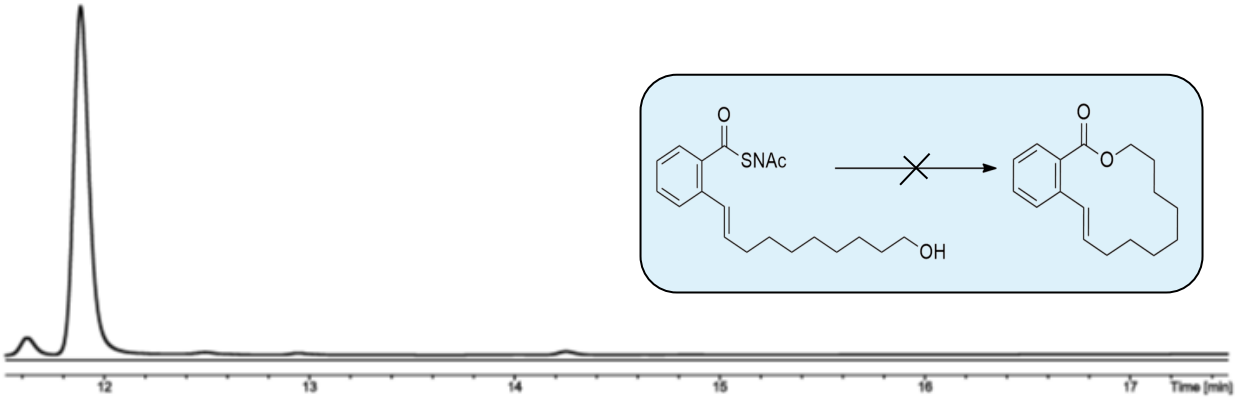


Figure 16: Negative Control Assay HPLC Chromatogram of Substrate Assay
Negative control assay, substituting water for Rdc1TE, with similar components to the *in vitro* assay were separated with HPLC and detected with UV, monitored at 280 nm. Reaction in blue outlines the negative assay, as the product could not be formed as Rdc1TE was lacking.

It is useful to note than an entire PURExpress reaction is not required to produce detectable protein or conduct *in vitro* biochemical experiments as ¼ of a reaction sufficed. Being able to use less of the kit allows the ability to conduct more experiments, 4X more than the manufacturers 10 experiments per kit. Furthermore, the PURExpressed RadTE protein was not purified from the PURExpress kit components but was still catalytically active. This indicates that the additional components in the kit do not interfere with this biochemical reaction and may not inhibit other biosynthetic reactions, thus simplifying the downstream biochemical assays. This is particularly important as PKS proteins have been observed to be unstable and unfold easily, especially during purification.[36]

2.2.2…Cell-Free Production of Fungal Monocillin II Biosynthetic Proteins

Based on the success of using the PURExpress to express and subsequently assay with the PKS catalytic domain RadTE, we undertook the evaluation of expression of full length PKS proteins, Rdc5, Rdc1 and the phosphopantetheinyl transferase required for their post-translational modification, SFP.

Rdc5 and Rdc1 are multi-domain proteins and significantly larger than the RadTE, as shown in ***Table 2***. While the PURExpress kit has been shown to express proteins up to approximately 100 kDa, such as β-galactosidase, multidomain 200+ kDa proteins had not been reported with this system. As the large domain proteins require more amino acids, it is expected that more resources will be required for every protein produced, thus taxing the CFPS. We were thus concerned that the limited, or no, protein production would be detected and significant optimization of the CFPS protocol might be required.

Table 2: Amino Acid Compositions and kDa of Rdc5, Rdc1 and SFP

Protein	# of Amino Acids	kDa
Rdc5	2383	260.19
Rdc1	2090	228.61
SFP	233	26.25

Our initial proof of concept experiments involved using ¼ PURExpression scale reactions to produce Rdc1 from the plasmid pFM51, Rdc5 from the plasmid pFM50 and SFP from the plasmid pMEH22. (***Figure 17***).

Visualization of stain free SDS-PAGE of the resulting PURExpressions reactions showed a haze of signal in the 200-250 kDa range where Rdc5 and Rdc1 are expected. This rendered it impossible to see the bands (**Figure 17***)*. Staining the gel overnight with coomassie brilliant blue, followed by destaining, proved more sensitive than stain-free imaging and enabled identification of discrete,

though faint, bands for Rdc5 and Rdc1 (***Figure 18***). The much smaller proteins, Rdc2 and SFP, were readily visualized in both the stain-free and coomassie stained gels.

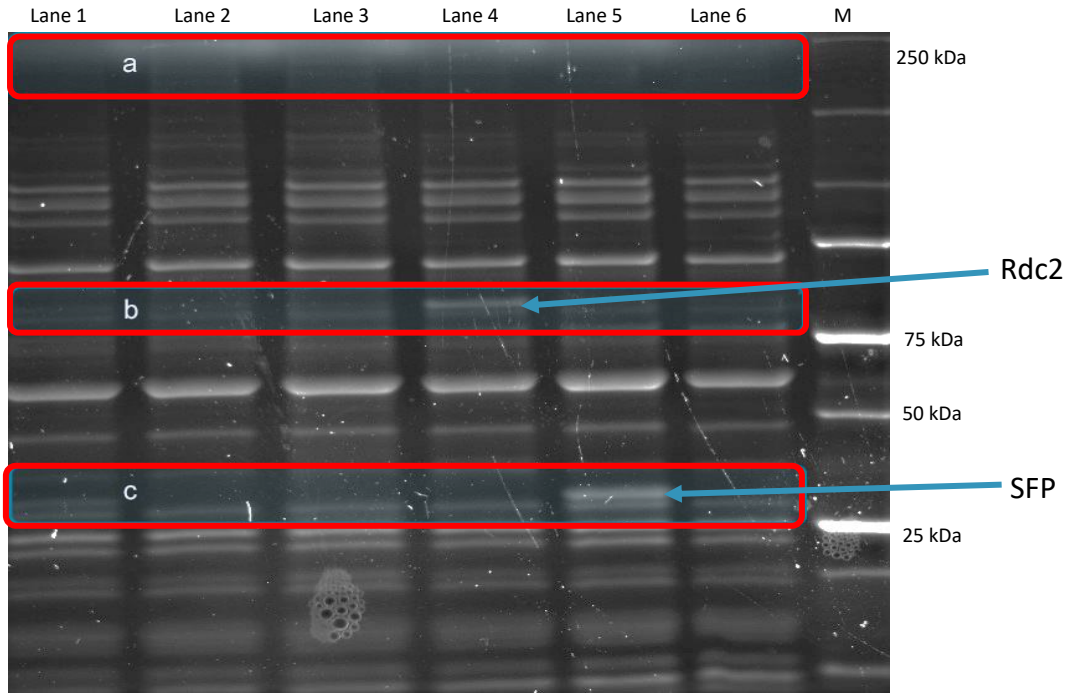


Figure 17: Initial Visualization of PURExpressed Radicicol biosynthetic pathway proteins via SDS-PAGE
Proteins Rdc1, Rdc2, Rdc5 and SFP PURExpressed and subsequently separated via SDS-PAGE to determine if the proteins can be produced with PURExpress CFPS kit **a)** "Fog" not allowing visualization of Rdc5 and Rdc1 **b)** Region showing production of Rdc2 in lane 1 & lane 4 **c)** Region showing production of SFP in lane 1 and 4. **Lane 1**: All 4, **Lane 2**: Rdc5, **Lane 3:** Rdc1, **Lane 4**: Rdc2, **Lane 5**: SFP, **Lane 6**: Negative Control, **M**: Precision Plus Protein™ All Blue Prestained Protein Standard. Along with the lanes labelled, the important kDa markers shown next to ladder bands. Proteins Rdc2 and SFP labelled with arrows

Analysis of the CFPS for Rdc5 and Rdc1 demonstrated that bands consistent with their expected sizes were produced. As none of the PURExpress components are of similar size, and no bands of this size are observed in the no DNA control experiments, these bands are dependent on DNA encoding Rdc5 and Rdc1 and thus likely represent the proteins Rdc5 and Rdc1.

It is significant that the PURExpress kit can produce these extremely large proteins as this work now expands the range of protein sizes for which CFPS can be applied to. While we were concerned with the low levels of PKS proteins produced, it is well documented that the expression of these proteins in native, and heterologous, hosts typically generates only trace protein levels which are at best poorly detected by SDS-PAGE. Thus, these results were not

inconsistent with heterologous host expression levels that led to production of detectable natural products.[8,37] To increase productivity of the CFPS system, we can use more of the kit or optimize the expression conditions, such as PURExpress incubation length and temperature. As proteins work optimally at specific temperatures[38], it's possible that more abundant PURExpressed protein can be synthesized by manipulating the expression temperature.

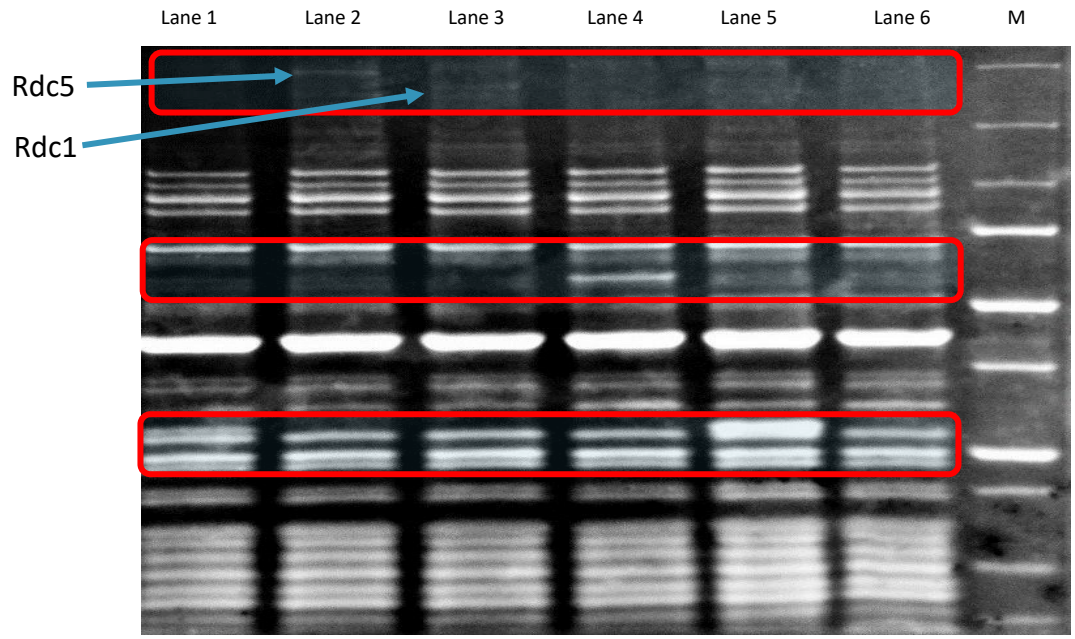


Figure 18: Coomassie stain visualization of PURExpressed Radicicol biosynthetic pathway proteins via SDS-PAGE
Coomassie brilliant blue staining of previous SDS-PAGE gel to remove fog and visualize large PKS and determine if they were produced. **Lane 1**: All 4, **Lane 2**: Rdc5, **Lane 3**: Rdc1, **Lane 4**: Rdc2, **Lane 5**: SFP, **Lane 6**: Negative Control, **M**: Precision Plus Protein™ All Blue Prestained Protein Standard. Now visible Rdc5 and Rdc1 protein bands are indicated with arrows.

The manufacturer recommends PURExpression at 37˚C but many polyketide synthases proteins are expressed in heterologous hosts at reduced temperatures. An incubation temperature study was conducted where PURExpression is performed at room temperature, 30˚C, 37˚C and 40˚C.

The results from the temperature incubation studies **(*Figure 19a*)** show that there is a considerable increase in proteins being produced when an incubation temperature of 30˚C is

used. This can be clearly seen with Rdc1, where the band at 30˚C is more prominent, and defined, as compared to the bands produced at elevated temperatures.

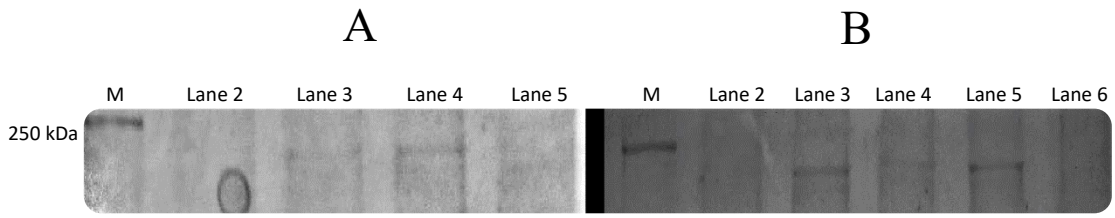


Figure 19: Expression, and *in vitro*, incubation temperature manipulation studies with Rdc1
A) SDS-PAGE separation and coomassie staining of PURExpressed Rdc1 expressed either at 40˚C (**Lane 2**), 37˚C (**Lane 3**), 30˚C (**Lane 4**), and room temperature (**Lane 5**) to determine if expression temperature impacts production **B)** SDS PAGE separation and coomassie staining of PURExpressed Rdc1 expressed at 30˚C for 2 hours (**Lane 2**), 3.55 hours (**Lane 3**), 5 hours (**Lane 4**), 6.5 hours (**Lane 5**), and overnight (**Lane 5**) to determine if length of expression impacts production. Both expression temperature, and length of expression impact production as can be seen.

After identifying a more optimal temperature for expression, a time course was performed to optimize PURExpression incubation times using the manufacturer recommended 2 hour incubation as well as 3.5 hours, 5 hours, 6.5 hours and overnight incubations ***(Figure 19b***).

The 3.5-hour incubation produced a more prominent band compared to the other conditions. The band became less prominent with 5 hours but darkened with 6.5 hours and disappeared with the overnight incubation. Based on this data, we selected 3.5 hours as the optimal incubation time of CFPS of large PKS proteins.

2.2.3…Visualization of proteins synthesis through incorporation of BODIPY-Lysine

One of the inherent advantages of cell-free protein synthesis is the fact that it's an open system unlike cellular hosts. This open nature allows the addition of various supplements to the *in vitro* transcription and translation reactions. For example, tRNAs loaded with of non-canonical amino acids can be added to the PURExpress system enabling incorporation of the non-canonical amino acid into the protein product. In this experiment we take advantage of this property to incorporate a fluorescently labelled Lys into the growing polyketide synthases.

The FluoroTect™ GreenLys *in vitro* Translation Labeling System, available by Promega Corp, is an amino acylated lysine tRNA linked to a BODIPY fluorescent dye. BODIPY fluoresces green at 512 nm when excited with light, 503 nm.[39] As this tRNA has the anticodon UUU, it can compete with the native non-fluorescently labelled tRNA for AAA codons. Thus, once protein expression has occurred, any gene possessing an AAA codon should produce protein with some fluorescently labelled Lys.[40] This experiment therefore gives a direct fluorescent readout of proteins produced by *in vitro* translation. As only the target PURExpressed protein should be produced under these conditions, the only fluorescent bands on an SDS-PAGE should be the target protein. The CFPS reactions of Rdc1, Rdc5, and SFP with the incorporations of BODIPY-Lysine were separated via SDS PAGE and imaged using the Alexa488 blot protocol of the ChemiDoc XRS+ imaging system (***Figure 20***). The bands for Rdc5, Rdc1 and SFP are clearly visible, especially as compared to the coomassie stained counterparts. Based on the gel, the large PKS proteins produce a band consistent with full length protein, as well as a significant number of smaller fragments, likely due to stalling of translation.

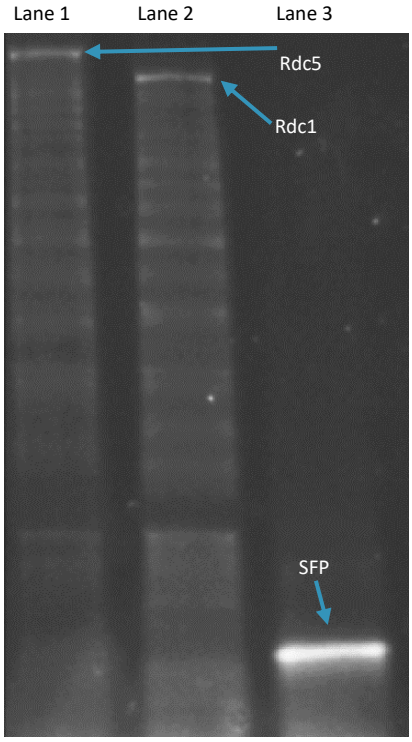


Figure 20: Visualization of BODIPY-Lysine incorporated PURExpressed proteins
PURExpressed protein backbones were integrated with BODIPY-Lysine for visualization post SDS-PAGE using ChemiDoc imaging system. **Lane 1**: Rdc5, **Lane 2**: Rdc1, **Lane 3**: SFP. Allows detection, and confirmation, of PURExpressed proteins

This result confirms that, as expected, protein of 200-250 kDa, consistent with Rdc1 and Rdc5, are produced by the PURExpress reaction.

2.2.4…Phase I Monocillin II Production Assays

With successful detection of bands for PURExpressed Rdc5, Rdc1 and SFP, an initial attempt to assay for monocillin II from the PKS proteins was attempted. Proteins were expressed, using the same conditions as previously described. PURExpressed SFP and Coenzyme A (CoA), were added to Rdc1 and Rdc5 to ensure post-translational modification of the ACP domains (***Figure 9***). After the addition of NADPH and Malonyl CoA, the assays were incubated and evaluated by HPLC analysis for monocillin II production.

Literature reports show successful production of monocillin II *in vitro* from recombinant Rdc1 and Rdc5 using 2 mM NADPH and 2 mM Malonyl-CoA.[31] These were thus the NADPH and Malonyl-CoA concentrations used for our initial tests for cell-free monocillin II production.

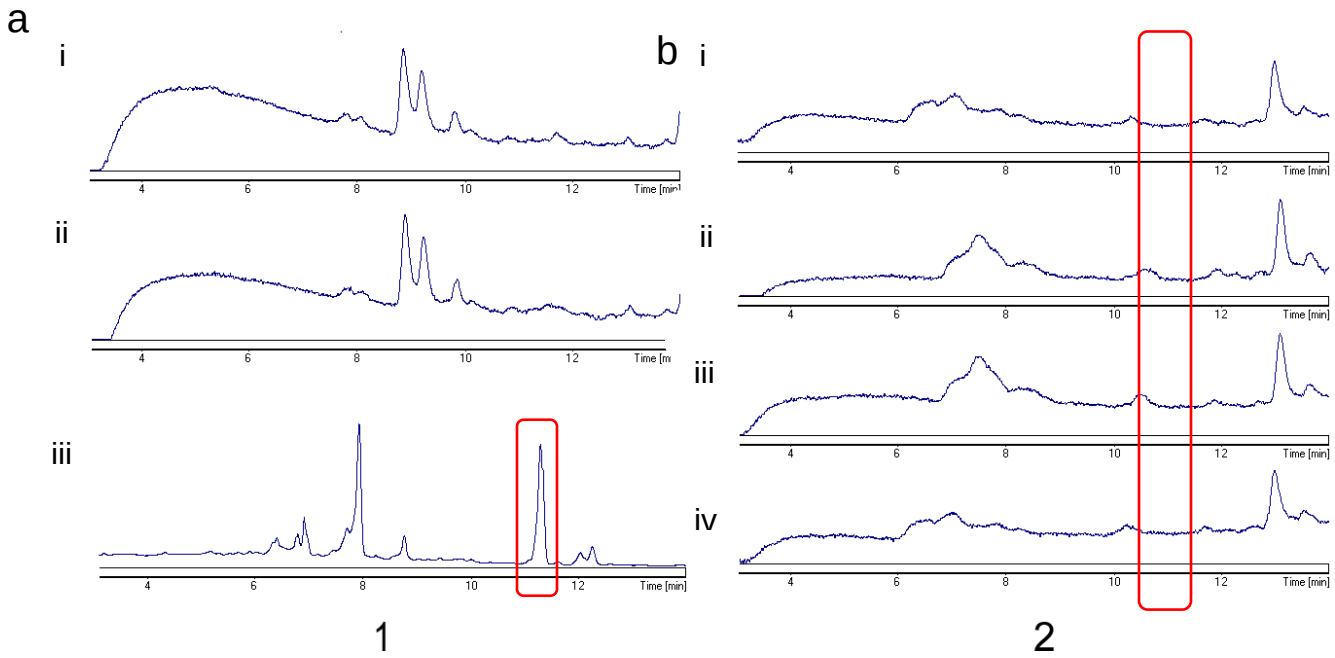


Figure 21: Initial Monocillin II Production Assays HPLC Chromatograms
Monocillin II production assays, separated with HPLC and detected with UV, comprised of either **a)** Enzymatic assays using 2 mM NADPH and Malonyl CoA. **i)** Enzymatic Assay **ii)** Negative control **iii)** Positive control Monocillin II produced in *S. cerevisiae* strain BJ5464-NpgA with Monocillin II peak outlined in red **b)** Elution of Monocillin II production assays using variable substrate amounts and buffers. **i)** No DMSO or phosphate buffer added. 2 mM Malonyl-CoA, 4 mM NADPH added **ii)** Negative control of i. **iii)** DMSO and Phosphate added, double Malonyl-CoA and NADPH concentration **iv)** Negative control of iii. Region outlined in red indicated region where Monocillin II should be detected, if successfully produced, based on previous standard. Lack of peak in red box indicates no detectable production of monocillin II.

As can be seen in ***Figure 21***, none of the experiments conducted showed any discernable difference between the enzymatic assays and their respective negative controls. Additionally, ***Figure 21aiii*** is a chromatogram of a yeast produced authentic standard of monocillin II. The peak highlighted in red is Monocillin II. No detectable monocillin II was produced, however, the sensitivity of the HPLC assay used in the experiment is limited and thus it is possible a more sensitive detection method such as LCMS could detect monocillin II.

2.2.5…Visualization of the modification of the ACPs with BODIPY-CoA and SFP

To rule out lack of post-translational modification of the ACP domains in Rdc1 and Rdc5 as the reason that no monocillin II was produced, we evaluated if the proteins were correctly modified. To detect this modification, a CoA derivative is needed that will enable visualization of the post-translational modification of the ACP. By reacting BODIPY-maleimide with CoA, we produced a BODIPY-CoA substrate that can then be incorporated onto the ACP (***Figure 22***) by the action of SFP. Following incorporation, the polyketide synthase proteins should become fluorescent, confirming the correct post-translational modification. Production of the BODIPY-CoA substrate has been previously reported in the literature.[7] This protocol was used to make the BODIPY-CoA.

Visualization of the post-translational modification, ***Figure 23a***, clearly displayed bands corresponding to Rdc1 and Rdc5 on the SDS-PAGE which are illuminating via fluorescence, consistent with a BODIPY linked phosphopantetheinyl arm added onto the ACP of the multi-domain proteins.

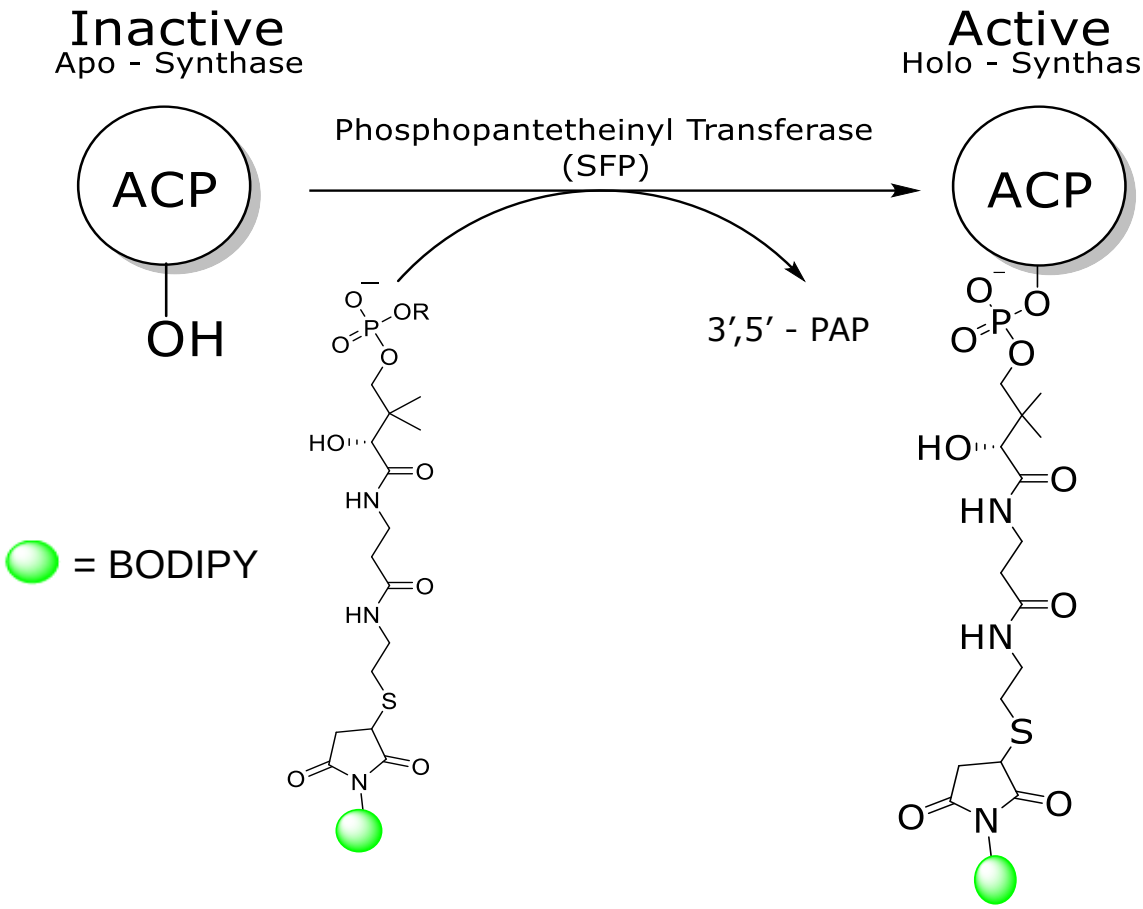


Figure 22: Overview of the modification of ACP
Modification of ACP through inclusion of BODIPY-CoA, which will be integrated onto the ACP to allow fluorescence detection.

Figure 23b confirms that the bands are SFP dependent. For example, in lane 4, in the absence of SFP, no labelling of Rdc1 is observed. Note, labelling is very faint in Rdc5 lane 1. This data is consistent with the ACP domains being successfully modified in an SFP-dependent fashion by a phosphopantetheinyl group. Additionally, these experiments are further confirmation that full length Rdc5 and Rdc1 are produced as the ACP domains are at the C-terminus of these multi-domain proteins. Thus, for SFP to recognize the ACP, the full-length protein must be produced, and the ACP domain must fold correctly.

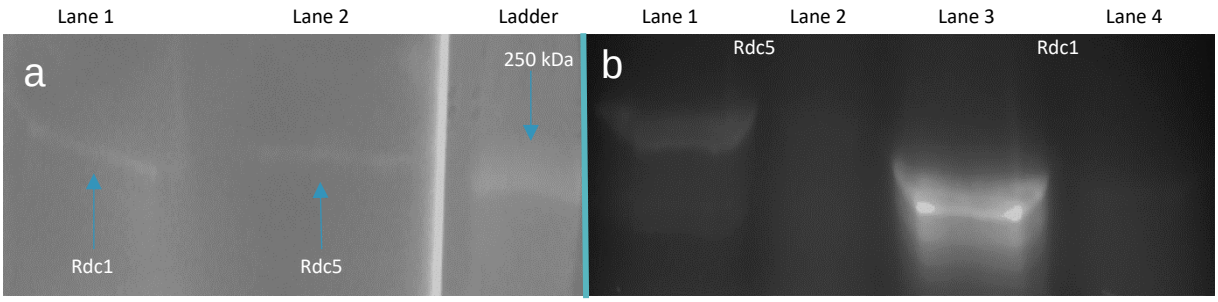


Figure 23: Visualization of Post-Translational modification of Rdc5 and Rdc1 ACPs using SFP and BODIPY-CoA
(a) Initial visualization of ACP using BODIPY-CoA. BODIPY-CoA produced using MES Acetate as the buffer, and *in vitro* post translational modification conducted at room temperature for 30 minutes. **Lane 1:** Post-translationally modified Rdc1 **Lane 2:** Post translationally modified rdc5 **Lane 3:** Precision Plus Protein™ Unstained Protein Standard visualized UV detection post activation. (**b)** Updated visualization of ACP using BODIPY-CoA where buffer was 50 mM Tris-HCl, 50 mM NaCl, 100 mM MgCl. *In vitro* post translational modification occurred at 30°C over 3 hours. **Lane 1:** *In vitro* post-translational modification of Rdc 5 **Lane 2:** Negative control modification assay of Rdc5 **Lane 3:** *In vitro* post-translational modification of Rdc 1 **Lane 4:** Negative control modification assay of Rdc1. Presence of bands within the post-translational modification assay, and lacking in the negative control assay, confirms modification is SFP dependant.

The experiments presented in ***Figure 24b*** show significantly improved labelling for Rdc1 as compared to ***Figure 24a***. This is due to a change in buffer from MES-Acetate to Tris-HCl, NaCl and $MgCl_2$, a change in temperature from room temperature to 30°C, and an incubation time of 3 hours, instead of 30 minutes. These protocols are similar to those previously described in the literature.[41] Both the increase in temperature and incubation times are expected to lead to increased SFP mediated labelling of the PKS ACP domains. In addition, the production of Rdc1 and Rdc5 in ***Figure 24b*** was generated using modified cell-free production. Comparing ***Figure 23a*** *to* ***Figure 23b***, one can clearly see a distinct difference in the presence, and intensity, of the bands. This can be attributed to the changes in the incubation conditions ***(Table 3)***, which significantly increase the levels of proteins produced (described in section 2.2.6) thus helping to better demonstrate labelling of Rdc1.

Table 3: Modern Standard PURExpress conditions

Solution A	4.0
Solution B	3.0
Solution C	2.0
DNA	0.5
RNAse Inhibitor	0.5
Total	10.0 µL

2.2.6…Supplementing PURExpression to increase protein production

Having demonstrated that the lack of monocillin II production from cell-free production of Rdc5 and Rdc1 is not due to a lack of pos-translational modification of the ACP domains, we hypothesized that there could be insufficient PKS proteins present to produce enough monocillin II to detect by our HPLC assay. We thus investigated supplementing the PURExpress reaction to boost PKS protein production.

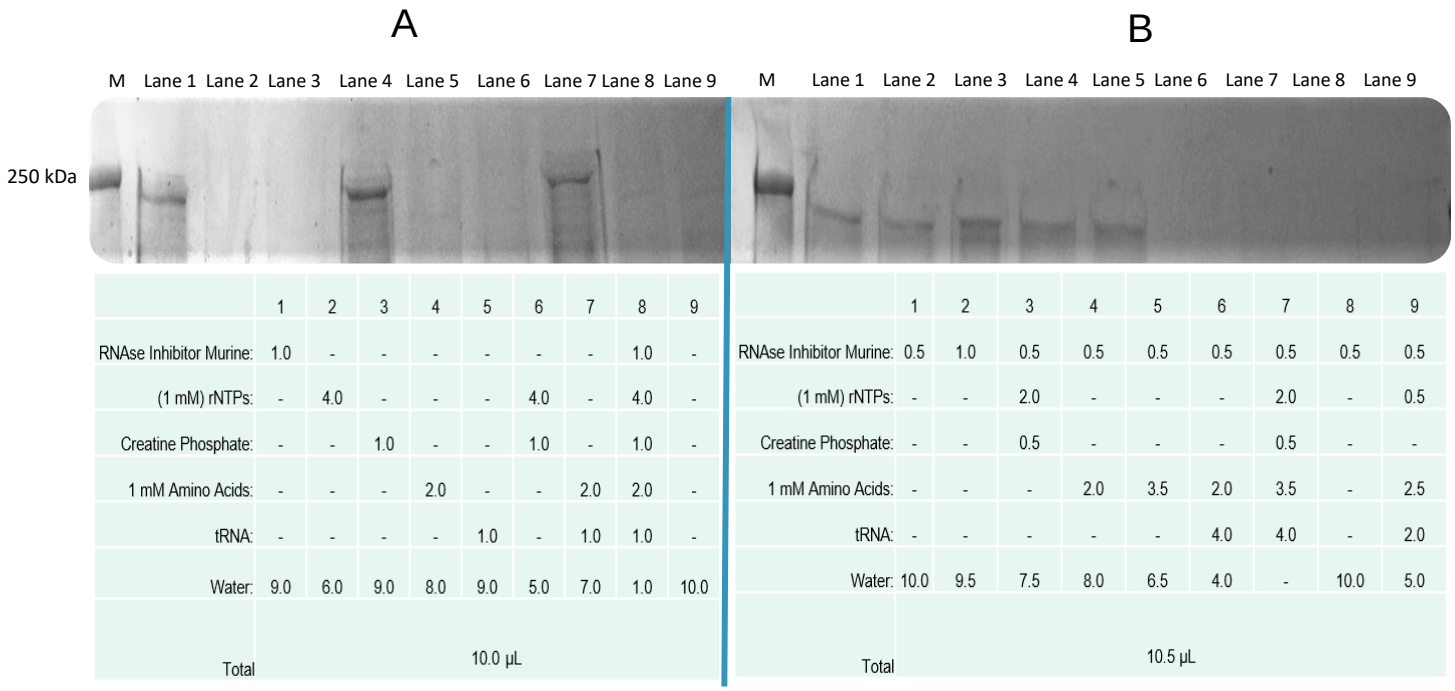
Transcription and translation are cellular processes that produce RNA and proteins respectively.[41] These processes are constant and require the same substrates and energy components to fuel them. By supplying additional components for transcription/translation such as tRNAs or amino acids, these processes should work more optimally and produce more product in a cell-free setting. Within the kit, T7 RNA polymerase is used to transcribe the RNA using NTPs along with ATP as the energy source.[23] ATP is provided as both readily available ATP, and stored ATP as creatine phosphate. Creatine phosphate is used by creatine kinase to convert ADP to ATP, thus regenerating ATP. This regeneration of ATP enables the phosphate levels in the *in vitro* transcription and translation assay to remain fairly constant rather than increasing significantly as the reactions proceed. For translation of proteins, tRNAs along with amino acids are used to build the proteins based on the RNA strand produced.

Table 4: PURExpress conditions for Supplemental Reactions

Solution A	2.0
Solution B	1.5
DNA	1.5
Supplements	X.X
Water	X.X
Total	15.0 µL

Knowing the general contents of the kit, and what is ultimately used to produce either mRNA or polypeptides, this investigation was conducted to see if the various components can be

supplied to see a positive effect on the amount of proteins produced. This would allow for more target proteins to be produced. For this investigation, Rdc1 was used as the target protein. The low level of production of Rdc1 under our standard conditions enables ready detection of improved production by supplementation of individual, or a combination, of components. The reactions follow the protocol described in ***Table 4***.



	1	2	3	4	5	6	7	8	9
RNAse Inhibitor Murine:	1.0	-	-	-	-	-	-	1.0	-
(1 mM) rNTPs:	-	4.0	-	-	-	4.0	-	4.0	-
Creatine Phosphate:	-	-	1.0	-	-	1.0	-	1.0	-
1 mM Amino Acids:	-	-	-	2.0	-	-	2.0	2.0	-
tRNA:	-	-	-	-	1.0	-	1.0	1.0	-
Water:	9.0	6.0	9.0	8.0	9.0	5.0	7.0	1.0	10.0
Total	10.0 µL								

	1	2	3	4	5	6	7	8	9
RNAse Inhibitor Murine:	0.5	1.0	0.5	0.5	0.5	0.5	0.5	0.5	0.5
(1 mM) rNTPs:	-	-	2.0	-	-	-	2.0	-	0.5
Creatine Phosphate:	-	-	0.5	-	-	-	0.5	-	-
1 mM Amino Acids:	-	-	-	2.0	3.5	2.0	3.5	-	2.5
tRNA:	-	-	-	-	-	4.0	4.0	-	2.0
Water:	10.0	9.5	7.5	8.0	6.5	4.0	-	10.0	5.0
Total	10.5 µL								

Figure 24: Coomassie staining of PURExpressed Rdc1 with Supplementary Solutions of Substrates
a) Initial supplemental reactions of PURExpression of Rdc1 using either RNAse Inhibitor Murine, rNTPS, Creatine Phosphate, Amino Acidsor tRNA. Concentrations, and volumes, of each supplement outlined in ***Table 8*** **b)** Supplemental reactions consistently having RNAse Inhibitor and having either rNTPs, Creatine Phosphate, Amino Acids or tRNA. Concentrations, and volumes, of each supplement outlined in ***Table 9***. The PURExpressed proteins were separated by SDS-PAGE and imaged using the Biorad stain free in gel fluorescence system. Shows possibility to increase production of desired protein through supplementing of certain expression components, ex. Amino Acids.

Initial experiments (***Figure 24a***) showed that addition of Murine RNAse Inhibitor increased the production of Rdc1 substantially, compared to the no supplement control lane. As the RNAse inhibitor stabilizes and prevents degradation of mRNA, this result is easily rationalized as an increase in the transcript level. Additionally, increased protein was produced by supplementing with of additional amino acids. This makes sense as more amino acids should equate to more proteins as there is a larger pool of amino acids to be used for each growing polypeptide. This is likely particularly important for long polypeptides as these PKS proteins. The presence of

protein was negatively impacted when NTPs, creatine phosphate, or tRNA were supplemented individually or in combinations. The reasons for this negative impact could be due to a saturation of ATP, which has been shown to inhibit protein production.[41] A prominent band was present when amino acids and tRNA was added, though comparing that band to the band produced with only amino acids shows that the inclusion of tRNA likely decreased the intensity of the band. Inclusion of Murine RNAse Inhibitor was deemed a necessary addition to all further PURExpression experiments.

A second round of experiments included a constant amount of Murine RNAse Inhibitor, except for a single lane which had double the amount (***Figure 24b, lane 2***), supplemented with additional reaction components to investigate the impact that the addition had on protein production. These experiments (***Figure 24b***) showed that the doubling of Murine RNAse Inhibitor didn't affect the intensity of the desired protein product band. The combination of increased NTPs, and creatine phosphate, did not have a substantial effect on protein production. Surprisingly, the supplementation with additional amino acids, while effective without RNAse inhibitor, did not substantially increase the protein levels. The combined supplementation of amino acids and tRNA eliminated production of the desired protein band suggesting the inclusion of excess tRNAs is disadvantageous.

Table 5: Standardized PURExpression Composition moving

Solution A	4.0
Solution B	3.0
Solution C	2.0
DNA	0.5
RNAse Inhibitor	0.5
Total	10.0 µL

Using the knowledge acquired from both these experiments, a standard PURExpress protocol was produced that would be ultimately used

throughout the future experiments unless otherwise stated. This new protocol included the addition of Murine RNAse Inhibitor, as well as amino acids, which are combined and labeled as "Solution C".

2.2.7…Production of an authentic Monocillin II standard in *S. cerevisiae*

An authentic standard of monocillin II was clearly needed to confirm the ability of our analytical methods to detect this natural product. To obtain a standard, monocillin II was heterologously expressed in *S. cerevisiae*.

A two-plasmid system was used to express Rdc1 and Rdc5 in *S. cerevisiae*. Each plasmid was selected for via a different auxotrophic marker.[42] Expression of genes was under the control of the *ADH2* promoter which represses expression in the presence of glucose.[43]

Following multiple repetitions of monocillin II production experiments, followed by, extraction and partial purification, an adequate sample was obtained. Comparing HPLC traces of newly isolated monocillin II with an old authentic standard, FM-15, showed that the chromatograms were very similar (***Figure 25)***. Attempts were made to further purify the samples; however, they were unsuccessful.

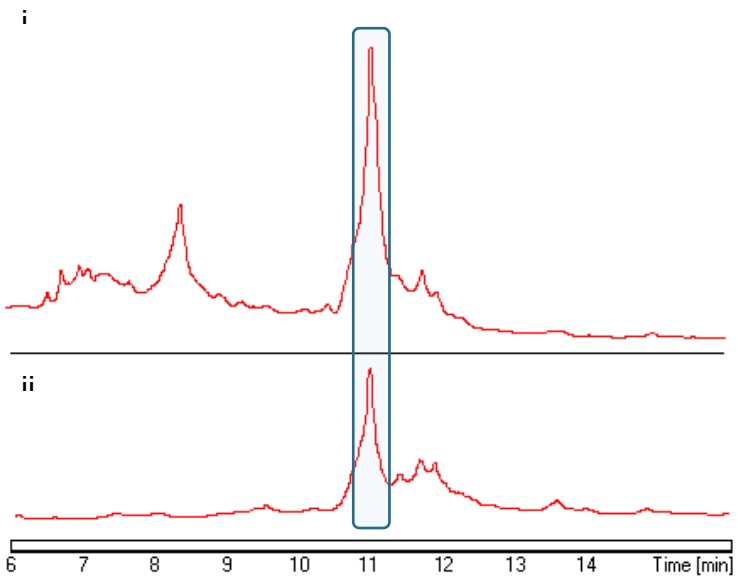


Figure 25: HPLC Chromatograms of Monocillin II standards
Authentic standard samples separated using HPLC and detected at 300 nm with UV to then be used as a positive control in terms of detection of monocillin II. **(i)** Yeast extract of *S. cerevisiae* strain BJ5464-NpgA harboring pKJ91 and pZH223 following 3 days of expression in YPD, isolated with EtOAc, dried and re-dissolved in MeOH **(ii)** Chromatogram of previously confirmed Monocillin II produced using a similar method. Alignment of the peaks indicates region of interest when detecting presence of monocillin II via HPLC-UV.

2.2.8… Phase II Monocillin II Production Assays

To improve the sensitivity of our monocillin II detection, we developed an LCMS assay to test for monocillin II production. As LC-MS is far more sensitive compared to HPLC with UV detection. Typically, sensitivity is increased by orders of magnitude moving from UV detection to MS detection. Thus, it may be possible that some previously conducted reactions produced monocillin II below the threshold for detection by the UV detector but above the detection threshold for the MS detector.

Semi-purified extract containing monocillin II was subjected to LC-MS analysis in scan mode using a positive electron spray ionization (ESI) and negative ESI. The positive scan clearly showed an *m/z* of 317 ($M+H^+$), and 339 ($M+Na^+$), consistent with monocillin II at ~16.4 minutes of retention time. The negative scan produced an m/z signal at 315 (M-H) also consistent with monocillin ii ~16.4 minutes. This data strongly confirms the authenticity of the monocillin II

standard. While both modes readily detected monocillin II, the positive mode provided a signal of greater intensity, whereas the negative mode was significantly more selective (***Figure 26***). As such, both modes were used for further analysis.

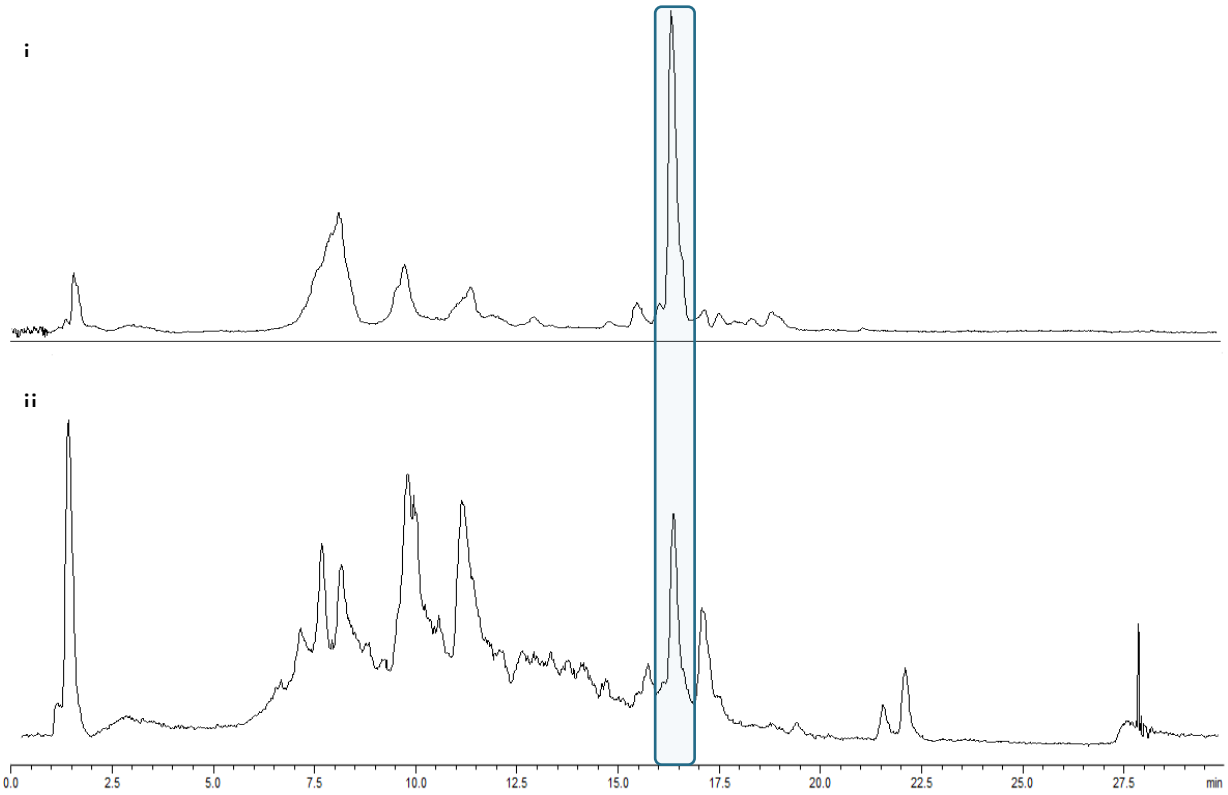


Figure 26: LC-MS Chromatograms of Yeast Extracts containing Monocillin II
Authentic standard samples separated using HPLC and detected at 300 nm with UV to then be used as a positive control in terms of detection of monocillin II.**(i)** Negative scan of yeast extracts, scan showing elution's of m/z 315, M-H **(ii)** Positive Scan of yeast extracts, scans showing elution's of m/z 317, M+H. Alignment of the peaks indicates region of interest when detecting presence of monocillin II via LC-MS

With the authentic standard analysis parameters identified, *in vitro* enzymatic assays to produce monocillin II were evaluated. Standard *in vitro* assay transcription, and translation, conditions for production of post-translationally modified PURExpressed Rdc1 and Rdc5, supplemented with 2 mM NADPH, 2mM Malonyl-CoA were performed. LC-MS analysis of the assays, **Figure 27**, showed no discernable difference between the enzymatic assay and the negative control under the more sensitive ESI positive conditions.

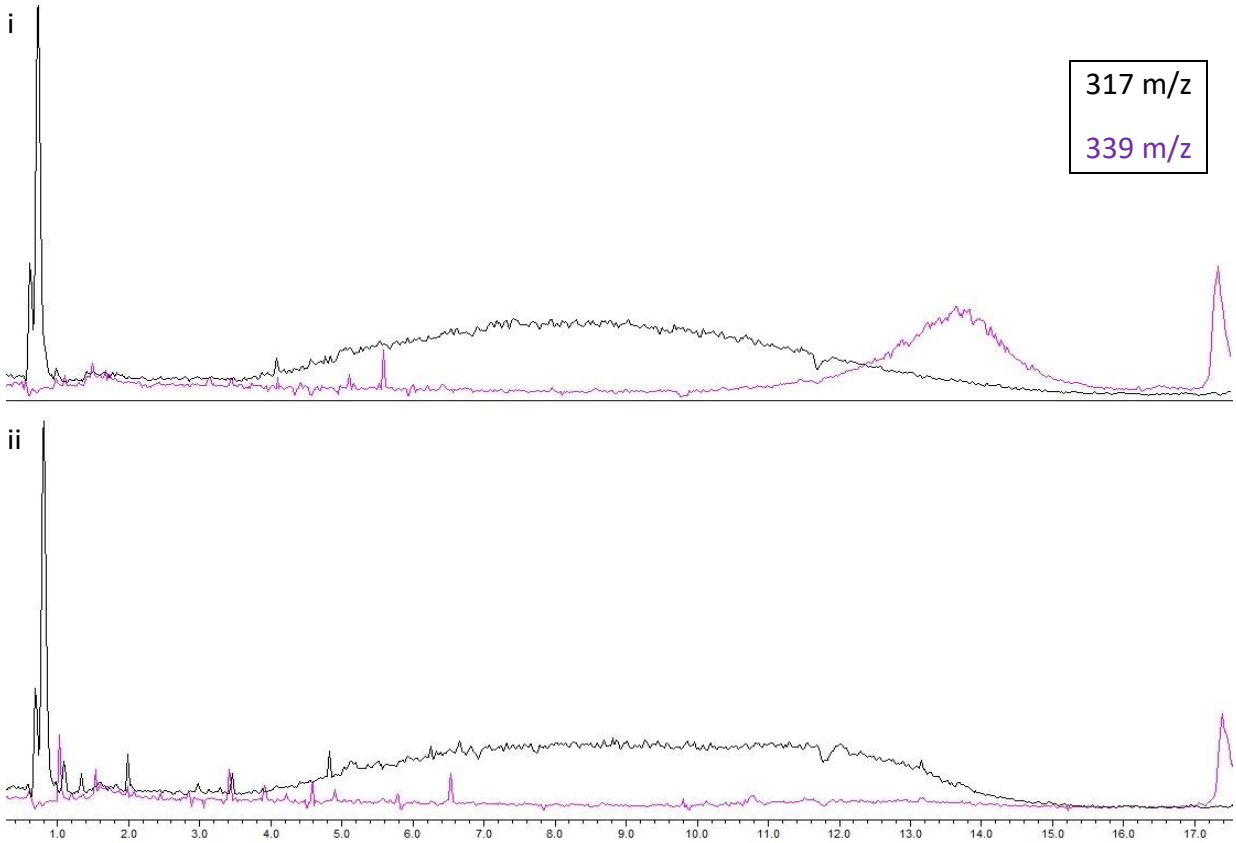


Figure 27: LC -MS Positive Scan Chromatogram of Initial Monocillin II *In vitro* production assay
In vitro biochemical assays separated with LC, detected using MS, scanning for m/z of 317 (Black;M+H) 339 (Purple;M+Na), to detect presence of Monocillin II **(i)** *In vitro* enzymatic assay **(ii)** Negative Control Assay. Lack of distinct peak indicates no detectable monocillin II.

Additionally, modifying the reaction buffer for the assay failed to improve the assay (***Figure 28***). Switching between MES acetate buffer at pH 6.0 to phosphate buffer at pH 7.4 also did not lead to production of monocillin II.

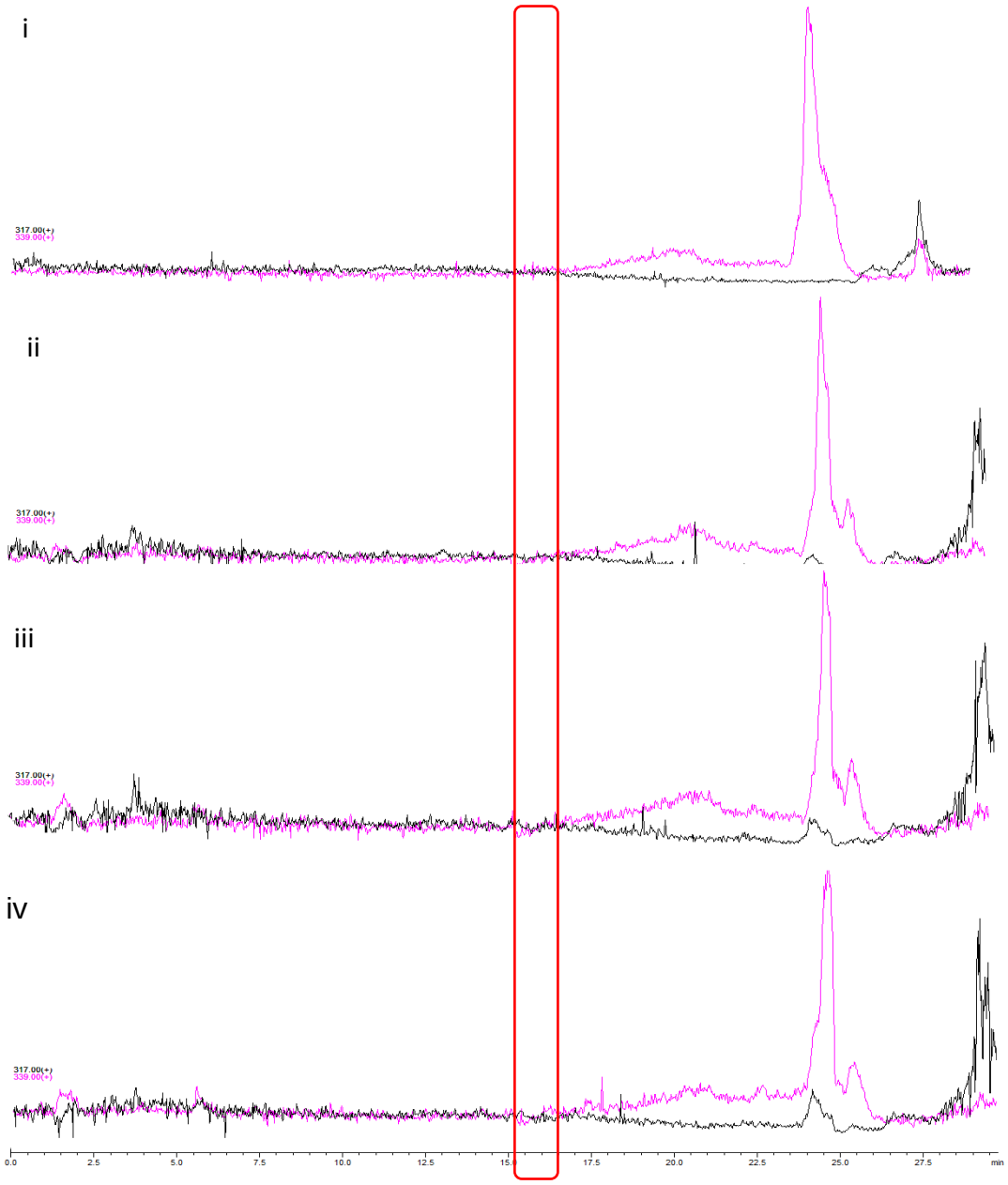


Figure 28: LC-MS Positive Scan Chromatograms for *In vitro* Experiments with different supplementary buffers
In vitro biochemical assays separated with LC, detected using MS, scanning for m/z of 317 (M+H) 339 (M+Na), to detect presence of Monocillin II **(i)** *In vitro* enzymatic assay with additional MES Acetate pH 6.0 **(ii)** *In vitro* enzymatic assay with additional Tris-HCl, NaCl pH 8.0 **(iii)** *In vitro* enzymatic assay with additional phosphate buffer pH 7.4 **(iv)** Negative control assay with additional Tris-HCl, NaCl pH 8.0. Lack of peaks within region of interest, red box, indicates no levels of monocillin II were detected.

2.2.9…Combination experiments using purified and Cell-Free multi-domain proteins

As none of the *in vitro* enzymatic assay's yielded monocillin II, we hypothesized that the one, or both, of the PURExpressed proteins may be misfolded, or inactive.

Since monocillin II has been produced in *S. cerevisiae* through heterologous expression, both Rdc5 and Rdc1 are known to express in functional form in yeast. We thus proposed to isolate each of these proteins from a yeast expression system and complement them with the PURExpressed post-translationally modified Rdc5 and Rdc1. This should enable us to identify if either PURExpressed protein was active. The combinations would be as follows:

- *S. cerevisiae* Rdc5 with *S. cerevisiae* Rdc1
- *S. cerevisiae* Rdc5 with PURExpressed Rdc1
- *S. cerevisiae* Rdc1 with PURExpressed Rdc5
- PURExpressed Rdc1 with PURExpressed Rdc5

The positive control *in vitro* enzymatic assay between the two *S. cerevisiae* expressed and purified proteins should provide authentic monocillin II.

The expression in *S. cerevisiae* of Rdc5 and Rdc1 would be performed under conditions similarly to the production of the authentic standard, however both proteins would be tagged with a hexa-His sequence to enable purification via immobilized metal affinity chromatography.

S. cerevisiae strain BJ5464-NpgA was thus transformed separately, and in combination, with Rdc5 (pKJ91) and/or Rdc1 (pKJ61) and incubated under expression conditions. The recombinant proteins were then isolated and purified using a similar protocol to protein isolation and purification in *E. coli.*

Combinations of Rdc5 and Rdc1 proteins from either *S. cerevisiae* or cell-free production from *in vitro* assays were conducted with NADPH and Malonyl-CoA under our standard conditions and assayed for monocillin II production by LCMS.

As can be seen in ***Figure 29***, none of the LCMS traces showed production of monocillin II.

Thus, it can be inferred none of these combinations of proteins, PURExpressed and/or purified, lead to Monocillin II production.

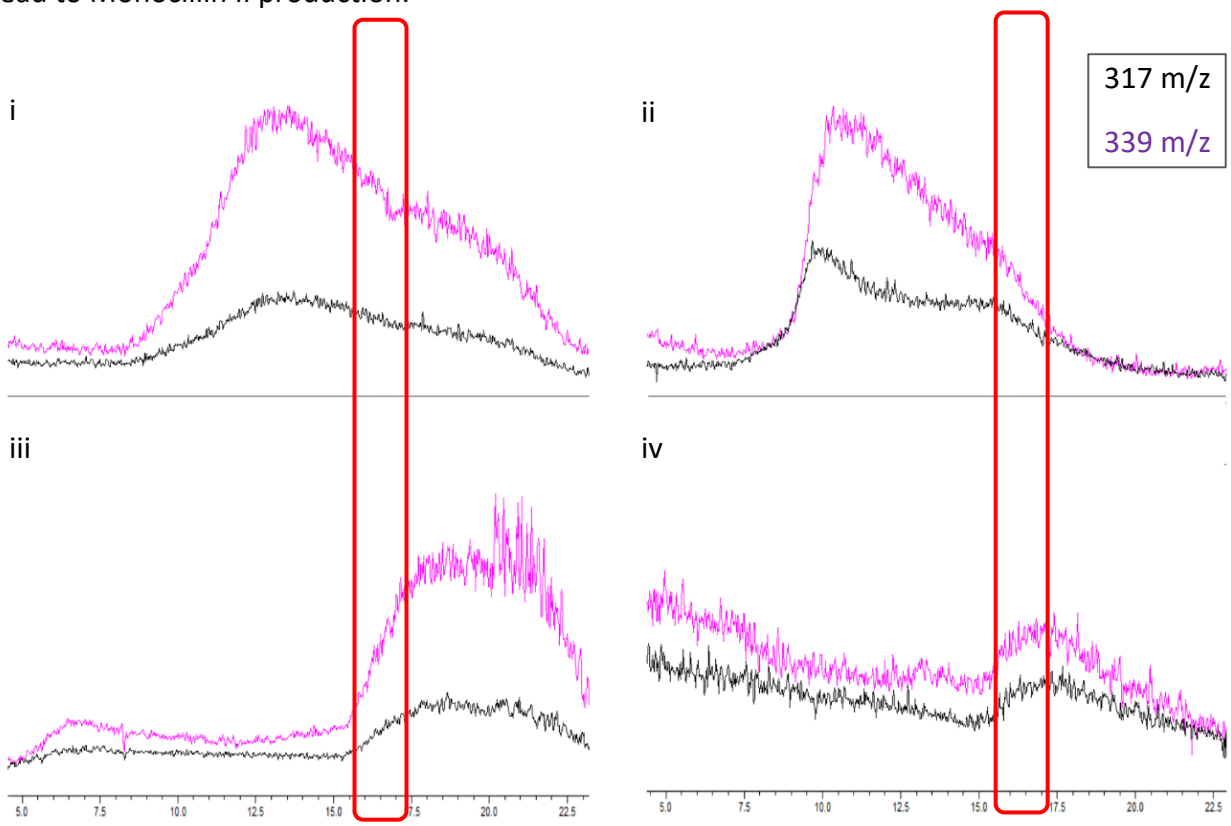


Figure 29: LC-MS Positive Scan Chromatogram for combination assays with PURExpressed Rdc5 & Rdc1 with heterologously expressed, and purified Rdc5 and Rdc1
In vitro biochemical assays separated with LC, detected using MS, scanning for m/z of 317 (Black;M+H) 339 (Purple;M+Na), to detect presence of Monocillin II**. (i)** PURExpressed Rdc5 with PURExpressed Rdc1 **(ii)** PURExpressed Rdc1 with Purified Rdc5 **(iii)** Purified Rdc5 and Purified Rdc1 **(iv)** PURExpressed Rdc5 with Purified Rdc1. Lack of peaks within region of interest, red box, indicates no levels of monocillin II were detected.

It's been noted that PKS proteins are inherently unstable[36] as they contain multiple domains, each needing to fold independently of one another. Significant loss of activity has been observed for bacterial PKS proteins upon purification. We thus hypothesized that both Rdc1

and Rdc5 are likely inactivated during purification as is seen for PKS proteins from the pikromycin biosynthetic pathway.[44]

An additional concern was the presence of his tags on the proteins. While tags, such as a HIS tag can facilitate purification of the protein of interest[41], they can in some cases destabilize the fold and inhibit protein activity.[45] For Rdc5 and Rdc1, the tags are located at the N-terminus of the proteins. It is thus possible that the tags may also disrupt chain transfer from Rdc5 to Rdc1.[35] Therefore, removal of the tags was explored to test if this improved the activity of the system.

2.2.10…Production of TAGless Rdc5 and Rdc1 expression plasmids for Phase I Production assays of Monocillin II using TAGless Rdc5 and Rdc1

N and/or C terminal tags are powerful tools for purification of heterologously expressed proteins. This tag allows one employ highly selective, and specific, purification strategies such as immobilized metal affinity chromatography for His tagged proteins.[41] As we hypothesize that the N-terminal His tag on Rdc1 and Rdc5 are inhibiting activity, we propose to remove these tags.

The coding region for Rdc1 and Rdc5 were digested from their parent vectors pFM50, with NheI and NotI, and pFM51, using NdeI and NotI, and cloned into linearized pET-21b, making new pCAS01 (Rdc5) and pCAS02 (Rdc1) . PURExpression was conducted on these tagless constructs following their production. As seen in ***Figure 30,*** the TAGless constructs produced full length protein as seen in the case of the tagged constructs.

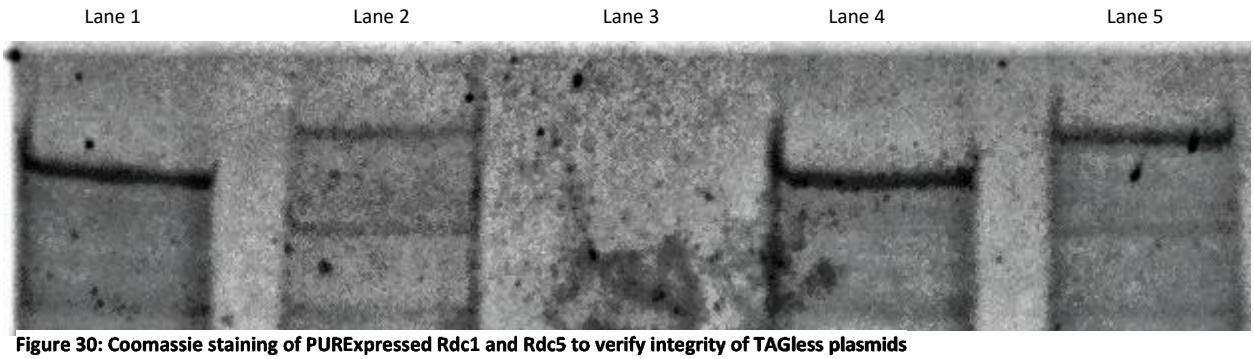


Figure 30: Coomassie staining of PURExpressed Rdc1 and Rdc5 to verify integrity of TAGless plasmids
PURExpressed proteins separated with SDS-PAGE and imaged by staining the Biorad protein gels with coomassie brilliant blue. Presence of bands in respective lanes for TAGless Rdc5 and Rdc1 compared to the positive controls using Tagged Rdc1 and Rdc5 shows protein expression of plasmid intact. Lane 1: Rdc1 (pFM51) , Lane 2: Rdc5 (pFM50) , Lane 3: Empty, Lane 4: Rdc1 (pCAS01), Lane 5: Rdc5 (pCAS02)

Unfortunately, the *in vitro* assays containing tagless Rdc1 and Rdc5 with SFP, malonyl-CoA and NADPH did not produce any detectable monocillin II by LC-MS. ***Figure 31*** shows the results of those experiments, with both the enzymatic assay and the negative control.

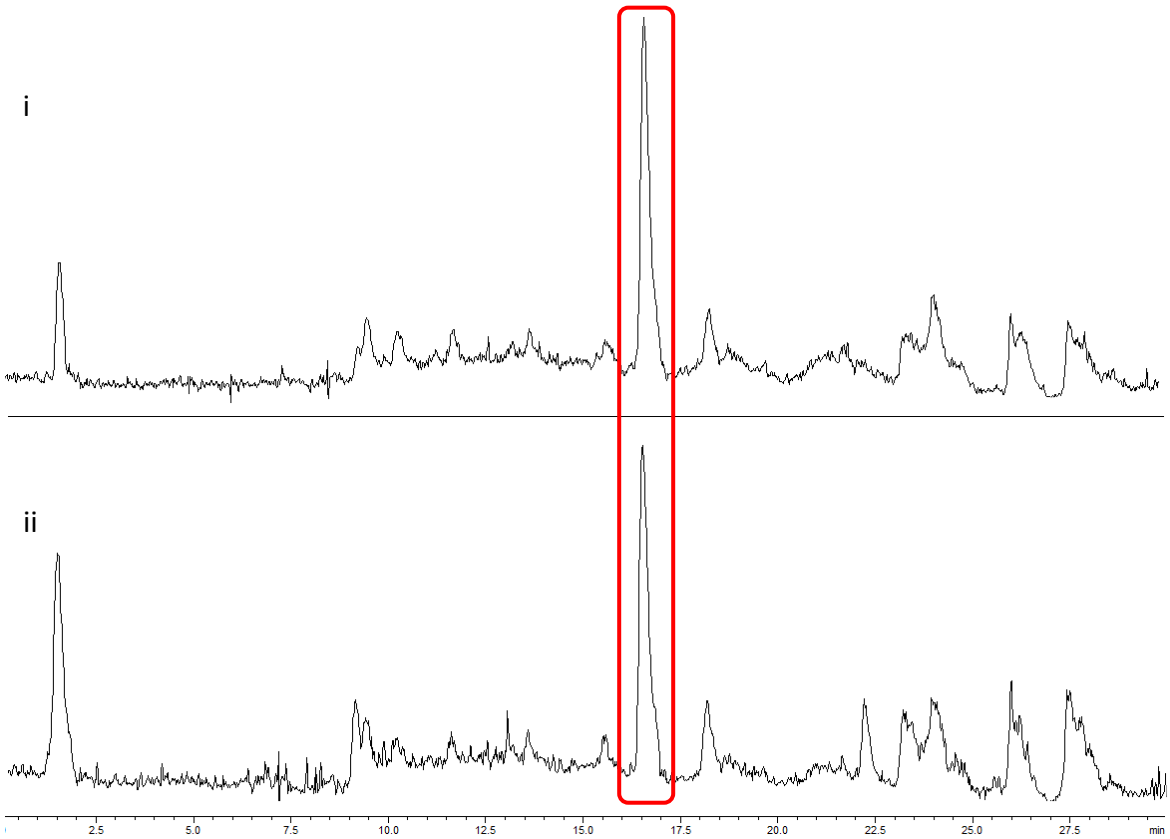


Figure 31: LC-MS Negative Scan Chromatogram of *in vitro* assays using TAGless Rdc5 and TAGless Rdc1
In vitro biochemical assays, **(i)** Enzymatic Assay **(ii)** Negative Control Assay, separated with LC, detected using MS, scanning for m/z of 315 (M-H), to detect presence of Monocillin II**.** Presence of peak within region of interest, red box, in both chromatograms indicates no levels of monocillin II were detected.

2.2.11… Production of Monocillin II within *E. coli* strain BAP1

Monocillin II has either been synthesized synthetically or using *S. cerevisiae* heterologously expressed Rdc5 and Rdc1, especially using strain BJ5464-NpgA. There has yet to be publications of Monocillin II being produced in *E. coli* and these experiments were to determine exactly why. It could be the underlying issue to explain why all the changes within the possible variables do not positively impact the production of Monocillin II.

For this to be possible, BAP1 would need to harbor the plasmids for Rdc1 and Rdc5, both of which have a kanamycin resistance gene which would make it hard to determine if both plasmids are correctly transformed within the host. One of the genes would need to be transferred into another expression vector with a different resistance marker, in this instance ampicillin. This would allow the selection for the double transformants on LB Kan Amp plates. Afterwards, the colonies would be grown up and IPTG would be added to the cultures to induce expression of the proteins, and ultimately generate monocillin II. After the whole process, a fraction of the pellet would be used to detect RNA production of the proteins involved, another fraction of the remaining pellet would be used for protein isolation/ purification through normal conditions, and a final fraction of the pellet would be subjected to denatured extraction of proteins where urea would be added to not only lyse the cells but to denature the proteins.

In a perfect scenario of heterologous expression of the biosynthetic Monocillin II proteins, RNA should be transcribed for both Rdc1 and Rdc5 signifying successful transcription, proteins should be detected via both SDS-PAGE and anti-his tag antibody western blotting within the soluble fraction, signifying successful translation of folded protein, and Monocillin II should be present within the broth, signifying the proteins are active and functional.

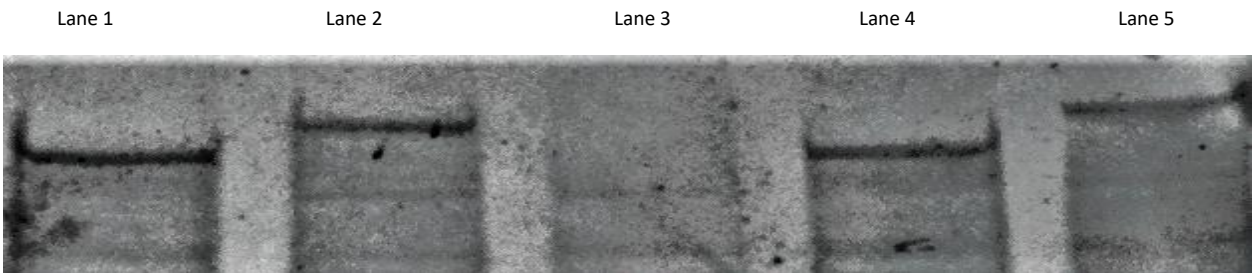


Figure 32: Coomassie staining of PURExpressed Rdc1 and Rdc5 to verify integrity of plasmid switching
PURExpressed proteins separated with SDS-PAGE and imaged by staining the biorad protein gels with coomassie brilliant blue. Presence of bands in respective lanes for TAGless Rdc5 and Rdc1 compared to the positive controls using Tagged Rdc1 and Rdc5 shows protein expression of plasmid intact. Lane 1: TAGless Rdc1 (pCAS01) , Lane 2: TAGLess Rdc5 (pCAS02) , Lane 3: Negative Control (pET28) , Lane 4: TAGless Rdc1 in pET28 (pCAS03), Lane 5: HIS Tagged Rdc5 in pET21 (pCAS04)

The results from these experiments should aid in the understanding of why, although protein expression is successful when using PURExpress, Monocillin II isn't produced in any of the *in vitro* enzymatic assays conducted.

Construction of expression vectors was easily accomplished, and the expression was deemed successful through visualization of PURExpressed bands. The TAGless Rdc5 and Rdc1 vectors produced clear bands on the protein gel (***Figure 32***), whereas no band was detected for the negative control.

Following heterologous expression of HIS tagged Rdc1 and HIS tagged Rdc5 in BAP1, experiments were conducted to detect if mRNA encoding the proteins are being produced in *E.coli*. If mRNA can be detected then, following the central dogma, the respective proteins should be made. The results (***Figure 33***) showed that mRNA is being generated for both Rdc1 and Rdc5, following PCR amplification of the reversely transcribed mRNA from the BAP1 pellets. As RNA is being successfully generated, then proteins should be generated as well.

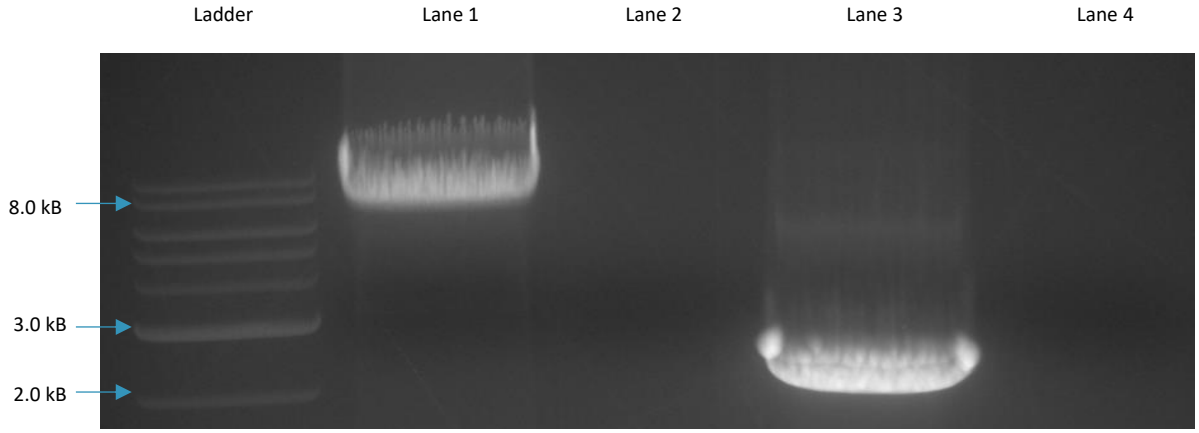


Figure 33: Agarose Gel Electrophoresis separation of PCR amplified reversely transcribed extracted RNA from BAP1 expressing Rdc5 and Rdc1
Separation of reversely transcribed amplicons conducted using 1% agarose gel electrophoresis. Lane 1: Band produced through Rdc5 scout amplification of reversely transcribed RNA Lane 2: Negative control assay of Rdc5 scout lacking reverse transcriptase Lane 3: Band produced through Rdc5 scout amplification of reversely transcribed RNA Lane 4: Negative control assay of Rdc1 scout lacking reverse transcriptase. Presence of bands within the assays, while lacking in the negative wells, indicate successful detection of mRNA corresponding to Rdc1 and Rdc5.

Normally, if a cytosolic protein is correctly folded it would be detected within the soluble fraction. When analyzing the western blot for Rdc5 and Rdc1 expression (***Figure 34***), no bands were detected within the soluble fraction while the insoluble fraction showed a set of bands at the appropriate molecular weight as well as lower molecular weight streaking.

This result shows that the proteins are being heterologously expressed in *E.coli*, as they are the only proteins that should be detected by the anti-HIS antibody. The result is interesting as for PKS proteins to be stable and active, they must be within the soluble fraction. This result thus suggests that they unfortunately are not folding properly within *E. coli* BAP1. It could be possible that the cellular systems within *E. coli* is placing stress on the expression of the protein, leading to misfolding and aggregation of protein rendering it insoluble due to formation of inclusion bodies.[46] Due to this, we predict that no monocillin II should be produced by BAP1.

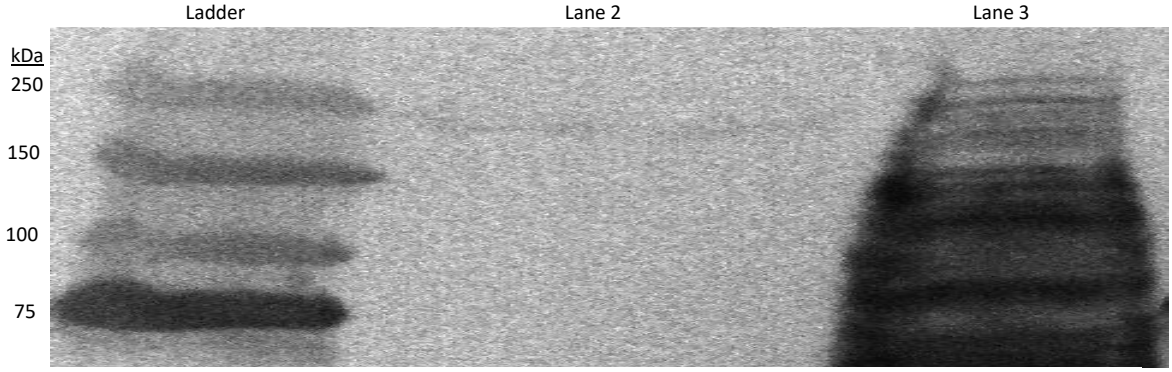


Figure 34: Western detection of HIS tagged proteins within soluble and insoluble fractions of BAP1 harboring pFM50 and pCAS04
Western was conducted post separation with SDS PAGE, which was transferred to a nitrocellulose membrane. Blocking was conducted using 5% Milk TBST, tagged with HRP linked Antibody. Visualization done using Millipore Immobilon Western Chemiluminescence. **Lane 1**: Precision Plus Protein™ All Blue Prestained Protein Standard **Lane 2**: Soluble BAP1 fraction, proteins isolated using normal lysis **Lane 3**: Insoluble BAP1 fraction, proteins isolated using 8M urea. Presence of bands within insoluble fraction, but lacking in the soluble fraction, indicates production issues for these proteins within *E.coli.*

The results from the mRNA and protein detection assays show that the proteins are definitely being expressed but are not folding properly. This result can explain why neither monocillin II ,nor radicicol, have been previously heterologously produced *in vivo* in *E. coli.*

As the PURExpress® *In vitro* Protein Synthesis Kit contains defined concentrations of T7 polymerase, as well as ribosomes, that are used within *E. coli*, it's possible to infer that a reason none of the *in vitro* enzymatic assays between Rdc5 and Rdc1 have been successful is due to a similar reason as to why proteins aren't being properly expressed in BAP1.

2.3…Discussion and Final Thoughts

The goal of this thesis was to determine if *in vitro* transcription and translation could be harnessed to effect cell free heterologous expression of complex polyketide biosynthetic pathways. While much was done to troubleshoot the heterologous production of Monocillin II using the PURExpress® *In vitro* Protein Synthesis Kit, full length biosynthetic proteins could be generated but based on our inability to detect the expected natural product, these proteins are likely misfolded and non-functional.

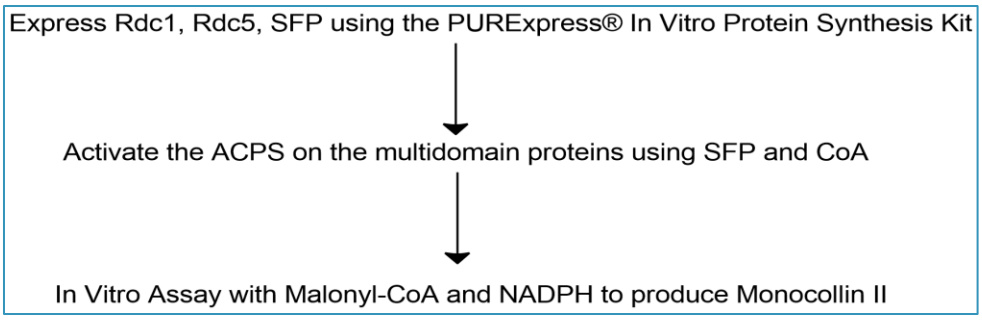


Figure 35: Brief Overview on the experiment

Our results show that low levels of PKS proteins are produced by the cell-free *in vitro* transcription, and translation, system. However, this is not unlike the expression levels seen in native *in vivo* systems, thus we hypothesize that sufficient protein levels are being produced.

PKS proteins contain ACP domains which need to be post-translationally modified. We were able to clearly show that both Rdc5 and Rdc1 produced via the cell-free *in vitro* transcription/translation system could be phosphopantetheinylated by the enzyme SFP in the presence of CoA.

No biochemical activity however could be reconstituted from Rdc5 and Rdc1 under any of the assay conditions investigated. *In vivo*, in *E. coli*, we showed that Rdc5 and Rdc1 were not

expressed as soluble proteins but rather in inclusion bodies. If the cell-free *in vitro* transcription/translation system, which also relies on the *E. coli* ribosome, functions similar to the *E. coli* ribosome *in vivo*, then misfolded protein is likely generated in the cell-free system as well. This would be consistent with our observed lack of biochemical activity.

Protein misfolding, particularly in, the multi-domain proteins is inherently different in different bacterial species. A clear example of this ability of *S. cerevisiae* but not *E. coli* to heterologously produce Monocillin II *in vivo*. A potentially important difference is the translation rates and its codon dependency. Both *E. coli* and *S. cerevisiae* are biased towards specific codons for each amino acid incorporated within the protein. A codon that could favour one organism can be unfavoured by the other one causing the insertion of the amino acyl-tRNA to be faster within the favoured and slower in the unfavoured. It could be entirely possible that the average translation rate, which is the rate of addition of each individual tRNA averaged together, is significantly different within *S. cerevisiae* compared to *E. coli* which could favour a faster translation average rate. It could be faster on a multitude of levels rendering it difficult for each domain to fold properly before the next one is spit out. This could render longer chains of the protein to be translated causing it to not fold as an individual domain, but as a hybrid of the domains.[47]

Additionally, it has been shown that there is a difference in folding energies between *E. coli* and *S. cerevisiae*, which has shown that not only does it affect translation initiation but elongation speeds.[47] This could also play a huge role in the inherent differences between the two cellular organisms which renders some multi-domain PKS proteins to be successfully expressed, and others not. This is similar to the example presented within the introduction of the heterologous

expression of CE, although significant amounts of the CE proteins were detected in *E. coli*, low levels of enzyme activity were observed.

Solutions to these issues could be to conduct a similar study using a defined kit comprised of cellular tools from *S. cerevisiae*. If it is possible to express the proteins and the *in vitro* experiments are successful, using the same conditions and concentrations as used in this study, then it could further point towards these inherent differences of translation rates, and folding energies, being an inherent problem when translating these multi-domain PKS proteins of fungal origins.

Although no Monocillin II was produced through *in vitro* assays, a lot of important new data was obtained from the study regarding the applications of PURExpress® *In vitro* Protein Synthesis Kit and multi-domain PKS proteins. We showed we could reliably miniaturize the cell-free assay and produce enough protein to perform *in vitro* enzymatic assays. Furthermore, we found supplementing the cell-free system with RNAse Inhibitor Murine and additional amino acids consistently improved protein production.

We also took advantage of the "open" nature of CFPS. The open system allowed integration of non-canonical amino acids into the protein backbone by addition of a synthetic amino acyl-tRNA. We used a fluorescently labeled Lys to aid in visualizing newly synthesized protein, however, it is likely that any modified amino acyl-tRNA could be used.

To conclude, although experiments to produce Monocillin II, a fungal natural product, using cell-free *in vitro* protein synthesis, of bacterial origin, have not produced Monocillin, CFPS is still a very versatile heterologous expression tool that, with some tuning, can be significantly more advantageous and useful.

2.4…Methods

Expression of Rdc1 Thioesterase as CFPS proof of concept

2.4.1.A……Cell-Free *In vitro* Expression of Rdc1 TE with PURExpress® *In vitro* Protein Synthesis Kit to visualize protein production

Cell-free *in vitro* protein synthesis was conducted using the contents of the PURExpress® *In vitro* Protein Synthesis Kit. The pMW28 plasmid, ~100 mM, was used to produce Rdc1TE while no plasmid was to be used for the negative control. The reactions were produced using the protocol listed in ***Table 6***.

Table 6: Reaction protocol for CFPS production of proteins.

	Rdc1TE 1 (µL)	Rdc1TE 2 (µL)	Negative Control (µL)
Solution A	5.00	2.500	2.500
Solution B	3.75	1.875	1.875
DNA	1.50	1.500	0.000
Water	2.25	0.375	1.625
Total	12.5	6.25	6.25

2.4.1.B……SDS PAGE separation of PURExpressed Rdc1 TE protein compared to Negative Control

Aliquot 6.25 µL from Rdc1TE 1 and freeze at -20˚C, while adding 6.25 µL of 1X Protein Loading buffer to the rest of Rdc1TE 1 and to Rdc1TE 2 and the negative control. The proteins are then boiled at 95 ˚C for 5 minutes to denature the proteins, all of which are then loaded into a 4-20% Mini-PROTEAN® TGX Stain-Free™ Gel, 12-well, 20 µl from Bio-Rad and separated via SDS PAGE at constant 175 volts, current of 500, for 45 minutes.

2.4.1.C……Cell-Free *In vitro* Protein Synthesis of Rdc1 TE with the PURExpress® *In vitro* Protein Synthesis Kit using different incubation lengths

Cell-free *in vitro* protein synthesis was conducted using the contents of the PURExpress® *In vitro* Protein Synthesis Kit. Three reactions of Rdc1 TE, pMW28 at ~100 mM, were produced, each either being incubated for 2 hours, 3.5 hours or 7.5 hours while no plasmid was to be used for the negative control. All reactions were incubated at 37˚C. The amount of solution a,

solution b, DNA and water are similar to the amounts used for Rdc1TE 1 in ***Table 6***. Following protein production half of each mixture was used in for SDS PAGE while the other half was frozen for future *in vitro* assays.

2.4.1.D……SDS PAGE separation of PURExpressed Rdc1 TE protein compared to Negative Control

Following PURExpression, half of the protein mixtures from ***2.4.1.B*** are prepared for SDS PAGE separation with protein loading buffer, boiled at 95 ˚C for 5 minutes, all contents were loaded into a 4-20% Mini-PROTEAN® TGX Stain-Free™ Gel, 12-well, 20 µl from Bio-Rad and separated via SDS PAGE at constant 175 volts, current of 500, for 45 minutes.

2.4.1.E……Visualizing of protein separation via SDS PAGE

The resulting gel is separated from its casing and placed into the Alpha Innotech Alpha Imager EV. The gel is subjected to a minute of UV light at 325 nm, and the produced bands are then captured and imaged.

2.4.2.A……*In vitro* Enzymatic Assay of Rdc1 TE and Substrate

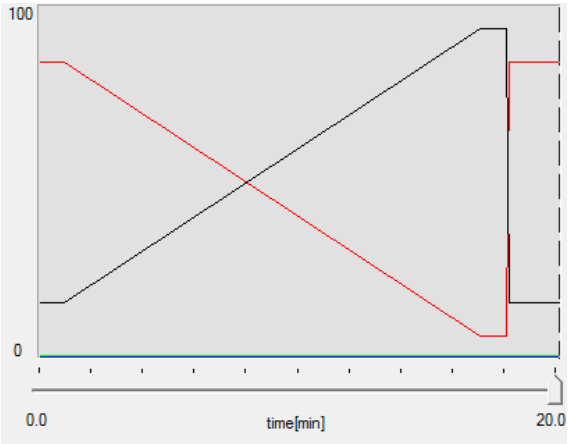
The frozen aliquot produced in section ***2.4.1.B*** was thawed combined with 6.25 µL of PURExpressed Rdc1 TE protein with 6.75 µL H_20, 5 µL of 200 mM Phosphate Buffer, 1 µL DMSO, and 1 µL substrate. Negative control was produced by substituting the PURExpressed Rdc1 TE with water instead. Mixtures were lightly mixed through pipetting and incubated overnight at room temperature.

2.4.2.B……Separation, via HPLC, and detection, via UV detector, of product

The assay was transferred to a plastic vial and the sample was separated using the Agilent Technologies 1260 Infinity HPLC. The ProntoSIL 120-5 Eurobond C18 5uM 125 x 4.00 mm was

used for the separation at an oven temperature of 30 ˚C. The sample was separated at a 1.0 mL / minute flow rate for a 20-minute run.

The method of separation was done using H_2O: Acetonitrile as the solvents following ***Figure 36***



0.00	Solvent Mix	85.0 15.0 0.0 0.0
1.00	Solvent Mix	85.0 15.0 0.0 0.0
17.00	Solvent Mix	5.0 95.0 0.0 0.0
18.00	Solvent Mix	5.0 95.0 0.0 0.0
18.10	Solvent Mix	85.0 15.0 0.0 0.0
20.00	Solvent Mix	85.0 15.0 0.0 0.0

Figure 36: HPLC Method of Separation for TE substrate assay
Separation method uses a mixture of water, A (Red line), and acetonitrile, B (Black Line) where initially it starts at a ratio of 85:15 (H_20: ACN) for a minute, swapping to a ratio of 5:95 over the course from minute 1 – minute 18 of the run and holding it for a minute. It immediately drops down to 85:15 from within 10 seconds and holds it till the end of the 20.

The chromatograms were produced using the using the 1260 DA VL measuring at wavelengths 280, 254, 220, 210 nm.

2.4.3.A…… *In vitro* Assays to investigate relationship between PURExpress incubation length and protein activity

The frozen aliquots produced in section ***2.4.1.C*** was thawed combined with 6.25 µL of PURExpressed Rdc1 TE protein with 6.75 µL H_20, 5 µL of 200 mM Phosphate Buffer, and 1 µL DMSO. Negative control was produced by substituting the PURExpressed Rdc1 TE with water instead. 1 µL of 14-membered ring substrate was added to each reaction. Mixtures were lightly mixed through pipetting and left to run overnight at room temperature.

2.4.3.B……Separation, via HPLC, and detection, via UV detector, of product

Separation and detection of product was conducted using the same method as reported in section ***2.4.2.B***.

Cell-Free Monocillin II Production

2.4.4.A……PURExpression of Rdc5, Rdc1, Rdc2, and SFP with a Negative Control

Individual Rdc5 (pFM50) , Rdc1 (pFM51), Rdc2 (pFM52) and SFP (pMEH22) proteins were expressed using the ¼ PURExpress conditions as shows in ***Table 1***, each with 2 µL of their respective plasmids. For the negative control, DNA was substituted with water and for the reaction containing all 4 plasmids, 0.5 µL of each was added instead. Reactions were incubated at 37˚C for 3.5 hours and placed on ice immediately afterwards.

2.4.4.B……SDS PAGE separation of PURExpressed proteins Rdc5, Rdc1 and SFP

Protein mixtures processed with protein loading buffer, boiled at 95 ˚C for 5 minutes to denature the proteins and then contents are loaded into a Mini-PROTEAN® TGX Stain-Free™ Protein Gels, 12 well, 20 µl and separated via SDS PAGE at 175 volts, current 400, for 45 minutes.

2.4.4.C……Coomassie staining of proteins to visualize Rdc5, Rdc1, and SFP

Following SDS PAGE separation, the gel is removed from its shell and put into a 10-uL tip tray. Fix solution is added to submerge the gel, which is then microwaved for two pulses of 30 seconds. The fix solution is removed and Stain/ Destain solution covers the gel followed by addition of 700 µL of coomassie brilliant blue stain. The solution is microwaved for 30 seconds, which is then placed on the counter. A Kimwipe® EX-L Delicate Task Wipe is placed over the gel to absorb the unused stain overnight.

The gel is then placed into the Alpha Innotech AlphaImager™ EV, and an image of the gel is taken.

2.4.5.A……PURExpression of Rdc1 at different temperatures

Rdc1, 1.875 µL of pFM51, was PURExpressed using ¼ PURE protocol **(*Table 1*)** at either room temperature, 30˚C, 37˚C, or 40˚C for 3.5 hours. No water was added, and the reaction was placed on ice afterwards.

The proteins were prepared with protein loading buffer, boiled and separated with SDS PAGE at a constant 170 volts, 400 mA for 45 minutes, followed by coomassie staining following the same protocol as section ***2.4.4.C*** and imaged with the Alpha Innotech AlphaImager™ EV.

2.4.5.B……PURExpression of Rdc1 using different expression lengths at 30˚C

Rdc1, 1.875 µL of DNA was PURExpressed using ¼ PURE protocol, ***Table 1***, either incubated for 2 hours, 3.5 hours, 5 hours, 6.5 hours or overnight at 30˚C. No water was added, and the reaction was placed on ice afterwards.

The proteins were prepared with protein loading buffer, boiled and separated with SDS PAGE at a constant 170 volts, 500 mA for 45 minutes, followed by coomassie staining following the same protocol as section ***2.4.4.C*** and imaged with the Alpha Innotech AlphaImager™ EV.

Visualization of proteins synthesis through incorporation of BODIPY-Lysine

2.4.6.A……PURExpression of Rdc5, Rdc1, SFP and a Negative Control with BODIPY-Lysine

Individual Rdc5 (pFM50) , Rdc1 (pFM51), and SFP (pMEH22) proteins were PURExpressed using the ¼ PURExpress conditions as shows in ***Table 1***, each with 1.5 µL of their respective plasmids and for the negative control, pET28 was used. BODIPY-Lysine, 0.375 µL, along with water, 0.375

µL, were also included in the reaction. Reactions were incubated at 30˚C for 3.5 hours and placed on ice immediately afterwards.

2.4.6.B……SDS PAGE separation of BODIPY-Lysine incorporated PURExpressed proteins

Protein mixtures are mixed with protein loading buffer, boiled at 95 ˚C for 5 minutes to denature the proteins and then contents are loaded into a Mini-PROTEAN® TGX Stain-Free™ Protein Gels, 12 well, 20 µl and separated via SDS PAGE at 175 volts, current 400, for 45 minutes.

2.4.6.C……Visualization of BODIPY-Lysine incorporated protein separation

Following SDS, the gel was removed from its shell and the gel was transferred to a ChemiDoc XRS+ System for visualization with Image Lab 6. The gel was subjected to the Alexa488 Blot option and the resulting image was captured.

Phase I Monocillin II Production Assays

2.4.7.A……PURExpression of Rdc5, Rdc1, SFP for *in vitro* production of Monocillin II

Individual Rdc5 (pFM50) , Rdc1 (pFM51), and SFP (pMEH22) proteins were expressed using the ¼ PURExpress conditions as shows in ***Table 1***, each with 2 µL of their respective plasmids. For the negative control, DNA was substituted with water. Reactions were incubated at 30˚C for 3.5 hours and placed on ice immediately afterwards.

2.4.7.B……Initial *In vitro* Enzymatic Assays between Rdc5, Rdc1 and SFP

Following PURExpression, the protein mixtures are combined together and Malonyl-CoA and NADPH, both in water, are added to reach a concentration of ~5 mM in the solution. This combination is incubated at room temperature overnight

2.4.7.C……Separation, via HPLC, and detection, via UV detector, of initial *in vitro* assays

The assay was transferred to a plastic vial and the sample was ran through the Agilent Technologies 1260 Infinity HPLC. The ProntoSIL 120-5 Eurobond C18 5uM 125 x 4.00 mm was used for the separation at an oven temperature of 30 ˚C. The sample was separated at a 1.0 mL / minute flow rate for an 18-minute run.

The method of separation was done using H_2O: Acetonitrile as the solvents using a separation method outlined in ***Figure 37.***

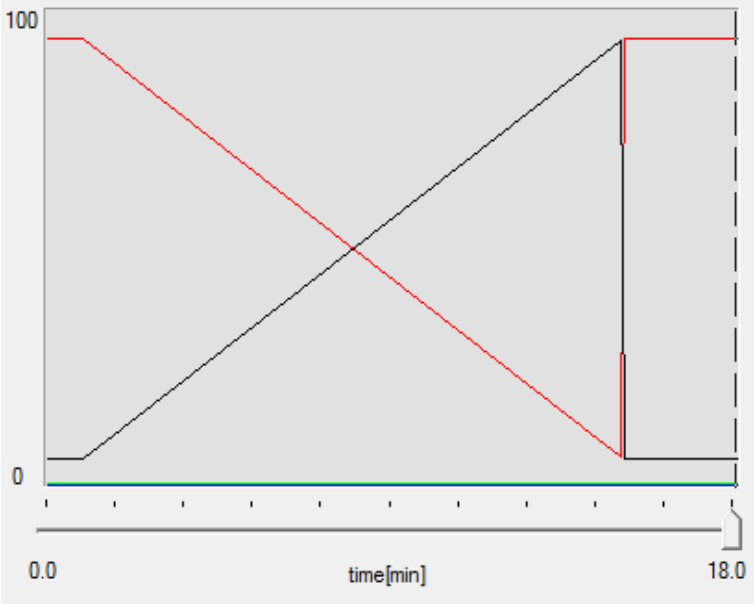


Figure 37: HPLC separation conditions for scouting on Monocillin II
Separation method uses a mixture of water, A (Red line), and acetonitrile, B (Black Line) where initially it starts at a ratio of 95:5 (H_20: ACN) for a minute, swapping to a ratio of 5:95 over the course from minute 1 – minute 15 of the run followed by immediately swapping to 95:5 within 10 seconds and holding it for the remainder of the 18 minute run.

The chromatograms were produced through UV detection, following HPLC separation, using the 1260 DA VL measuring at 300 nm.

2.4.7.D…… PURExpression of Rdc5, Rdc1, and SFP for double substrate concentration *in vitro* experiments

4 PURExpression reactions were produced, 1 for each protein with one negative control, with the same method as ***2.4.7.A.***

2.4.7.E……*In vitro* Enzymatic Assays between Rdc5, Rdc1 and SFP with double substrate concentrations

Table 7: Content volumes of reactions in 2.7.3.B

	1	2	Negative
PURE Rdc5	2.0	2.0	0.0
PURE Rdc1	2.0	2.0	0.0
PURE SFP	2.0	2.0	0.0
PURE pET28	0.0	0.0	6.0
50 mM Malonyl-CoA	2.0	4.0	2.0
50 mM NADPH	2.0	6.0	2.0
200 mM Phosphate Buffer	10.0	5.0	10.0
Total	20.0 µL		

Following PURExpression, the contents of each mixture were divided to produce *in vitro* assays as shown in ***Table 7.*** Reactions were incubated at 30°C for 3.5 hours, immediately placed on ice afterwards.

2.4.7.F……Separation, via HPLC, and detection, via UV detector, of product from section 2.7.3.D

The *in vitro* assays were extracted with EtOAc, evaporated to dryness and re-dissolved in Optima grade MeOH. Separation was conducted the same method as ***2.4.7.C***

Visualization of the modification of the ACPs with BODIPY-CoA and SFP

2.4.8.A…… Production of BODIPY-CoA for SFP modification

A scarce amount of BODIPY-Maleimide was added to 100 µL of DMSO, where 10 µL was added to a 1.9 mL solution of MES acetate and 100 mM Mg $(OAc)_2$ at pH 6.0 containing 300 µL of DMSO. Additionally, another scarce amount of CoA was added to the solution, and the mixture was vortexed briefly, cooled for 30 minutes on ice and then left to warm up on the counter for 15 minutes.

2.4.8.B……PURExpression of Rdc5, Rdc1, Rdc2, and SFP for Post-Translational Modification

Individual Rdc5 (pFM50) , Rdc1 (pFM51), and SFP (pMEH22) proteins were expressed using the ¼ PURExpress conditions as shows in ***Table 1***, each with 2 µL of their respective plasmids. For the negative control, DNA was substituted with water and for the reaction containing all 4 plasmids, 0.5 µL of each was added instead. Reactions were incubated at 30˚C for 3.5 hours and placed on ice immediately afterwards.

2.4.8.C…… Modification of Rdc5 and Rdc1 with BODIPY-CoA and SFP

Half of the PURExpress SFP mixture was added to the individual Rdc5 and Rdc1 PURExpress mixtures. 10 µL of the BODIPY-CoA mixture was added to the now Rdc5/SFP and Rdc1/SFP mixtures and left to incubate at room temperature for 30 minutes in darkness, eg. Inside a drawer.

2.4.8.D……SDS PAGE separation of BODIPY-CoA modified PURExpressed proteins

Protein mixtures are mixed with protein loading buffer, boiled at 95 ˚C for 5 minutes to denature the proteins and then contents are loaded into a 4-20% Mini-PROTEAN® TGX Stain-Free™ Gel, 12-well, 20 µl from Bio-Rad and separated via SDS PAGE at constant 175 volts, current of 500 mA, for 45 minutes.

2.4.8.E……Visualization of post-translational modification via BODIPY-CoA and SFP

Following SDS, the gel was removed from its shell and the gel was transferred to a ChemiDoc XRS+ System for visualization with Image Lab 6. The gel was subjected to the Alexa488 Blot option and the resulting image was captured.

2.4.8.F……Updated production of BODIPY-CoA for SFP modification

Approximately 1 mg of BODIPY-Maleimide was dissolved in 100 µL of DMSO and added to a 2 mL solution of 50 mM Tris-HCl, 50 mM NaCl, 100 mM MgCl pH 8.0. Additionally, ~1 mg of CoA was added to the solution, and the mixture was vortexed briefly, cooled for 45 minutes on ice and then left to warm up on the counter for 30 minutes.

2.4.8.G……Updated PURExpression of Rdc5, Rdc1, Rdc2, and SFP for Post-Translational Modification

Individual Rdc5 (pFM50), Rdc1 (pFM51), and SFP (pMEH22) proteins were expressed using newer PURExpression conditions, ***Table 5***. Reactions were incubated at 30˚C for 3.5 hours and placed on ice immediately afterwards.

2.4.8.H……Modification of Rdc5 and Rdc1 with BODIPY-CoA and SFP

Half of the PURExpress SFP mixture was added to the individual Rdc5 and Rdc1 PURExpress mixtures. 5 µL of the updated BODIPY-CoA mixture **(2.4.8.F)** was added to the now Rdc5:SFP and Rdc1:SFP mixtures and incubated at 30°C for approximately 3 hours.

2.4.8.I……SDS PAGE separation of Updated BODIPY-CoA modified PURExpressed proteins

Protein mixtures are mixed with protein loading buffer, boiled at 95 ˚C for 5 minutes to denature the proteins and then contents are loaded into a 4-20% Mini-PROTEAN® TGX Stain-Free™ Gel, 12-well, 20 µl from Bio-Rad and separated via SDS PAGE at constant 200 volts, current of 500 mA, for 45 minutes.

2.4.8.J……Visualization of post-translational modification via BODIPY-CoA and SFP

Following SDS, the gel was removed from its shell and the gel was transferred to a ChemiDoc XRS+ System for visualization with Image Lab 6. The gel was subjected to the Alexa488 Blot option and the resulting image was captured.

Phase II Monocillin II Production Assays

2.4.9.A……PURExpression of Rdc5 and Rdc1 at various temperatures.

Three groups of Rdc1 (pFM51), Rdc5 (pFM50) and SFP (pMEH22) were PURExpressed, each group incubated at a different temperature. using the ½ PURExpress conditions as shows in ***Table 1***, each with 2 µL of their respective plasmids. Group 1 was incubated at 28°C, group 2 incubated at 30°C and group 3 incubated at 37°C for 3.5 hours. Immediately placed on ice afterwards.

2.4.9.B……Post-Translational modification of PURExpressed of Rdc5 and Rdc1 with PURExpressed SFP and CoA

Half of the contents of the PURExpressed SFP reaction is divided equally between both the Rdc5 and Rdc1 reaction mixtures along with 5 µ L of 50 mM CoA (100 mM MES-Acetate pH 6.0) and incubated at room temperature for 30 minutes.

2.4.9.C……*In vitro* Combination assays of purified Rdc5 and Rdc1 with PURExpressed Rdc5 and Rdc1

Within each individual group, the Rdc5 and Rdc1 mixtures were combined and 10 µL NADPH and 10 µL Malonyl-CoA, both at ~50 mM concentrations in 100 mM MES-Acetate pH 6.0.

The *in vitro* combination reactions would be incubated at room temperature overnight.

2.4.9.D……Separation, via HPLC, and detection, via Mass Spectroscopy, of assay products

The *in vitro* assays were extracted with EtOAc, evaporated to dryness and re-dissolved in Optima grade MeOH. Separation was conducted the same method as ***2.4.7.C***

2.4.9.E……PURExpression of Rdc5 and Rdc1 at various temperatures.

Two groups of Rdc1 (pFM51), Rdc5 (pFM50) and SFP (pMEH22) were PURExpressed using the ¼ PURExpress conditions as shows in ***Table 1***, each with 2 µL of their respective plasmids. All reactions incubated at 30°C for 3.5 hours. Immediately placed on ice afterwards.

2.4.9.F……Post-Translational modification of PURExpressed of Rdc5 and Rdc1 with PURExpressed SFP and CoA

Half of the contents of the PURExpressed SFP reaction is divided equally between both the Rdc5 and Rdc1 reaction mixtures along with 5 µ L of 50 mM CoA (100 mM MES-Acetate pH 6.0) and incubated at room temperature for 30 minutes.

2.4.9.G……*In vitro* Combination assays of purified Rdc5 and Rdc1 with PURExpressed Rdc5 and Rdc1

Within each individual group, the Rdc5 and Rdc1 mixtures were combined and 10 µL NADPH and 10 µL Malonyl-CoA, both at ~50 mM concentrations in 100 mM MES-Acetate pH 6.0.

Group 1 was incubated at 30°C overnight while group 2 was incubated at 37°C overnight.

2.4.9.H……Separation, via HPLC, and detection, via Mass Spectroscopy, of assay products

The *in vitro* assays were extracted with EtOAc, evaporated to dryness and re-dissolved in Optima grade MeOH. Separation was conducted the same method as ***2.4.7.C.***

Supplementing PURExpression to increase protein production

2.4.10.A…… Initial PURExpression of Rdc1 with Supplemental Solutions of Substrates

Table 8: Composition matrix of Supplements in Supplementary Solution for Corresponding Lane in Figure 12a

PURExpress of Rdc1 was conducted using the reaction conditions outlined in ***Table 4***. While those conditions remained constant, the supplemental and water content varied following ***Table 8***.

	1	2	3	4	5	6	7	8	9
RNAse Inhibitor Murine:	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
(1 mM) rNTPs:	0.0	4.0	0.0	0.0	0.0	4.0	0.0	4.0	0.0
Creatine Phosphate:	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0
1 mM Amino Acids:	0.0	0.0	0.0	2.0	0.0	0.0	2.0	2.0	0.0
tRNA:	0.0	0.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0
Water:	9.0	6.0	9.0	8.0	9.0	5.0	7.0	1.0	10.0
Total	10.0 µL								

PURExpress reactions were then incubated at 30°C for 3.5 hours.

2.4.10.B……SDS PAGE separation of PURExpressed proteins

Protein mixtures are mixed with protein loading buffer, boiled at 95 ˚C for 5 minutes to denature the proteins and then contents are loaded into a Mini-PROTEAN® TGX Stain-Free™ Protein Gels, 12 well, 20 µl and separated via SDS PAGE at 175 volts, current 400, for 45 minutes.

2.4.10.C……Coomassie Staining of SDS-PAGE gel to visualize supplemented Rdc1

Coomassie staining, and visualization, was conducted similar to section ***2.4.4.C***.

2.4.10.D…… Second PURExpressions of Rdc1 with Supplemental Solutions of Substrates

PURExpress of Rdc1 was conducted using the reaction conditions outlined in ***Table 4***. While those conditions remained constant, the substrate and water content varied following ***Table 9***

PURExpress reactions were then incubated at 30°C for 3.5 hours.

Table 9: Composition matrix of Supplements in Supplementary Solution for Corresponding Lane in Figure 12b

	1	2	3	4	5	6	7	8	9
RNAse Inhibitor Murine:	0.5	1.0	0.5	0.5	0.5	0.5	0.5	0.5	0.5
(1 mM) rNTPs:	0.0	0.0	2.0	0.0	0.0	0.0	2.0	0.0	0.5
Creatine Phosphate:	0.0	0.0	0.5	0.0	0.0	0.0	0.5	0.0	0.0
1 mM Amino Acids:	0.0	0.0	0.0	2.0	3.5	2.0	3.5	0.0	2.5
tRNA:	0.0	0.0	0.0	0.0	0.0	4.0	4.0	0.0	2.0
Water:	10.0	9.5	7.5	8.0	6.5	4.0	0.0	10.0	5.0
Total	10.5 µL								

2.4.10.E……SDS PAGE separation of PURExpressed proteins

Protein mixtures are mixed with protein loading buffer, boiled at 95 ˚C for 5 minutes to denature the proteins and then contents are loaded into a Mini-PROTEAN® TGX Stain-Free™ Protein Gels, 12 well, 20 µl and separated via SDS PAGE at 175 volts, current 500, for 45 minutes.

2.4.10.F…… Coomassie Staining of SDS-PAGE gel to visualize second round of supplemented Rdc1

Coomassie staining, and visualization, was conducted similar to section ***2.4.4.C***.

Production of an authentic Monocillin II standard in *S. cerevisiae*

2.4.11.A……Production of pKJ61 + pKJ91 Minimal Media Agar Plates

Yeast Synthetic Drop-out Medium Supplements without Uracil, Leucine, and tryptophan, 0.584 g, Yeast Nitrogen Base without amino acids, 2.696g, and Agar, 8.00g, would be added to 400 mL of water. 0.2 mL of 1.9 g/ mL of Leucine and 0.2 mL of 16 mL of 40g/ mL, 50% (w/v) Glucose would be filtered and added to the flask. The leucine solution would need to be previously autoclaved while the glucose solution would be dissolved and then autoclaved. The flask would then be autoclaved with plates being poured afterwards, once the flask has cooled down to safely handle. The plates would then be stored at 4°C

2.4.11.B……Growing of *S. cerevisiae* strain BJ5464-NpgA harboring pKJ91 and pZH223 colonies

A frozen aliquot of *S. cerevisiae* strain BJ5464-NpgA harboring pKJ91 and pZH223 was taken out of the -80°C freezer, scraped and streaked on a previously produced minimal media plate (SMM) with leucine. The SMM + Leu plate was left to incubate at 30°C for 3 days.

2.4.11.C……Overnight culture production using colony of *S. cerevisiae* strain BJ5464-NpgA harboring pKJ91 and pZH223

A 5 mL overnight culture of YPD media was produced using previously autoclaved YP media and 0.2 mL of filtered glucose (40g/ mL, 50% (w/v)). A single colony of *S. cerevisiae* strain BJ5464-NpgA harboring pKJ91 and pZH223 from the SMM + Leucine plate was incubated in the overnight culture and left to grow at 30°C, being shaken at 200 rpm overnight.

2.4.11.D……Inoculation of overnight culture into YPD media for protein expression and Monocillin II production

2 mL of the overnight culture was inoculated into a flask of 400 mL YPD in a 1 L flask. The flask was left to grow at 30°C with 200 rpm of shaking for 3 days. An additional 100 mL of YPD would be added to the flask and left to incubate an additional two days.

2.4.11.E……Isolation of Monocillin II

Following the 5 days of incubation, NaOH was added to the flask to drop the pH to 5.0 to lyse the cells. The insoluble contents of the flask would be pelleted through centrifugation at 4,000 x g for 30 minutes. The supernatant would be extracted twice with EtOAc, evaporated to dryness and re-dissolved in methanol.

2.4.11.F……Separation, via HPLC, and detection, via UV detector, of Monocillin II

Separation and detection of Monocillin II using HPLC follows the exact protocol as in section ***2.4.7.C***.

Phase III Monocillin II Production Assays

2.4.12.A……PURExpression of Rdc5, Rdc1, and SFP for LC-MS detection

Individual Rdc5 (pFM50) , Rdc1 (pFM51), and SFP (pMEH22) proteins were expressed using the PURExpress conditions as shows in ***Table 5,*** with the negative control using pET28 as its DNA source. Reactions were incubated at 30˚C for 3.5 hours and placed on ice immediately afterwards.

2.4.12.B……Post-Translational Modification of Rdc5 and Rdc1 via SFP and Malonyl-CoA

Contents of the SFP reaction is divided equally between Rdc5, Rdc1 and the negative control mixtures along with 5 µ L of 50 mM CoA in 100 mM MES-Acetate pH 6.0 being added to each. Reactions are incubated at room temperature for 30 minutes.

2.4.12.C……*In vitro* Enzymatic Assay for production of Monocillin II for LC-MS detection

The contents of Rdc5 and Rdc1 are combined and 5 µL NADPH and 5 µL Malonyl-CoA, both at 100 mM concentrations in 100 mM MES-Acetate pH 6.0, are added. 5 µL are also added to the negative control assay. Reactions are incubated overnight at room temperature.

2.4.12.D……Separation, via HPLC, and detection, via Mass Spectroscopy, of assay products

Following *in vitro* assay incubation, the enzymatic and negative control assays were transferred to plastic vials which were placed into the tray of the Shimadzu Prominence UFLC, which is then separated in a ProntoSIL 120-5 Eurobond C18 5uM 125 x 4.00 mm column at an oven temperature of 30 ˚C. The sample was separated with a flow rate of 1.0 mL / minute for an 18-minute run.

The method of separation was done using H_2O: Acetonitrile as the solvents using a separation method outlined in ***Figure 38***.

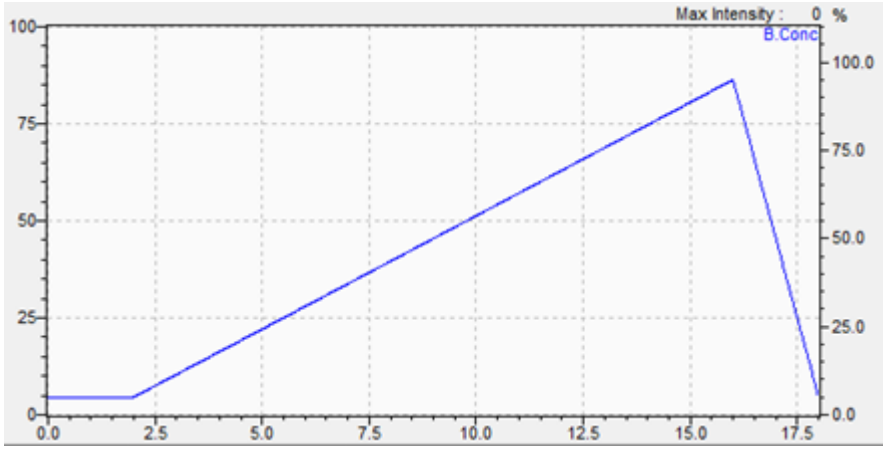


Figure 38: Initial separation method of compounds in Shimadzu UFLC for Mass Spectroscopy in LCMS-2020
Separation method involved mixtures of H_2O: Acetonitrile as the solvents, where initially it starts at a ratio of 95:5 for 2 minutes, swapping to a ratio of 5:95 over the course from minute 2 – minute 16 of the run, dropping down to 5:95 from minute 16 to the end of the run. The blue graph shows a graphical representation of the composition of acetonitrile during the 18-minute run.

The separated compounds would elute into the Shimadzu LCMS-2020 for mass spectroscopy.

The positive scan would scan through the range of 310 – 400 m/z.

2.4.12.E……Separation, via HPLC, and detection, via Mass Spectroscopy, of Monocillin II in Yeast Extract using new method

The yeast extracts would be diluted 1:10 with MeOH and transferred to a plastic vial which were placed into the tray of the Shimadzu Prominence UFLC, which is then separated in a Dikma Technologies Inc. Leapsil 2.7 µm C18, 100 x 2.1 mm column at an oven temperature of 30 ˚C. The sample was separated at a 0.25 mL / minute flow rate for a 30-minute run. The method of separation was done using H_2O: Acetonitrile as the solvents using a separation method outlined in ***Figure 39***.

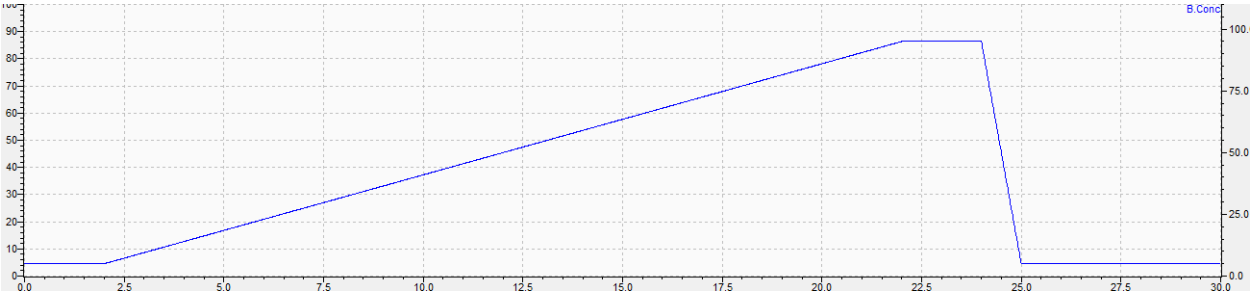


Figure 39 Updated separation method of compounds in Shimadzu UFLC for Mass Spectroscopy in LCMS-2020
Separation method involved mixtures of H_2O: Acetonitrile as the solvents, where initially it starts at a ratio of 95:5 for 2 minutes, swapping to a ratio of 5:95 over the course from minute 2 – minute 22 of the run, remaining at 5:95 for 2 minutes. It would then revert to 95:5 from minute 24 to minute 25 and remaining at 95:5 until the end of the run. The blue graph shows a graphical representation of the composition of acetonitrile during the 30-minute run.

The separated compounds would elute into the Shimadzu LCMS-2020 for mass spectroscopy. Usually a positive scan, scanning a range of 300 – 370 m/z at a speed of 72 u/second, is run first followed by a negative scan, scanning a range of 300 – 325 m/z at a speed of 27 u/second.

2.4.12.F……PURExpression of Rdc5, Rdc1, and SFP with a Negative Control for *in vitro* assays using different buffers

Individual Rdc5 (pFM50) , Rdc1 (pFM51), and SFP (pMEH22) proteins were expressed using 2x the PURExpress conditions as shows in ***Table 5,*** except for SFP***,*** with the negative control using pET28 as its DNA source. Reactions were incubated at 30˚C for 3.5 hours and placed on ice immediately afterwards.

2.4.12.G……Post-Translational Modification of Rdc5 and Rdc1 via SFP and Malonyl-CoA

Contents of the SFP reaction is divided equally between Rdc5, Rdc1 and the negative control mixtures along with 5 µ L of 50 mM CoA (50 mM Tris-HCl, 50mM NaCl, 100 mM MgCl, pH 8.0) and incubated at room temperature for 30 minutes.

2.4.12.H……*In vitro* Enzymatic Assay of for production of Monocillin II using different buffers

Enzymatic assays were conducted following ***Table 10***. The buffer used in the negative control assay was Tris-HCl, NaCl. Concentrations of buffers were: 200 mM Phosphate buffer (pH 7.4), 50 mM Tris-Cl, NaCl (pH 8.0) and 100 mM MES-Acetate (pH 6.0). Assays were incubated at room temperature overnight.

	Phosphate	Tris-HCl,NaCl	MES	Negative
Table 10: Composition of investigation of supplemented buffer on production of Monocillin II				
Rdc5	9	9	9	0
Rdc1	9	9	9	0
Negative	0	0	0	18
Buffer	6	6	6	6
NADPH	5	5	5	5
M-CoA	5	5	5	5
H_20	1	1	1	1
Total		35 µL		

2.4.12.I……Separation, via HPLC, and detection, via Mass Spectroscopy, of *in vitro* assay results using different buffers

In vitro assays were extracted with EtOAc, evaporated to dry and re-dissolved in Optima grade MeOH. Separation and detection would occur similarly as section ***2.4.12.E***

Combination experiments using purified multi-domain with Cell-Free multi-domain proteins

2.4.13.A……PCR of pZH233 to produce HIS tagged linear insert of Rdc1

General distribution of components followed ***Table 11***, but before the reactions were produced 2.5 of 100 mM primer was diluted in 22.5 µL water. PCR's were conducted using pMRH08 as the template using the method outline in ***Table 12*** with varying temperatures for variable 1 and variable 2, which were ((Tm forward Primer + Tm Reverse Primer)/2)-5. Primers "HIS Rdc5 FW" and "HIS_pZH223_RV"were used for this PCR while the method used 64°C for both variable temperatures.

Following PCR, the reactions were subjected to 1% agarose gel electrophoresis. The successful reactions, where a band was present at the right MW region, were excised from the gel and the linear DNA fragment was purified using the E.Z.N.A.® Gel Extraction Kit.

Table 11: Standard PCR concentrations used

H_2O	35.5
Herculase II Buffer (5x)	10.0
Forward Primer (10 mM)	1.25
Reverse Primer (10 mM)	1.25
DNA	1.00
dNTPs	0.50
Herculase II Fusion Polymerase	0.50
Total :	50.00 µL

Table 12: Standard PCR method with 2 Variable Temperatures

Step 1	98˚C	5:00
Step 2	98˚C	0:20
Step 3	Variable Temp 1	0:20
Step 4	68˚C	1:00
Step 5	Go to Step 2, -1 ˚C /cycle	14 X
Step 6	98˚C	0:20
Step 7	Variable Temp 2	0:20
Step 8	72˚C	1:00
Step 9	Go to Step 6	16 X
Step 10	72˚C	10:00
Step 11	12˚C	∞

2.4.13.B……Cloning of linear HIS tagged insert of Rdc1 into expression vectors

The linear fragments produced by PCR were digested, alongside pZH233, with NheI-HF at 37˚C overnight with Cut Smart being used as the buffer. The next day, 5 µL of 500 mM NaCl is added to the digests along with adding 1 µL of SwaI and left to incubate at 25˚C for 4 hours. Digested bands were separated via 1% agarose gel electrophoresis, excised from the gel, and purified with using the Monarch® DNA Gel Extraction Kit. The digested linear fragment and digested pZH233 fragment were ligated together using T4 ligase. Ratios of Insert: Vector were calculated to be 5: 1 and the ligation reactions were run overnight at 16˚C.

Expression vector stocks were produced through transformation of ligated vectors were into chemically competent XL1-BLU *E. coli* cells and plated on LB-Kan plates. Colonies were picked and overnight cultures were produced. The plasmid DNA was isolated using the E.Z.N.A.® Plasmid Mini Kit I. Check digests were conducted on each plasmid isolated to ensure that the purified DNA contains the right inserts. This was done by a small quantity of each with NheI-HF, separated with 1% agarose gel electrophoresis and visualized.

2.4.13.C……Transforming pkJ91 and pKJ61 into chemically component *S. cerevisiae* BJ5464-NpgA

The protocol for transformation closely follows that of literature. (R Daniel Gietz, 2007)

From previously plated *S. cerevisiae* BJ5464-NpgA, ~10 colonies are transferred to a 50 µL solution of sterile water within a 1.5 mL microcentrifuge tube. 950 µL additional water for resuspension of cells, which were then pelleted via centrifugation at 13,000 x g.

To the pellet, 240 µL of PEG, 36 µL of LiAc, 50 µL of pre-boiled (95°C for 5 minutes) single stranded salmon sperm DNA, 34 µL of water is added along with 0.75 µL of pKJ91 and 0.75 µL of pKJ61. Contents was vortexed for a minute and incubated at 42°C for ~3 hours.

The pellet is then centrifuged, and the supernatant is replaced with 500 µL water, vortexed till resuspended with 250 µL then being plated onto SMM + Leu plates. The plates were incubated at 30 °C for 3 days.

2.4.13.D……Growing up transformed *S. cerevisiae* BJ5464-NpgA for protein isolation

Overnight cultures, 5 mL, of pKJ61 *S. cerevisiae* BJ5464-NpgA and pKJ91 *S. cerevisiae* BJ5464-NpgA were produced using SMM + Leucine. The following day, 5% of the overnight culture was inoculated into 400 mL of YPD.

The 1 L flask was then shaken at 200 rpm at 30˚C for two days. Post two days add another 100 mL of YP to each flask and let it continue shaking at 200 rpm at 30˚C for 2 more days.

2.4.13.E……Lysis and isolation of protein soluble fraction

The contents of both flasks were transferred into centrifuge buckets, which were weighed and balanced. The cells would be pelleted at 4,000 rpm for 25 minutes. The supernatant is discarded, and the pellets are resuspended in 5 mL lysis buffer and transferred into 50 mL centrifuge tube. From this point, the centrifuge tube is constantly on ice. The cells would be lysed via sonication, with 3 cycles of 45 seconds on and 30 seconds off, followed by the insoluble contents pelleted at 10,000 rpm for 60 minutes.

2.4.13.F……Isolation of *S. cerevisiae* BJ5464-NpgA expressed Rdc1 and Rdc5 proteins via Ni-NTA column chromatography

The soluble fraction supernatants are slowly transferred to their individual 15 mL centrifuge tube, where 800 mL of 50% Ni-NTA resin is added to each. The centrifuge tubes are then subjected to spinning at 4˚C for 1.5 hours.

The contents of the tubes are transferred to their own columns and fractions, which are produced by going from low concentrations of imidazole to higher, are acquired with the imidazole concentration, volumes and fraction amounts of ***Table 13***.

Table 13: Standard imidazole fractions used for protein purification

Imidazole Concentrations:	0.00 mM	20 mM	100 mM	250 mM
Volume:	4.5 mL		2.25 mL	
# of fractions:	1		2	

2.4.13.G……Condensing fractions and Amicon filtering of Rdc1 and Rdc5 proteins

Aliquots, 20 µL, of each fraction for both proteins were prepared and loaded into a STAIN FREE gel for SDS PAGE separation.

Fractions both 100 mM imidazole and 250 mM imidazole fractions, of both proteins, were combined into their respective Amicon® Ultra-15 Centrifugal Filter Units, which were then centrifuged at 5,000 x g for 45 minutes.

Following centrifugation, the filtered solution is discarded and ~ 10 mL of dialysis buffer (50 mM NaCl, 50 mM Tris – HCl, pH 8.0) which was then centrifuged at 5,000 x g for 45 minutes, and the filtered solution removed. This step occurred 2 more times and finally concentrated into ~500 µL of supernatant. 60% glycerol was diluted into both solutions to reach 10%, which the tubes were then aliquoted into 35 µL fractions and flash frozen with liquid nitrogen, and stored at -80˚C.

2.4.13.H……PURExpression of Rdc5 and Rdc1

Rdc1 (pFM51), Rdc5 (pFM50) and SFP (pMEH22) were PURExpressed using double the conditions in ***Table 5*** and incubated at 30°C for 3.5 hours. Immediately placed on ice afterwards

2.4.13.I……Post-Translational modification of PURExpressed of Rdc5 and Rdc1 with PURExpressed SFP and CoA

Half of the contents of the PURExpressed SFP reaction is divided equally between both the Rdc5 and Rdc1 reaction mixtures along with 5 µ L of 50 mM CoA (50 mM Tris-HCl, 50mM NaCl, 100 mM MgCl, pH 8.0) and incubated at room temperature for 30 minutes.

2.4.13.J……*In vitro* Combination assays of purified Rdc5 and Rdc1 with PURExpressed Rdc5 and Rdc1

The PURExpressed, and post-translationally modified, Rdc5 and Rdc1 were mixed with either *S. cerevisiae* Rdc5 or Rdc1 based on the combinations below.

- *S. cerevisiae* Rdc5 with *S. cerevisiae* Rdc1

- *S. cerevisiae* Rdc5 with PURExpressed Rdc1
- *S. cerevisiae* Rdc1 with PURExpressed Rdc5
- PURExpressed Rdc1 with PURExpressed Rdc5

5 µL protein mixtures in each combination were contributed to each of the combination, along with an addition of 5 µL NADPH and 5 µL Malonyl-CoA, both at 100 mM concentrations in 50 mM Tris-HCl, 50 mM NaCl, pH 8.0.

The *in vitro* combination reactions would be incubated at room temperature overnight.

2.4.13.K……Separation, via HPLC, and detection, via Mass Spectroscopy, of assay products

In vitro assays were extracted with EtOAc, evaporated to dry and re-dissolved in Optima grade MeOH. Separation and detection would occur similarly as section ***2.4.12.E.***

Production of TAGless Rdc5 and Rdc1 expression plasmids for Phase I Production assays of Monocillin II using TAGless Rdc5 and Rdc1

2.4.14.A……Cloning of Rdc5 and Rdc1 gene into expression vectors that would remove the TAG

pFM50 along with pET21 were digested with NheI using 2.1 Buffer while pFM51 and additional pET21 were digested using NdeI, both groups incubated at 37°C overnight. Following initial digestion NotI was added to pFM50, and it's respective pET21, and incubated at 37°C while SwaI was added to pFM51, and it's respective pET21, and incubated at 25°C for 2 hours.

Digested bands were excised from the gel, purified with using the E.Z.N.A.® Gel Extraction Kit and the digested linear vector and insert combinations were ligated together using T4 ligase. Ratios of Insert: Vector were calculated to be 5: 1 and the ligation reactions were incubated overnight at 16˚C.

Expression vector stocks were produced through transformation of ligated vectors were into chemically competent XL1-BLU *E. coli* cells and plated on LB-Ampicillin plates. Colonies were picked and overnight cultures were produced. The plasmid DNA was isolated using the E.Z.N.A.® Plasmid Mini Kit I. Check digests were conducted on each plasmid isolated to ensure that the purified DNA contains the right inserts. This was done by a small quantity of pCAS1 being digested with NheI and pCAS2 digested with NdeI for an hour at 37°C. Confirmation was done through separation of digests with 1% agarose gel electrophoresis and visualized.

2.4.14.B……PURExpression of TAGless Rdc5 and Rdc1

Rdc1 (pCAS1), Rdc5 (pCAS2) and SFP (pMEH22) were PURExpressed using the conditions in ***Table 5*** and incubated at 30°C for 3.5 hours. Immediately placed on ice afterwards

2.4.14.C……Post-Translational modification of PURExpressed of Rdc5 and Rdc1 with PURExpressed SFP and CoA

SFP would be divided between Rdc5, Rdc1 and the negative control. Each mixture would be fed 2.5 µL of ~100 mM Malonyl CoA and 2.5 µL of 100 mM NADPH, both in 100 mM MES-Acetate (pH 6.0). The reactions would be incubated at room temperature for 30 minutes.

2.4.14.D……*In vitro* assays of PURExpressed TAGless Rdc5 and Rdc1

The PURExpressed, and post-translationally modified, TAGless Rdc5 would be combined with TAGless Rdc1. The enzymatic assay, along with the negative control assay, would be incubated overnight at room temperature.

2.4.14.E……Separation, via HPLC, and detection, via Mass Spectroscopy, of assay products

In vitro assays were extracted with EtOAc, evaporated to dry and re-dissolved in Optima grade MeOH. Separation and detection would occur similarly as section ***2.4.12.E***

Phase II Monocillin II Production Assays using TAGless proteins

2.4.15.A……PURExpression of Rdc5 and Rdc1 at various temperatures.

Three groups of Rdc1 (pFM51), Rdc5 (pFM50) and SFP (pMEH22) were PURExpressed, each group incubated at a different temperature. Conditions were similar to ***Table 5*** with group 1 incubated at 28°C, group 2 incubated at 30°C and group 3 incubated at 37°C for 3.5 hours. Immediately placed on ice afterwards.

2.4.15.B……Post-Translational modification of PURExpressed of Rdc5 and Rdc1 with PURExpressed SFP and CoA

Half of the contents of the PURExpressed SFP reaction is divided equally between both the Rdc5 and Rdc1 reaction mixtures along with 1 µ L of 50 mM CoA (50 mM Tris-HCl, 50mM NaCl, 100 mM MgCl, pH 8.0) and incubated at 30°C for 30 minutes.

2.4.15.C……*In vitro* Combination assays of purified Rdc5 and Rdc1 with PURExpressed Rdc5 and Rdc1

Within each individual group, the Rdc5 and Rdc1 mixtures were combined and 5 µL NADPH and 4 µL Malonyl-CoA, both at ~50 mM concentrations in 50 mM Tris-HCl, 50 mM NaCl, pH 8.0 were added along with another 5 µL of 50 mM Tris-HCl, 50 mM NaCl, pH 8.0.

The *in vitro* combination reactions would be incubated at room temperature overnight.

2.4.15.D……Separation, via HPLC, and detection, via Mass Spectroscopy, of assay products

In vitro assays were extracted with EtOAc, evaporated to dryness and re-dissolved in Optima grade MeOH. Separation and detection would occur similarly as section ***2.4.12.E***

2.4.15.E……PURExpression of Rdc5 and Rdc1 for increased NADPH concentration or recycling

Two groups of Rdc1 (pFM51), Rdc5 (pFM50) and SFP (pMEH22) were PURExpressed, with conditions similar to ***Table 5***. Both groups incubated at 30°C for 3.5 hours. Immediately placed on ice afterwards.

2.4.15.F……Post-Translational modification of PURExpressed of Rdc5 and Rdc1 with PURExpressed SFP and CoA

Half of the contents of the PURExpressed SFP reaction is divided equally between both the Rdc5 and Rdc1 reaction mixtures along with 1 µ L of 50 mM Malonyl-CoA (50 mM Tris-HCl, 50mM NaCl, 100 mM MgCl, pH 8.0) and incubated at 30°C for 30 minutes.

2.4.15.G……*In vitro* Combination assays of purified Rdc5 and Rdc1 with PURExpressed Rdc5 and Rdc1 with either NADPH or GDH and glucose

Within group 1, the Rdc5 and Rdc1 mixtures were combined and 15 µL NADPH ~100 mM and 4 µL ~50 mM Malonyl-CoA in 50 mM Tris-HCl, 50 mM NaCl, pH 8.0 were added.

Within group 2, the Rdc5 and Rdc1 mixtures were combined with 4 µL ~50 mM Malonyl-CoA in 50 mM Tris-HCl, 50 mM NaCl, pH 8.0 being added. Additionally, 10 µL of heterologously expressed and purified GDH was added along with 10 µL of 50 mM NADPH/ Dextrose solution.

The *in vitro* reactions would be incubated at room temperature overnight.

2.4.15.H……Separation, via HPLC, and detection, via Mass Spectroscopy, of assay products

In vitro assays were extracted with EtOAc, evaporated to dryness and re-dissolved in Optima grade MeOH. Separation and detection would occur similarly as section ***2.4.12.E***

Production of Monocillin II within *E. coli* strain BAP1

2.4.16.A……Transfer of Rdc1 into pET28

pFM50 (Rdc1) along with pET21 were digested with NotI-HF and BglII overnight and the digested linear fragments were separated with 1% agarose gel electrophoresis. Similarly, pCAS01 (Rdc1) along with pET28 were digested with NotI-HF and BglII overnight and the digested linear fragments were separated with 1% agarose gel electrophoresis. The appropriate bands were excised and cleaned up using the E.Z.N.A.® Gel Extraction Kit.

The bands were ligated together using T4 DNA ligase with an Insert: Vector ratio of 3:1 overnight at 16˚C. Ligations were then transformed into XL1-Blue to produce stocks of plasmids, which were isolated using the E.Z.N.A.® Plasmid Mini Kit I.

2.4.16.B……PURExpression of Rdc1 and Rdc5 to verify integrity of transcription/ translation

PURExpression experiments were conducted using the modern conditions (***Table 5***) on pFM50, pFM51, pCAS01, pCAS02, pCAS03 and pCAS04 as well as a pET28 as a negative control. Reactions were incubated at 30˚C for 3.5 hours.

Following PURExpression, samples were prepared with protein loading buffer and loaded into a Mini-PROTEAN® TGX Stain-Free™ Protein Gels, 12 well, 20 µl. Proteins were separated at 200 volts for 45 minutes.

2.4.16.C……Transforming of HIS tagged Rdc5 and Rdc1 into BAP1 and Heterologous Expression of HIS tagged Rdc5 and Rdc1

HIS tagged Rdc5 (pFM50) and HIS tagged Rdc1 (pCAS04) were transformed into BAP1 and played on LB – Kan – Amp plates. Following overnight growing of colonies an overnight culture, 5mL, of *E. coli* strain BAP1 harboring pFM50 and pCSA04 was produced to be inoculated into

400 mL of LB, which would then be shaken at 200 rpm at 37˚C until OD_{600}= ~0.6. IPTG would be added, 400 µL of 1M IPTG, and the flask would be left to express at 30˚C overnight.

2.4.16.D……Distribution of culture contents

The following day, 5 mL of culture would be taken out and the cells would be pelleted in a 1.5 mL centrifuge tube. The RNA would then be extracted from this pellet using the RNeasy Mini Kit.

5mL of culture would also be taken out and the cells would be pelleted in a 1.5 mL microcentrifuge tube. This would then be subjected to denaturing conditions.

The remaining culture would be centrifuged in centrifuge buckets at 4,000 rpm for 30 minutes. The pellet would remain in the bucket for lysis while the brother is transferred into 1 L flask for extraction to isolate the broth.

2.4.16.E……RNA extraction using RNeasy Mini Kit

200 mL TE buffer (30 mM Tris Cl, 1mM EDTA, pH 8) would be made. An aliquot of 5 mL would be isolated and mixed with ~75 ng of lysozyme. Approximate 1 mL of Bacteria RNA Protect Reagent is added to the tubes, vortexed for 5 seconds and incubated at room temperature for 5 minutes. This would then be centrifuged for 10 minutes at 5,000 x g.

The resulting pellet would be resuspended in TE/Lysozyme, 200 µL, along with 20 µL of Proteinase K, 40 ng/ mL, which would then be incubated at room temperature with spinning, 200 rpm, for 10 minutes.

RLT buffer, 700 µL, would be added and the contents would be vortexed vigorously followed by addition of 500 µL of 99% ethanol. The solution would be transferred into a RNeasy column and the remaining protocol was followed per the kit's instruction.

2.4.16.F……Reverse Transcription of the isolated RNA

Two reverse transcription reactions would be produced containing 5 µL of eluted RNA, 2 µL of Random Primer 6, 4 µL of dNTPs (1 mM), and 5 µL of water. These would be heated for 4.5 minutes at 75˚C and immediately placed on ice afterwards.

Reverse transcriptase buffer, 2 µL of 10X, and RNAse Inhibitor Murine, 1 µL, would be added to both. One tube would get 1 µL of Reverse Transcriptase while the second would get 1 µL of water, to act as the negative control. These reactions would then be incubated at 42˚C for 1 hour followed by inactivation at 90˚C for 10 minutes.

2.4.16.G……PCR of Reverse Transcripts for Rdc5 and Rdc1 gene detection

Four reactions total, two for Rdc5 and 2 for Rdc1 gene detection, would be produced using the same PCR contents would be followed as ***Table 11*** using the primers "HIS Rdc5 FW" and "HIS_pZH223_RV" for Rdc5 samples and primers "Rdc1_SAT_NdeI_F" and "Rdc1_no_SAT_Bsu36I_R" for the Rdc1 samples. The method in ***Table 12*** was used with the variable temps for Rdc5 being 64˚C and variable temps for Rdc1 being 58.

2.4.16.H……Isolating proteins from BAP1 using Urea for Denatured Conditions

The pellet isolated for denatured conditions would then be resuspended in 200 µL of 8 M urea and incubated at room temperate for 10 minutes. The suspension would be centrifuged at 16,000 x g for 10 minutes and the supernatant would be carefully transferred to a clean 1.5 mL microcentrifuge tube.

2.4.16.I……Isolating proteins from BAP1 using normal conditions

Proteins would be isolated similar to section ***2.4.13.F*** using the same buffers and process. For this instance, the proteins were not purified using HIS tag affinity chromatography and instead following lysis the proteins were subjected to SDS PAGE separation and western detection of HIS tagged proteins.

2.4.16.J……SDS PAGE separation of denatured condition and normal conditioned proteins

A protein sample from each pool, denatured and normal conditions, were prepared for SDS PAGE and loaded into an SDS PAGE gel alongside Precision Plus Protein™ All Blue Prestained Protein Standard. Contents were separated at 200 volts for 45 minutes.

2.4.16.K……Western detection of HIS tagged Rdc5 and Rdc1 proteins

Following imaging, the gel was transferred onto a nitrocellulose membrane at 40 volts for 90 minutes. The membrane was then blocked with 5% milk TBST solution overnight at 4˚C.

The membrane was then rinsed with 25 mL TBST three times for ~10 minutes, probed with FLAG tag antibody, 10 µL of antibody in 25 mL 5% milk TBST, for 90 minutes.

Membrane was washed with 25 mL TBST three times for ~10 minutes again and then transferred to a new petri dish. 4 mL of chemiluminescence solution was added to the membrane, and then incubated at room temperature for 5 minutes.

The chemiluminescence of the membrane was then visualized with the imager using the chemiluminescence function.

2.4.16.L……Extraction of broth to acquire any Monocillin II

The broth from the culture was extracted twice with 99% ethyl acetate, with sodium sulfate added to the organic phase afterwards to remove any aqueous impurities. The organic extracts would be filtered and evaporated to dryness and then re-dissolved into methanol.

2.4.16.M……HPLC separation and LC-MS detection of Monocillin II in BAP1 broth extract

Separation and detection of assay products were exactly as outlined in section ***2.4.12.E.***

References

1. Yuzawa S, Backman TWH, Keasling JD, Katz L. Synthetic biology of polyketide synthases. *J Ind Microbiol Biotechnol*. 2018;45(7):621-633. doi:10.1007/s10295-018-2021-9
2. Julien B, Shah S, Ziermann R, Goldman R, Katz L, Khosla C. Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum. *Gene*. 2000;249(1-2):153-160. doi:10.1016/S0378-1119(00)00149-9
3. Wang GQ, Chen GD, Qin SY, et al. Biosynthetic pathway for furanosteroid demethoxyviridin and identification of an unusual pregnane side-chain cleavage. *Nat Commun*. 2018;9(1). doi:10.1038/s41467-018-04298-2
4. Hassan MI, Lundgren BR, Chaumun M, et al. Total Biosynthesis of Legionaminic Acid, a Bacterial Sialic Acid Analogue. *Angew Chemie - Int Ed*. 2016;55(39):12018-12021. doi:10.1002/anie.201606006
5. Owen JG, Charlop-Powers Z, Smith AG, et al. Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc Natl Acad Sci U S A*. 2015;112(14):4221-4226. doi:10.1073/pnas.1501124112
6. Morton CL, Potter PM. *Expression of Rabbit Liver CE 193 RESEARCH 193 Comparison of Escherichia Coli, Saccharomyces Cerevisiae, Pichia Pastoris, Spodoptera Frugiperda, and COS7 Cells for Recombinant Gene Expression Application to a Rabbit Liver Carboxylesterase*. Vol 16.; 2000.
7. La Clair JJ, Foley TL, Schegg TR, Regan CM, Burkart MD. Manipulation of Carrier Proteins in Antibiotic Biosynthesis. *Chem Biol*. 2004;11(2):195-201. doi:10.1016/j.chembiol.2004.02.010
8. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science (80-)*. 2001;291(5509):1790-1792. doi:10.1126/science.1058092
9. Morgan JG, Sukiennicki T, Pereira HA, Spitznagel JK, Guerra ME, Larrick JW. Cloning of the cDNA for the serine protease homolog CAP37/azurocidin, a microbicidal and chemotactic protein from human granulocytes. *J Immunol*. 1991;147(9):3210-3214. http://www.ncbi.nlm.nih.gov/pubmed/1919011. Accessed September 4, 2019.
10. Plasmids 101: Antibiotic Resistance Genes. https://blog.addgene.org/plasmids-101-everything-you-need-to-know-about-antibiotic-resistance-genes. Accessed December 17, 2019.
11. Heterologous Expression | Study.com. https://study.com/academy/lesson/heterologous-expression-chromosomes.html#/targetText=Heterologous gene expression is important,native stock of that crop. Accessed September 4, 2019.
12. Stevens DC, Hari TPA, Boddy CN. The role of transcription in heterologous expression of polyketides in bacterial hosts. *Nat Prod Rep*. 2013;30(11):1391-1411. doi:10.1039/c3np70060g
13. Nah HJ, Pyeon HR, Kang SH, Choi SS, Kim ES. Cloning and heterologous expression of a large-sized natural product biosynthetic gene cluster in Streptomyces species. *Front Microbiol*. 2017;8(MAR):394. doi:10.3389/fmicb.2017.00394
14. Ross AC, Gulland LES, Dorrestein PC, Moore BS. Targeted capture and heterologous

expression of the pseudoalteromonas alterochromide gene cluster in escherichia coli represents a promising natural product exploratory platform. *ACS Synth Biol*. 2015;4(4):414-420. doi:10.1021/sb500280q

15. Wösten MM. Eubacterial sigma-factors. *FEMS Microbiol Rev*. 1998;22(3):127-150. doi:10.1111/j.1574-6976.1998.tb00364.x
16. Stevens DC, Conway KR, Pearce N, Villegas-Peñaranda LR, Garza AG, Boddy CN. Alternative Sigma Factor Over-Expression Enables Heterologous Expression of a Type II Polyketide Biosynthetic Pathway in Escherichia coli. *PLoS One*. 2013;8(5). doi:10.1371/journal.pone.0064858
17. Luo X, Reiter MA, d'Espaux L, et al. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature*. 2019;567(7746):123-126. doi:10.1038/s41586-019-0978-9
18. Kohler R. The background to Eduard Buchner's discovery of cell-free fermentation. *J Hist Biol*. 1971;4:35-61. http://www.ncbi.nlm.nih.gov/pubmed/11609437. Accessed September 4, 2019.
19. Nirenberg M, Leder P, Bernfield M, et al. RNA codewords and protein synthesis, VII. On the general nature of the RNA code. *Proc Natl Acad Sci U S A*. 1965;53(5):1161-1168. doi:10.1073/pnas.53.5.1161
20. Deciphering the Genetic Code - National Historic Chemical Landmark - American Chemical Society. https://www.acs.org/content/acs/en/education/whatischemistry/landmarks/geneticcode.html. Accessed September 4, 2019.
21. Noireaux V, Libchaber A. *A Vesicle Bioreactor as a Step toward an Artificial Cell Assembly*.; 2004. www.pnas.orgcgidoi10.1073pnas.0408236101. Accessed September 4, 2019.
22. Gregorio NE, Levine MZ, Oza JP. A User's Guide to Cell-Free Protein Synthesis. *Methods Protoc*. 2019;2(1):24. doi:10.3390/mps2010024
23. Shimizu Y, Inoue A, Tomari Y, et al. Cell-free translation reconstituted with purified components. *Nat Biotechnol*. 2001;19(8):751-755. doi:10.1038/90802
24. PURExpress® *In vitro* Protein Synthesis Kit | NEB. https://international.neb.com/products/e6800-purexpress-invitro-protein-synthesis-kit#Product Information. Accessed September 4, 2019.
25. Pardee K, Slomovic S, Nguyen PQ, et al. Portable, On-Demand Biomolecular Manufacturing. *Cell*. 2016;167(1):248-259.e12. doi:10.1016/j.cell.2016.09.013
26. Goering AW, Li J, McClure RA, Thomson RJ, Jewett MC, Kelleher NL. *In vitro* reconstruction of nonribosomal peptide biosynthesis directly from DNA using cell-free protein synthesis. *ACS Synth Biol*. 2017;6(1):39-44. doi:10.1021/acssynbio.6b00160
27. Prasad C. Bioactive cyclic dipeptides. *Peptides*. 1995;16(1):151-164. doi:10.1016/0196-9781(94)00017-z
28. Nguyen PHB, Wu YY, Guo S, Murray RM. Design Space Exploration of the Violacein Pathway in Escherichia coli Based Cell-Free System. *bioRxiv*. January 2016:027656. doi:10.1101/027656
29. Goering AW, Li J, McClure RA, Thomson RJ, Jewett MC, Kelleher NL. *In vitro* Reconstruction of Nonribosomal Peptide Biosynthesis Directly from DNA Using Cell-Free

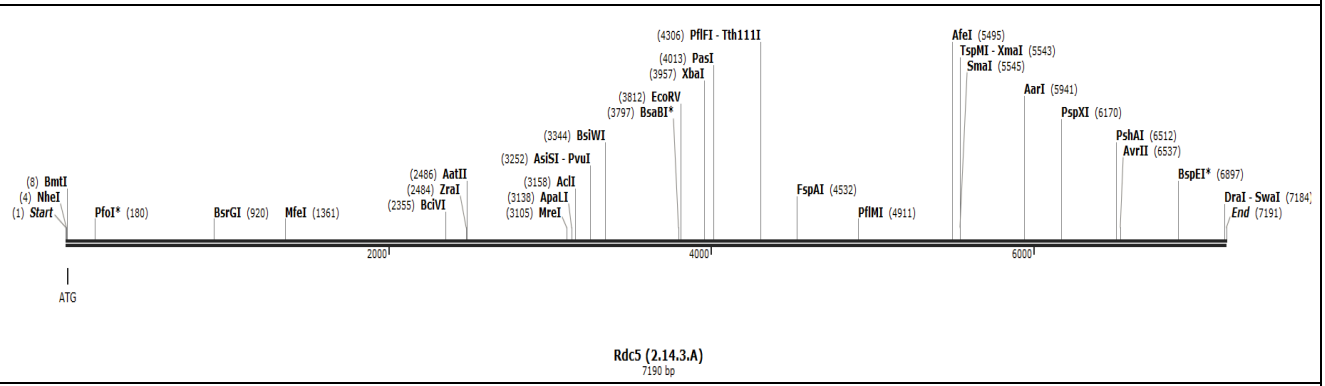
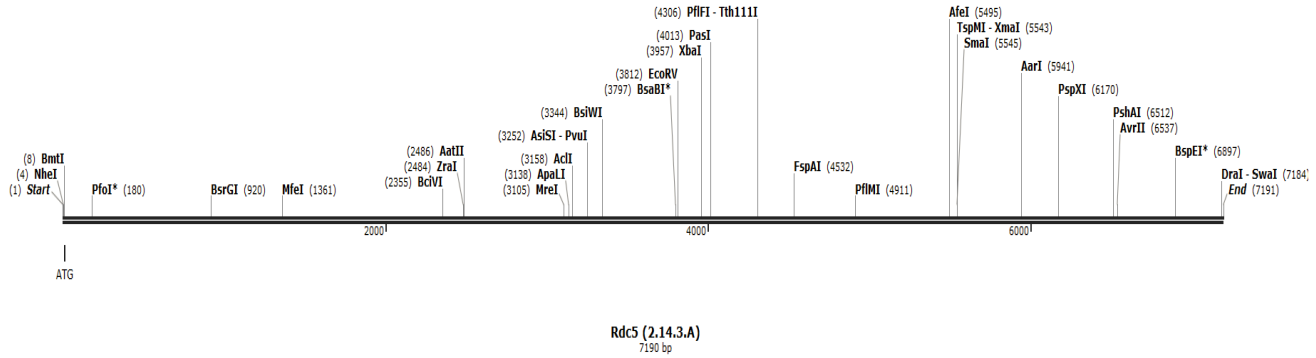
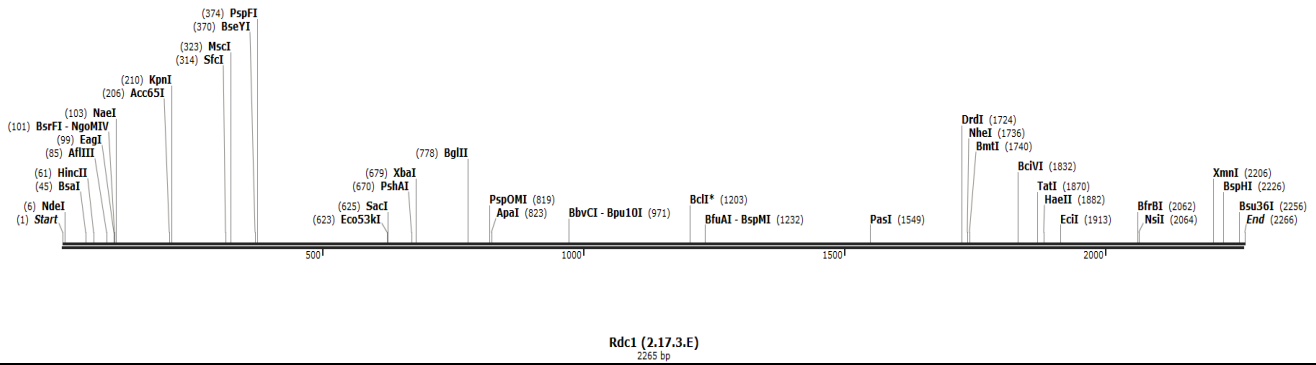
Protein Synthesis. *ACS Synth Biol*. 2017;6(1):39-44. doi:10.1021/acssynbio.6b00160
30. Shen B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol*. 2003;7(2):285-295. http://www.ncbi.nlm.nih.gov/pubmed/12714063. Accessed September 4, 2019.
31. Zhou H, Qiao K, Gao Z, Vederas JC, Tang Y. Insights into radicicol biosynthesis via heterologous synthesis of intermediates and analogs. *J Biol Chem*. 2010;285(53):41412-41421. doi:10.1074/jbc.M110.183574
32. Schopf FH, Biebl MM, Buchner J. The HSP90 chaperone machinery. *Nat Rev Mol Cell Biol*. 2017;18(6):345-360. doi:10.1038/nrm.2017.20
33. Quadri LEN, Weinreb PH, Lei M, Nakano MM, Zuber P, Walsh CT. Characterization of SFP, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carder protein domains in peptide synthetases. *Biochemistry*. 1998;37(6):1585-1595. doi:10.1021/bi9719861
34. Wang M, Zhou H, Wirz M, Tang Y, Boddy CN. A thioesterase from an iterative fungal polyketide synthase shows macrocyclization and cross coupling activity and may play a role in controlling iterative cycling through product offloading. *Biochemistry*. 2009;48(27):6288-6290. doi:10.1021/bi9009049
35. Xu W, Qiao K, Tang Y. Structural Analysis of Protein-Protein Interactions in Type I Polyketide Synthases. doi:10.3109/10409238.2012.745476
36. Fujii I. Heterologous expression systems for polyketide synthases. doi:10.1039/b817092b
37. Boddy CN, Hotta K, Tse ML, Watts RE, Khosla C. Precursor-directed biosynthesis of epothilone in Escherichia coli. *J Am Chem Soc*. 2004;126(24):7436-7437. doi:10.1021/ja048108s
38. Kumar S, Tsai C-J, Nussinov R. Factors enhancing protein thermostability. *Protein Eng Des Sel*. 2000;13(3):179-191. doi:10.1093/protein/13.3.179
39. BODIPY FL Dye - CA.
40. Kobs G, Hurst R, Betz N, Godat B. *FluoroTect ™ Green Lys in vitro Translation Labeling System*.
41. Kimple ME, Brill AL, Pasker RL. Overview of affinity tags for protein purification. *Curr Protoc Protein Sci*. 2013;(SUPPL.73). doi:10.1002/0471140864.ps0909s73
42. Commonly used auxotrophic markers - SGD-Wiki. https://wiki.yeastgenome.org/index.php/Commonly_used_auxotrophic_markers. Accessed September 4, 2019.
43. Lee KM, DaSilva NA. Evaluation of the Saccharomyces cerevisiae ADH2 promoter for protein synthesis. *Yeast*. 2005;22(6):431-440. doi:10.1002/yea.1221
44. Xue Y, Sherman - DH. *Review: Biosynthesis and Combinatorial Biosynthesis of Pikromycin-Related Macrolides in Streptomyces Venezuelae*. Vol 3.; 2001. www.idealibrary.com. Accessed December 8, 2019.
45. Booth WT, Schlachter CR, Pote S, et al. Impact of an N-terminal Polyhistidine Tag on Protein Thermal Stability. 2018. doi:10.1021/acsomega.7b01598
46. McCormick AM, Jarmusik NA, Endrizzi EJ, Leipzig ND. Expression, isolation, and purification of soluble and insoluble biotinylated proteins for nerve tissue regeneration. *J Vis Exp*. 2014;(83). doi:10.3791/51295
47. Kurland CG. Codon bias and gene expression. *FEBS Lett*. 1991;285(2):165-169. doi:10.1016/0014-5793(91)80797-7

Appendix

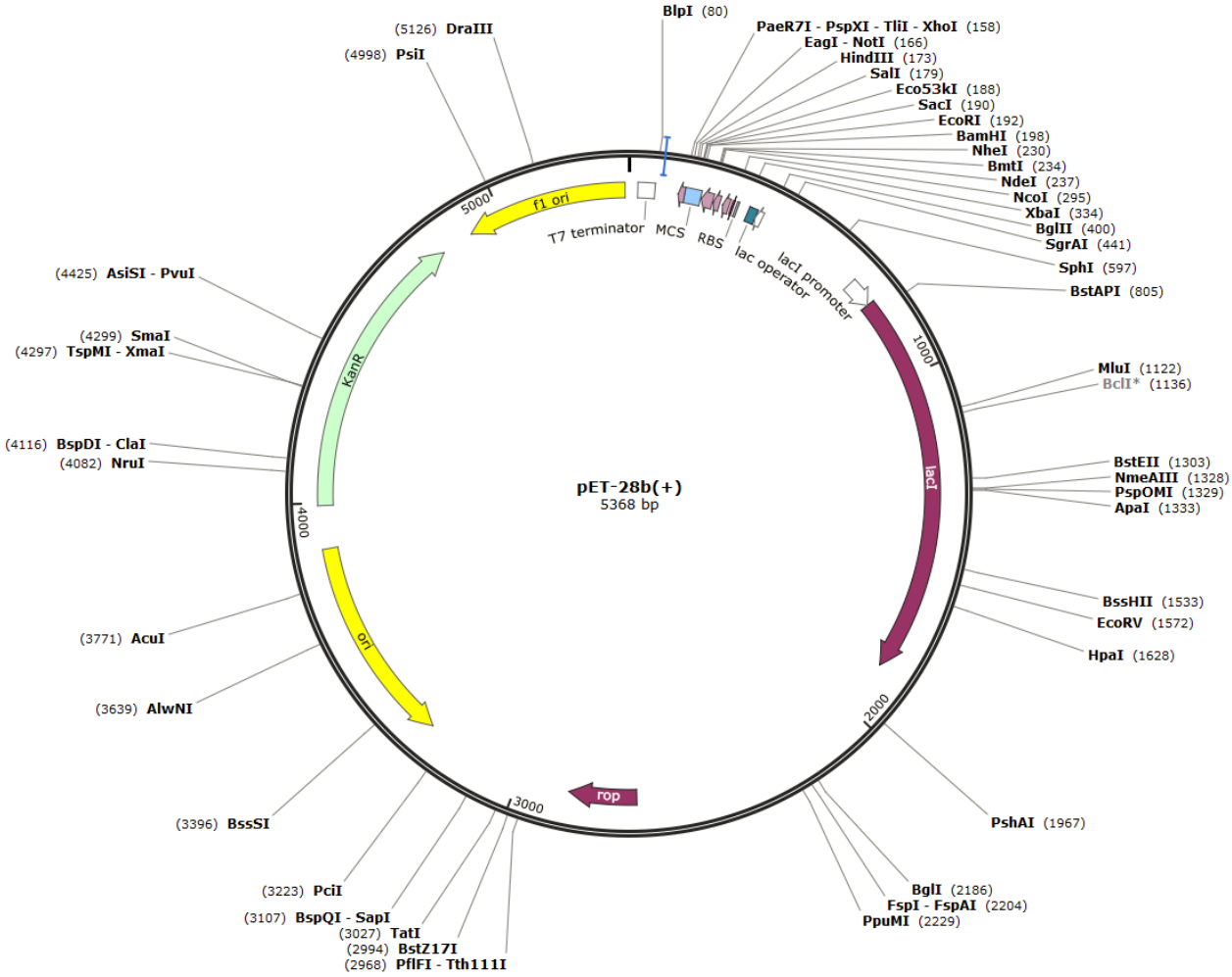
Primer List

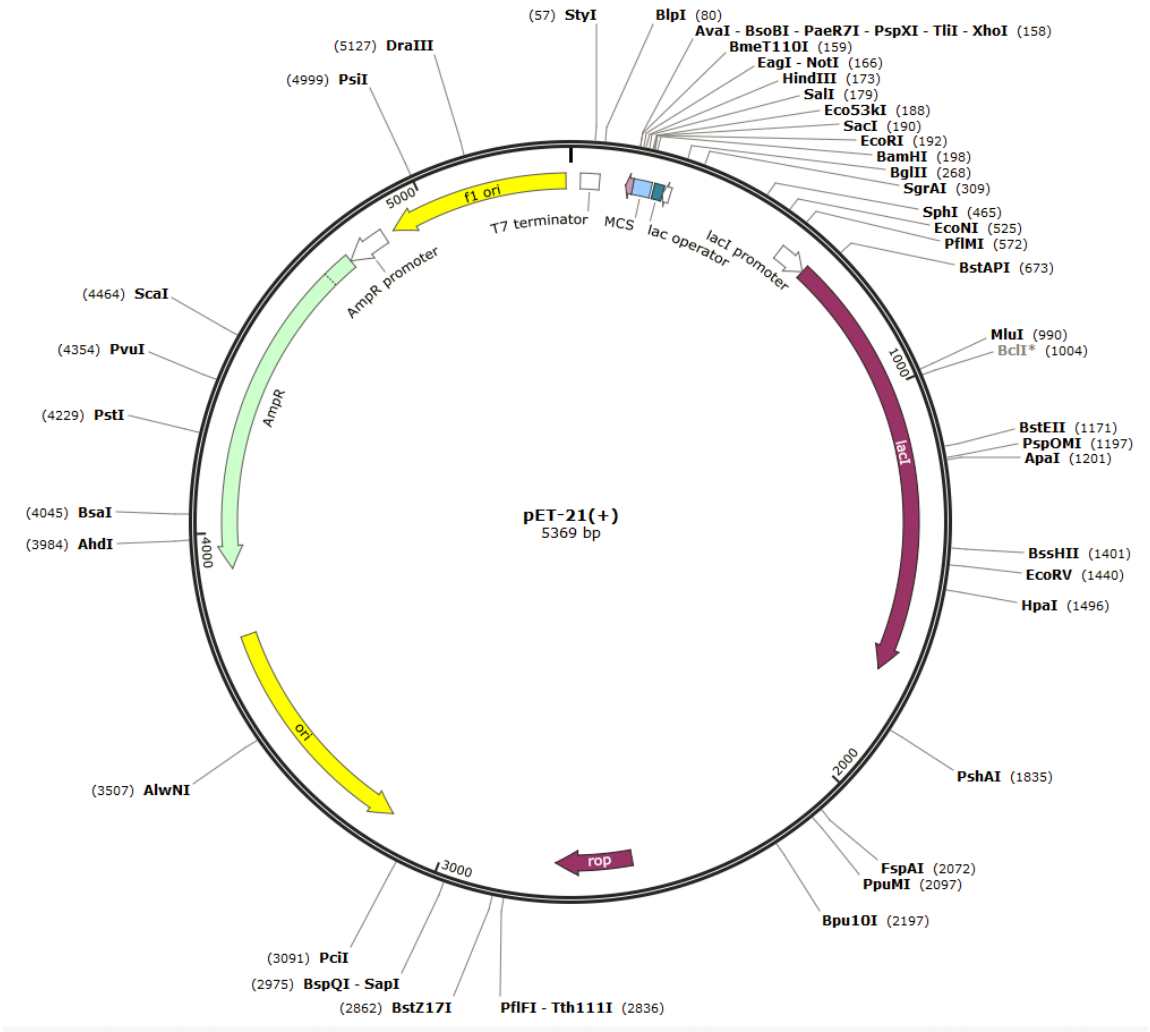
Protein (Section)	Forward Primer	Reverse Primer
Rdc5 (2.4.13.A)	AAG CTA GCA TGC CCT CCG CAA CTG CCC A	TTA TTT AAA TTC ATG TGG TGG TGG TGG TGG TGA GAA CCT CCC ACC AAC
Rdc5 (2.4.15.E)	AAG CTA GCA TGC CCT CCG CAA CTG CCC A	TTA TTT AAA TTC ATG TGG TGG TGG TGG TGG TGA GAA CCT CCC ACC AAC
Rdc1 (2.4.15.E)	CTT CAT ATG AAG TCG CAT ACT TCG GCT GCA	CCC GCC TGA GGT GGT ATC GTG

PCR Product Maps

Protein (Section)	Map
Rdc5 (2.4.13.A)	
Rdc5 (2.4.15.E)	
Rdc1 (2.4.15.E)	

Vector Maps


pET-28b(+)
5368 bp
BlpI (80)
PaeR7I - PspXI - TliI - XhoI (158)
EagI - NotI (166)
HindIII (173)
SalI (179)
Eco53kI (188)
SacI (190)
EcoRI (192)
BamHI (198)
NheI (230)
BmtI (234)
NdeI (237)
NcoI (295)
XbaI (334)
BglII (400)
SgrAI (441)
SphI (597)
BstAPI (805)
MluI (1122)
BclI* (1136)
BstEII (1303)
NmeAIII (1328)
PspOMI (1329)
ApaI (1333)
BssHII (1533)
EcoRV (1572)
HpaI (1628)
PshAI (1967)
BglI (2186)
FspI - FspAI (2204)
PpuMI (2229)
(2968) PflFI - Tth111I
(2994) BstZ17I
(3027) TatI
(3107) BspQI - SapI
(3223) PciI
(3396) BssSI
(3639) AlwNI
(3771) AcuI
(4082) NruI
(4116) BspDI - ClaI
(4297) TspMI - XmaI
(4299) SmaI
(4425) AsiSI - PvuI
(4998) PsiI
(5126) DraIII
f1 ori
T7 terminator
MCS
RBS
lac operator
lacI promoter
lacI
rop
ori
KanR
1000
2000
3000
4000
5000


pET-21(+)
5369 bp
(57) StyI
BlpI (80)
AvaI - BsoBI - PaeR7I - PspXI - TliI - XhoI (158)
BmeT110I (159)
EagI - NotI (166)
HindIII (173)
SalI (179)
Eco53kI (188)
SacI (190)
EcoRI (192)
BamHI (198)
BglII (268)
SgrAI (309)
SphI (465)
EcoNI (525)
PflMI (572)
BstAPI (673)
MluI (990)
BclI* (1004)
BstEII (1171)
PspOMI (1197)
ApaI (1201)
BssHII (1401)
EcoRV (1440)
HpaI (1496)
PshAI (1835)
FspAI (2072)
PpuMI (2097)
Bpu10I (2197)
PflFI - Tth111I (2836)
BstZ17I (2862)
BspQI - SapI (2975)
PciI (3091)
AlwNI (3507)
AhdI (3984)
BsaI (4045)
PstI (4229)
PvuI (4354)
ScaI (4464)
DraIII (5127)
PsiI (4999)
f1 ori
T7 terminator
MCS
lac operator
lacI promoter
AmpR promoter
AmpR
ori
rop
lacI
5000
1000
2000
3000
4000