EXPRESSION AND PURIFICATION OF POLYKETIDE SYNTHASE DOMAINS TO INVESTIGATE PROTEIN-PROTEIN INTERACTIONS IN POLYKETIDE BIOSYNTHETIC PATHWAYS

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Mujeeb Adeyinka Wakeel
August 2023

Copyright © 2023 by Mujeeb A. Wakeel
All rights reserved.

ABSTRACT

Polyketide synthases (PKSs) are multienzyme proteins catalyzing the production of the polyketide class of natural products with vast structural diversity. Type II PKSs catalyze polyketide biosynthesis using discrete (standalone) enzymes (domains), while type I PKSs use covalently linked enzymes arranged in a module to catalyze polyketide production. It is hypothesized that type II PKSs are the precursors to type I PKSs, proposing that the genes encoding individual discrete enzymes in type II PKSs fused to give the multienzyme modules of type I PKSs. We reasoned that the intramolecular-like reactions occurring in type I PKS modules might be a driving force for this evolution. This thesis reports the heterologous expression and purification of acyl carrier proteins (ACPs) and ketoreductase (KR) domains from the 6-deoxyerythronolide B synthase's (DEBS) first module and the amphotericin B synthase's second module as model enzymes to investigate this hypothesis. These enzymes were expressed as didomains (KR-ACP) and standalone domains in *E. coli* strains and purified by fast protein liquid chromatography (FPLC). The domain boundaries of the ACP domains were established to express these enzymes as soluble standalone domains. Chimeric didomains were also produced to evaluate the effect of protein-protein interactions in rate acceleration.

TABLE OF CONTENTS

CHAPTER ONE INTRODUCTION……...1
Polyketides……..1
Polyketide Synthases (PKSs)...……...1
Polyketide Synthase Classifications……...4
Investigation of Polyketide Synthase Modules and Domains……4
Engineering Opportunities with Polyketide Synthases……..8
Evolutionary Relationship Between Type I and II Polyketide Synthases……10
Ketoreductase Domains and This Work……...10
CHAPTER TWO MATERIALS AND METHODS……12
Experimental Techniques……..……...12
Polymerase Chain Reaction……12
Spectrophotometry…….……..12
DNA Cloning……...……12
Centrifugation……..14
Liquid Chromatography……...14
Electroporation……14
Sonication……....16
DNA Sequencing……16
Bradford Protein Assay……16
Software Analysis……....18
Experimental Procedures……..18
Cloning of DNA Constructs...……..18
Construction of pET15bAmpKR2ACP2 (1) ...……...18
Construction of pET15bAmpKR2EryACP1 (2) ...…….20
Construction of pET15bEryKR1AmpACP2 (3) …...…….20
Construction of pET15bAmpACP2 (4) …..……...20
Construction of pET15bEryACP1 (5) …..……..20
Protein Expression ..……....21
Protein Expression of AmpKR2ACP2 (6) …...……...21

Protein Expression of AmpKR2EryACP1 (7) …..…….21
Protein Expression of EryKR1AmpACP2 (8) …..…….22
Protein Expression of AmpKR2 (9) ……22
Protein Expression of Sfp (10) ……22
Protein Expression of AmpACP2 (11) ……22
Protein Expression of EryKR1 (12) ……23
Protein Expression of EryKR1ACP1 (13) …..……...23
Protein Expression of EryACP1 (14) ……...……...23
In vitro Conversion of *apo*-ACPs to Acetoacetyl-*holo*ACPs …...……..…….23
CHAPTER THREE RESULTS AND DISCUSSION ……..24
Constructs Design Strategy …….....24
Strategy for Constructing Chimeric Didomains …….....24
Strategy for Constructing Standalone AmpACP2 ……..24
Strategy for Constructing Standalone EryACP1 …….…....27
Protein Expression and Purification ……30
Purification of EryKR1AmpACP2 ……30
Purification of AmpKR2EryACP1 ……30
Purification of AmpACP2 ……...36
Purification of EryKR1ACP1 ……36
Purification of AmpKR2ACP2 ……...36
Purification of AmpKR2, EryKR1, and Sfp ……...39
Sfp-catalyzed Phosphopantetheinylation of *apo*-ACP Domains …….……39
Proposed NADPH-Dependent Kinetic Assay …….…….43
CHAPTER FOUR CONCLUSIONS ……47
REFERENCES ……48
APPENDIX ……...55
VITA ……67

LIST OF TABLES

Table 2.1. List of primers ……..19

Table 3.1. List of proteins biosynthesized and purified in this thesis …….……..34

Table 3.2. Phosphopantetheinylated ACP domains and KR-ACP didomains and their calculated molecular weight ……..…….42

LIST OF FIGURES

Figure 1.1. Examples of polyketides with their primary biological activities and producer organisms ……2

Figure 1.2. Schematic representation of an extension cycle in polyketide biosynthetic pathway……3

Figure 1.3. Phosphopantetheinyl transferase (PPTase)-catalyzed post-translational modification of *apo*-ACP (inactive) to *holo*-ACP (active) .....……..…….5

Figure 1.4. Representations of domain arrangements showing *intramolecular* interactions between adjacent domains in type I PKS and *intermolecular* interactions among domains in type II PKS ……..……5

Figure 1.5. Graphical representation of a complete type I PKS, 6-deoxyerythronolide B synthase (6-DEBS). Abbreviation: LDD, loading didomain; TE, thioesterase …….7

Figure 2.1. Schematic representation of a polymerase chain reaction …..……....13

Figure 2.2. Simplified structure of a spectrophotometer ……..13

Figure 2.3. Schematic representation of DNA cloning ……..……...15

Figure 2.4. Diagrammatic representation of Sanger sequencing technique ……..17

Figure 3.1. Schematic representation of the construction of chimeric didomains from native didomains ……..25

Figure 3.2. Model of AmpACP2 ……..……26

Figure 3.3. Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) of AmpACP2. Lane 3, lysate (11.5 KDa); lane 4, pellet (11.5 kDa); lane 5, lysate (10.4 kDa); and lane 6, pellet (10.4 kDa) ……...…….…...28

Figure 3.4. Model showing EryACP1 (white) linked to the N-terminal of EryKS2 (red) in the 182 aa sequence ….....…….....29

Figure 3.5. SDS-PAGE of EryACP1 (11.3 kDa). Lane 1, lysate; and lane 2, pellet ……...…….……29

Figure 3.6. Schematic representation of the biosynthesis of enzymes from their corresponding constructs (color selection is random) ……..31

Figure 3.7. SDS-PAGE of impure EryKR1AmpACP2 (70.8 Da). Lane 1, lysate; lanes 2-10, impure fractions; and lane 11, pellet ……..33

Figure 3.8. SDS-PAGE of EryKR1AmpACP2. Lanes 1-10, pure fractions of EryKR1AmpACP2 ……...….……....34
Figure 3.9. SDS-PAGE of impure AmpKR2EryACP1 (71.6 kDa). Lanes 1-13, impure fractions of AmpKR2EryACP1 ……..……..35
Figure 3.10. SDS-PAGE of AmpKR2EryACP1 (71.6 kDa). Lanes 1-11, pure fractions of AmpKR2EryACP1 ……37
Figure 3.11. SDS-PAGE of AmpACP2 (12.4 kDa). Lanes 1-13, pure fractions of AmpACP2 ……...…….37
Figure 3.12. SDS-PAGE of EryKR1ACP1 (71.6 kDa). Lanes 1-10, pure fractions of EryKR1ACP1 ……....……38
Figure 3.13. SDS-PAGE of AmpKR2ACP2 (70.7 kDa). Lanes 1and 2, purified fractions of AmpKR2ACP2 ……..……38
Figure 3.14. SDS-PAGE of EryKR1 (52.8 kDa). Lanes 1-12, pure fractions of EryKR1 ……40
Figure 3.15. SDS-PAGE of AmpKR2 (51.3 kDa). Lanes 1-11, pure fractions of AmpKR2 ……..……..40
Figure 3.16. SDS-PAGE of Sfp (28.2 kDa). Lanes 1-11, pure fractions of Sfp ……...…….41
Figure 3.17. Conversion of *apo*-ACPs to acetoacetyl-*holo*ACPs ……..41
Figure 3.18. MS^3 fragmentation method to detect acetoacetyl-*holo*ACPs ……44
Figure 3.19. Schematic representation of KR-catalyzed reduction of acetoacetyl-*holo*ACP in the presence of NADPH ……..44
Figure 3.20. Schematic representation of type I PKS-like (didomains) and type II PKS-like (standalone domains) reactions catalyzed by native domains ……45
Figure 3.21. Schematic representation of type I PKS-like (didomains) and type II PKS-like (standalone domains) reactions catalyzed by chimeric domains ……46

CHAPTER ONE
INTRODUCTION

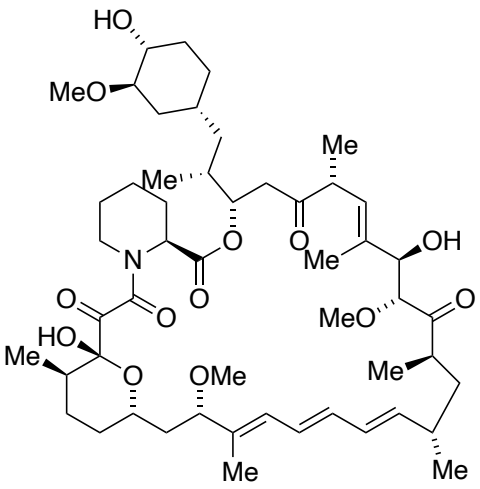
Polyketides

Bacteria, plants, and fungi produce a wide range of complex and structurally diverse polyketide natural products. Many of these compounds play crucial roles in various sectors, most especially the health sector.[1] Polyketides are secondary metabolites[2] synthesized from units of short acyl-CoA with the aid of complex proteins termed polyketide synthases (PKSs).[2,3] Many of these naturally produced compounds possess pharmacological properties, and some are applied as pharmaceuticals for the treatment of certain infections and diseases,[4] including the family of epothilone[5] used as antitumor agents, the yersiniabactin[6] used as iron chelator, the anticancer doxorubicin,[7] the immunosuppressant rapamycin[8] and FK506[4], the antibiotic erythromycin[9], and the potent antifungal and antibiotic polyene amphotericin B[10] (Figure 1.1).

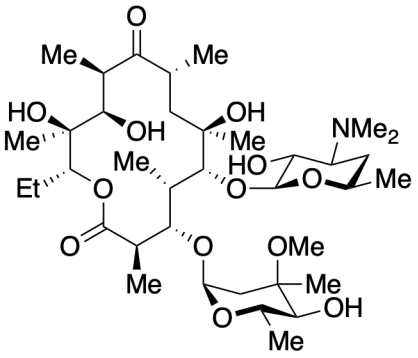
Polyketide Synthases (PKSs)

Polyketide synthases are large, multi-enzyme proteins that facilitate the biosynthesis of natural products belonging to the polyketide group.[9,11] A polyketide synthase (PKS) comprises a set of three essential enzymes (also called domains)[7], ketosynthase (KS), acyltransferase (AT), and acyl carrier protein (ACP), catalyzing a single round of polyketide chain elongation (Figure 1.2). A PKS may include β-carbon processing domains as optional components. These domains, which can include ketoreductase (KR), dehydratase (DH), and enoylreductase (ER), participate in modifying the newly added unit of a polyketide chain during growth.[11] Typically, a thioesterase (TE) facilitates the release of a growing polyketide chain from the synthase, and the released product may further undergo a series of post-PKS reactions such as oxidations, methylations, halogenations, and glycosylations.[3,12]

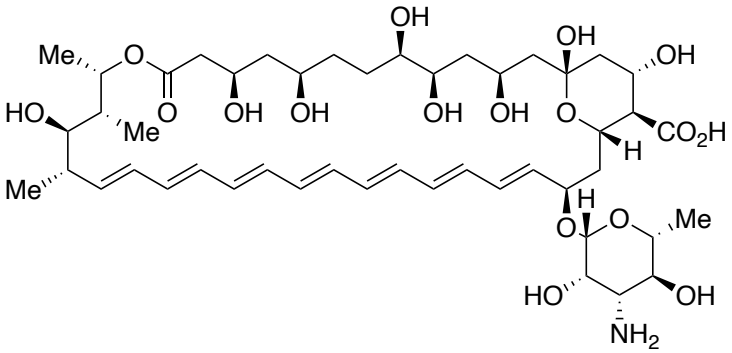
An AT domain acts as a monomer-selecting domain which selects the starter and extender units to be loaded on an ACP.[13,14] A ketosynthase (KS) enzyme facilitates a decarboxylative Claisen-like condensation reaction between a polyketide intermediate and a newly introduced extender unit bound to an acyl carrier protein (ACP).[1] An ACP is able to carry a growing polyketide chain when it is post-translationally modified to have the ~18 Å long



Rapamycin
immunosuppressant
Streptomyces hygroscopicus



Erythromycin
antibiotic
Saccharopolyspora erythraea



Amphotericin B
antibiotic
Streptomyces nodosus

Figure 1.1. Examples of polyketides with their primary biological activities and producer organisms.

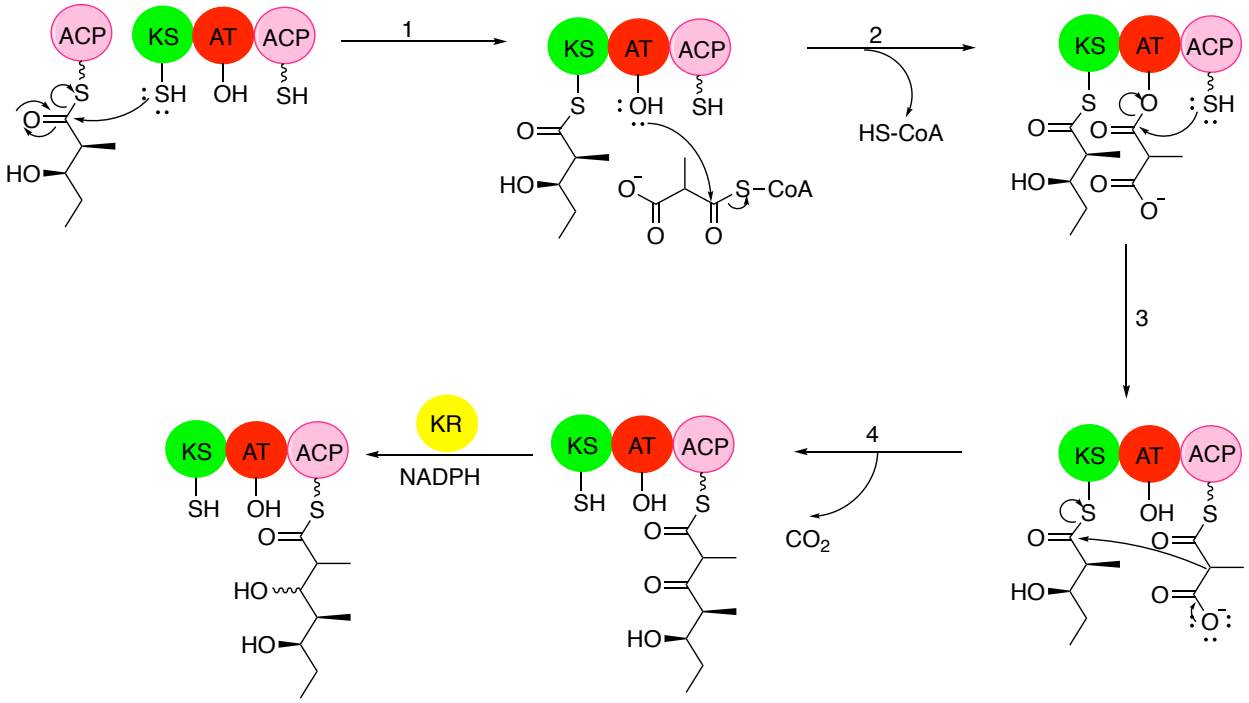


Figure 1.2. Schematic representation of an extension cycle in polyketide biosynthetic pathway.

phosphopantetheinyl (PPant) prosthetic group of coenzyme A (CoA) attached to its active site serine residue[3,15,16] (Figure 1.3). On both ends of an ACP domain are linkers, and these, together with the PPant arm, afford it adequate translational freedom to diffuse between active sites of nearby domains to deliver the polyketide chain.[3] The KR, DH, and ER domains are tailoring enzymes that catalyze the reduction of a growing polyketide chain's β-keto group to a hydroxyl group (-OH), dehydration of the hydroxyl group to an α,β-unsaturated thioester, and a further reduction of the double bond to a saturated, fatty acid-like carbon chain in that order.[3,17]

Polyketide Synthase Classifications

Polyketide synthases are generally classified into three main broad categories; type I, II, and III based on the structural arrangement of the functional domains present and the extender units' delivery method within the PKSs.[18] Type I PKSs are modular in nature and each module has a set of catalytic domains linked together to perform their individual enzymatic activity. Type I PKSs are further divided into non-iterative type I PKSs and iterative type I PKSs.[1819,20] The non-iterative type I PKSs are multimodular in nature, each module catalyzing a single round of extension and modification of a growing polyketide chain.[18,21,22] The iterative type I PKS (iPKS) is a single-module protein with a set of functional domains that catalyzes repeating chain extension and modification to produce a polyketide of a particular chain length.[23,24] Type II PKSs consist of distinct separable enzymes that function differently and iteratively to produce, mainly, multicyclic aromatic compounds (Figure 1.4). Type III PKSs biosynthesize aromatic polyketides (polyhydroxy phenols, chalcones, and stilbenes) at a single active site, utilizing acyl-CoA directly as substrates instead of depending on ACP-tethered extender units.[18,25,26]

Investigation of Polyketide Synthase Modules and Domains

Modular type I PKSs have received much attention due to their modular arrangement and tunability for combinatorial biosynthesis. Combinatorial biosynthesis is defined as the application of genetically modified natural products' biosynthetic pathways to biosynthesize structurally altered natural products or novel compounds.[11,12,27,28] Investigating polyketide synthase modules and individual domains comprises discovery of PKS biosynthetic gene clusters, engineering of host strains, site-directed mutagenesis, combinatorial biosynthesis, heterologous expression,

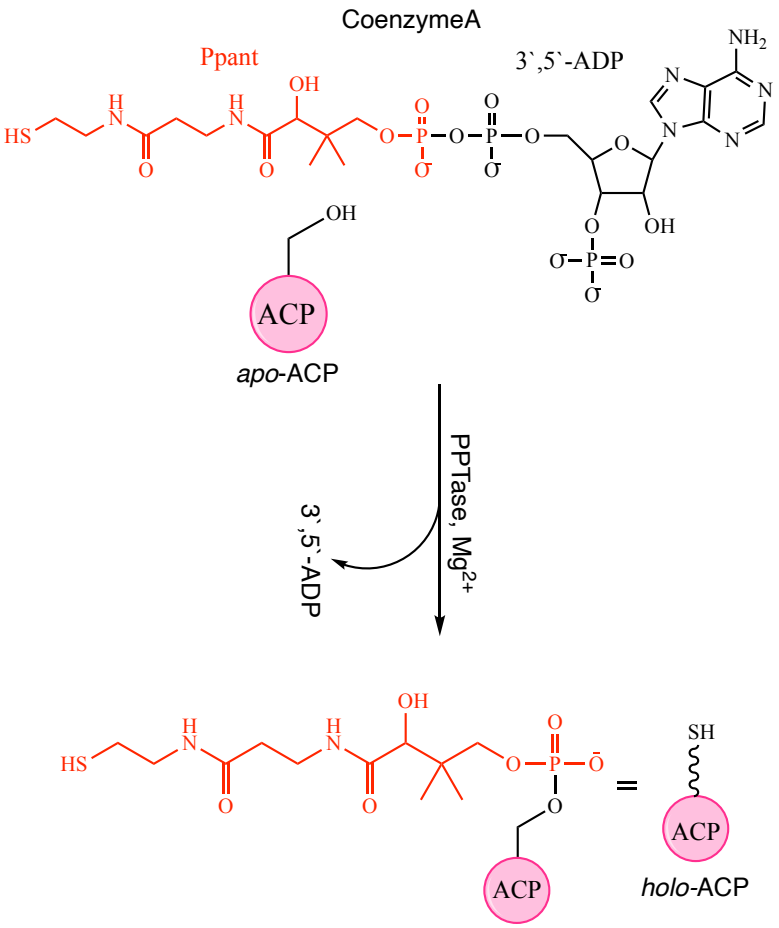


Figure 1.3. Phosphopantetheinyl transferase (PPTase)-catalyzed post-translational modification of *apo*-ACP (inactive) to *holo*-ACP (active).

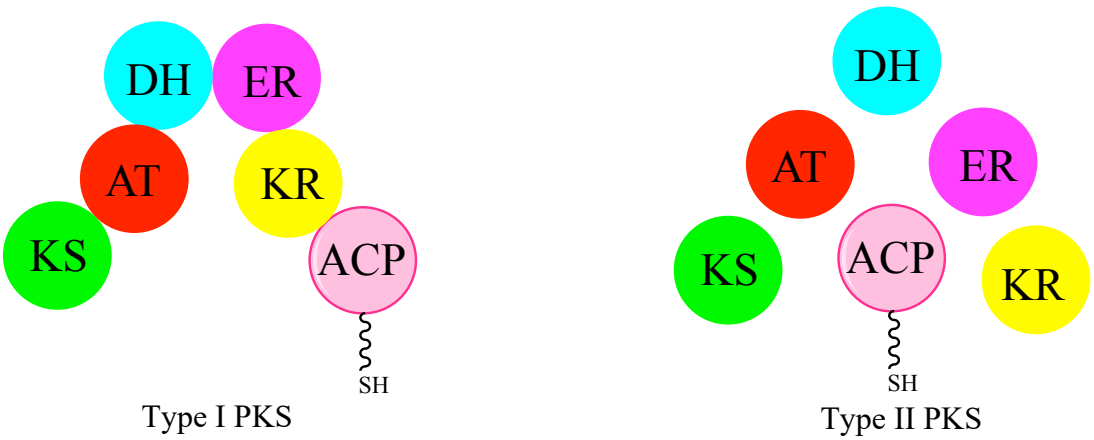


Figure 1.4. Representations of domain arrangements showing *intramolecular* interactions between adjacent domains in type I PKS and *intermolecular* interactions among domains in type II PKS.

development of high-throughput analytical methods, and advancement in technological tools for structural studies.[11,12,27–29]

Investigations into modules and domains from various type I PKSs have revealed information about factors that mediate catalysis such as protein-protein interactions between and within modules, active site residues, protein tertiary structures, and substrate specificity.[30–33] 6-DEBS (Figure 1.5) is a prototypical modular type I PKS whose studies have provided some details about the structural arrangements and protein-protein interactions within modular polyketide synthases. Three categories of protein linkers have been highlighted in 6-DEBS. The first kind of linkers are called *linker pairs,* existing at DEBS1 C-terminus and DEBS2 N-terminus and the analogous pair between DEBS2 and DEBS3. These linker pairs mediate in intermodular transfer of growing polyketide chain when successive modules are not covalently linked together, such as in modules 2-3 and the analogous modules 4-5. These linker pairs in DEBS, also called *docking domains*, are reported to be interchangeable without noticeable kinetic penalty.[34,35] The second category of protein linkers exists between an upstream acyl carrier protein and a downstream ketosynthase domain to facilitate intermodular transfer of growing polyketide chain between successive modules (e.g., modules 1-2, modules 3-4, and other corresponding successive modules). The third category of linkers are *interdomain linkers* that exist between domains within a PKS module (i.e., KS-AT-DH-ER-KR-ACP).[14]

Studies of the architectural features and mechanism of DEBS have also highlighted key amino acid residues, called *active site residues*, within each domain. For example, the conserved residues in DEBS KS, AT, and ACP domains are GH**S**xG, GPxxxxxTA**C**SS, and LGxD**S**LxxVE respectively (nucleophilic amino acid residues boldened). The KR and ER domains also have GxGxxAxxA motif as their active site for NADPH binding. The DEBS AT, KS, and ACP also have higher sequence similarity to their fatty acid synthase (FAS) counterpart domains,[14] Studies on DEBS also revealed that some PKS domains have substrate and partner domain specificity, while others are not as specific.[34] For example, KS domains in the second and sixth modules of DEBS (KS2 and KS6) can accept diketide substrates from the loading didomain (LDD) ACP and ACP2 without suitable linker pairs, whereas KS3 and KS5 are observed not capable of accepting diketide substrates without matching linker pairs.[34] High-resolution structures of protein domains and their ternary complexes have been deconvoluted using a combination of small-angle X-ray

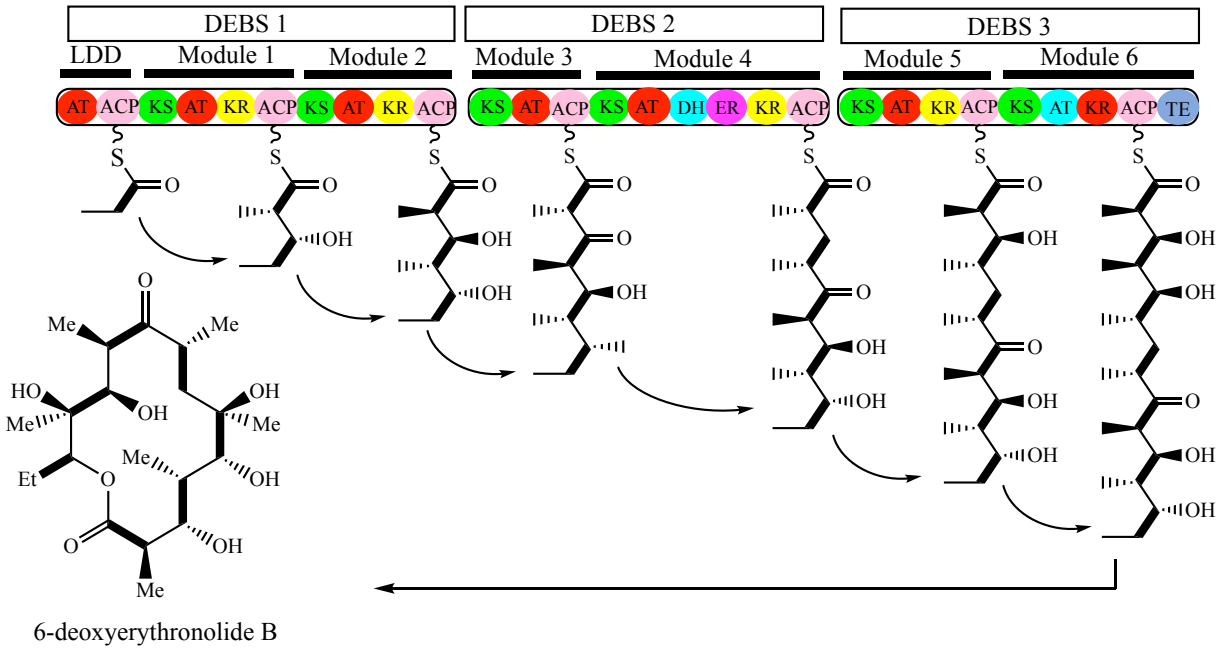


Figure 1.5. Graphical representation of a complete type I PKS, 6-deoxyerythronolide B synthase (6-DEBS). Abbreviation: LDD, loading didomain; TE, thioesterase.

scattering (SAXS), X-ray crystallography, and NMR spectroscopy. This approach has allowed for the identification of structural features and the analysis of protein-protein interactions.[7,36] Modules and domains from many other polyketide synthases have also been investigated, revealing key features such as active site residues and tertiary structures further supporting the hypothesis that domains of certain type have certain signature motif. These findings have made sequence alignment and structural elucidation become useful tools in recognizing domain types in PKS biosynthetic gene clusters.[1,10,36]

Engineering Opportunities with Polyketide Synthases

Discovery of polyketide natural products with useful medicinal properties has informed rational engineering of PKS biosynthetic pathways to increase the titers of pharmacologically active natural products and produce analogs with improved potency.[7] Availability of information about structural features, active site residues, and mechanism of catalysis of PKS domains has resulted into more rational design of chimeric modules tailored toward biosynthesis of structurally modified secondary metabolites. Subsets of certain PKSs (individual modules) have been utilized to produce "non-natural" products, supporting the hypothesis that individual PKS modules can be rationally designed and utilized to biosynthesize desired compounds of interest in the absence of other modules of their native PKS.[37]

It is hypothesized that the tandem ACPs in some polyketide synthases, such as batumin and macrolactin PKSs, increase the polyketide production rate of these PKSs. The tandem ACPs in polyunsaturated fatty acid (PUFA) synthases have been reported as the reason for the rare high production rate of the corresponding fatty acids by deep-sea bacteria. This inspired the design of artificial tandem ACPs in the module 6 of DEBS by triplicating the ACP naturally present in this module. This artificial tandem ACP was observed to have increased the yield of the polyketide product, a triketide ketolactone, by approximately 2.5 times fold in the *in vitro* study.[7]

Engineering of hybrid polyketide synthases from well characterized domains has also been hypothesized as promising alternative route to producing organic compounds that are not easily accessible using traditional methods of synthesis and commodity chemicals whose production generates toxic by-products.[38] In combinatorial biosynthesis, enzymes from different polyketide synthases can be combined to make chimeric modules to produce a particular compound of

interest. Adipic acid, a commodity chemical whose production results in production of N_2O – a greenhouse gas, has been produced by a chimeric module constructed by combining borrelidin PKS loading and first modules and a reductive loop (DH-ER-KR) consisting of borrelidin module 2 DH domain (BorDH2) and spinosyn B (SpnB) PKS's ER and KR domains. The production of adipic acid from an engineered PKS module has opened a potential non-toxic route of producing longer mono-carboxylic and dicarboxylic acids with desired substituent groups and level of oxidation using polyketide synthases.[38]

Heterologous expression platform has been extensively employed in expressing PKS biosynthetic gene clusters for combinatorial studies. This expression platform involves direct transfer of biosynthetic gene clusters into well characterized non-native hosts, such as *Saccharomyces cerevisiae* and *Escherichia coli*, capable of expressing foreign genes and culturable in laboratory settings. This strategy of expressing PKS genes has addressed some of the problems encountered in native hosts[38,39] and has created more synthetic opportunities to rapidly produce and screen polyketides derived from all classes of PKS. Heterologous expression has also fostered the development of chimeric modules via "design, build, and test" method in non-native hosts.[29,37]

However, expression of biosynthetic gene clusters in non-native hosts is also threatened with some challenges, cellular burden being a significant one, and these problems often result in poor maintenance of the biosynthetic gene clusters, resulting in poor expression. Advancements in manipulation of deoxyribonucleic acid (e.g., optimization of guanine-cytosine content), pathways cloning techniques,[40] and metabolic engineering (e.g., degradation of triacyglycerol product of primary metabolism in *Streptomyces* to increase titers of polyketides[41] and transformation of *E. coli* BL21 (DE3) strain with surfactin-producing (*sfp*) phosphopantetheinyl transferase gene present in *Bacillus subtilis* to make *E. coli* BAP1 strain[42]) continue to make heterologous expression more promising in engineering of polyketide synthases.[29]

Combinatorial biosynthesis has potential for production and optimization of compounds with promising pharmacological properties. However, more effort is required for this strategy to be accepted as a platform for making libraries of small molecules. This effort requires building a library of validated linkers, didomains, domains, and modules including their respective catalytic functions. This approach will guide decision-making during development of chimeric modules or

hybrid polyketide synthases for production of structurally altered natural products or novel compounds of interest on a large scale.[12,43]

Evolutionary Relationship Between Type I and II Polyketide Synthases

Type I PKSs are modular in nature, with a single module having its catalytic domains covalently linked to catalyze polyketide biosynthesis in an intramolecular-like reaction, while type II PKSs use discrete (standalone) domains to catalyze polyketide biosynthesis in an intermolecular-like fashion. It is hypothesized that type II PKSs are the precursors to type I PKSs, proposing that the genes encoding individual discrete domains in type II PKSs fused to give the multidomain modules of type I PKSs.[44] Modules of type I PKSs are also classified into *trans*-AT and *cis*-AT modules, *trans*-AT modules use a standalone AT domain to acylate their ACP domain with the extender unit[3] whereas *cis*-AT modules have their AT domain within the module. The existence of *trans*-AT modules supports the hypothesis that there is evolutionary relationship between type I and II polyketide synthases. The evolutionary driving force responsible for the formation of type I PKS multidomain modules from type II standalone domains despite cellular burden that may result from synthesizing large proteins remains unclear. The intramolecular-like reactions that occur within type I PKS module can be hypothesized to increase the rate of catalytic reactions. The ACP domain's proximity to other catalytic domains within a module allows swift delivery of its attached substrate to the active sites of those domains, resulting in an increased reaction rate. It is possible that the increase in rate of polyketide production compensates for the cellular burden of biosynthesizing a multidomain modular protein, hence, favoring the evolution of type I PKSs from type II PKSs. This does not preclude the idea that the need by the host organisms for the natural products biosynthesize by modular PKSs can also be responsible for this evolution.

Ketoreductase Domains and This Work

Ketoreductase (KR) domains have been a focus of engineering in altering the structures of polyketide natural products because these domains set majority of the stereocenters in polyketides. Structures of KR domain from the tylosin PKS' first module (TylKR1),[45] DEBS' first module,[46] and amphotericin PKS' second module[36] have been determined and investigated for their

mechanism of controlling stereochemistry of α-substituent and β-hydroxy group on growing polyketide chain.[47–49] A KR domain has NADPH co-factor binding site and utilizes hydride from the NADPH to reduce a β-keto group on a growing polyketide chain to a β-hydroxy group. It is hypothesized that all KR domains capable of reducing a β-keto group on a β-ketoacyl thioester intermediate have the NADPH co-factor bound to their binding site in the same orientation. An ACP domain, through its phosphopantetheine (PPant) arm, delivers a tethered β-ketoacyl thioester substrate to the catalytic groove (near the NADPH binding site) of a KR domain, and the stereochemistry of the resulting β-hydroxy group depends on which side the PPant tethered-substrate is guided into the catalytic groove.[36] The ability of many PKS domains to retain their enzymatic activity as isolated domain offers a great advantage to explore their engineering opportunities.[47,49,50] This thesis presented a work that will utilize heterologous expression platform to biosynthesize DEBS KR1-ACP1 didomain, standalone EryKR1, standalone EryACP1, amphotericin PKS KR2-ACP2 didomain, standalone AmpKR2, standalone AmpACP2, chimeric EryKR1-AmpACP2 didomain, and chimeric AmpKR2-EryACP1 didomain. These recombinant proteins will subsequently be used as models in NADPH dependent kinetic assay to establish kinetic relationships between intramolecular-like (didomain) and intermolecular-like (standalone domains) reactions in polyketide synthases to gain more understanding on the evolutionary relationship between type I and II PKSs. The proteins will also be investigated for their protein-protein interactions with non-native partner domains, essential for rational development of chimeric modules.

CHAPTER TWO
MATERIALS AND METHODS

Experimental Techniques

Polymerase Chain Reaction (PCR)

Polymerase chain reaction is a widely used method to rapidly produce many copies of a specific DNA region *in vitro*. This technique uses a thermostable DNA, a DNA template, the four DNA nucleotides (adenine, thymine, guanine, and cytosine), and two single-stranded DNA of short nucleotide sequences called primers. Primers are designed to specifically bind to a region of a single-stranded DNA to initiate a PCR. A polymerase chain reaction uses a repeating increasing and decreasing temperature cycle to multiply a DNA segment. The work reported in this thesis used the polymerase chain reaction for DNA amplification (Figure 2.1).

Spectrophotometry

Nucleic acids contain purine and pyrimidine backbones with conjugated double bonds having maximum absorption peak at 260 nm that be measured by a spectrophotometer. A spectrophotometer consists of a light source, a lens, a sample space, a wavelength selector, a photodetector, and a digital display (Figure 2.2). The work presented in this thesis used the spectrophotometry technique to monitor cell growth (OD_{600} measurement) and measure DNA and protein concentration.

DNA Cloning

DNA cloning is a widely adopted technique to make multiple identical copies of a specific DNA of interest *in vivo* using a cloning vector called plasmid. A plasmid is a circular piece of double-stranded DNA capable of replicating independently of the chromosomes. Plasmids naturally exist in bacteria and are sometimes found in archaea and eukaryotes such as plants. Artificial plasmids used in a laboratory typically contain a cloning site, a selection marker, a promoter region, and an origin of replication. Inserting DNA of interest into a plasmid can be achieved using any of the various available cloning methods, such as Gibson assembly, Golden Gate assembly, and restriction enzymes. The Gibson assembly method, developed by Gibson and

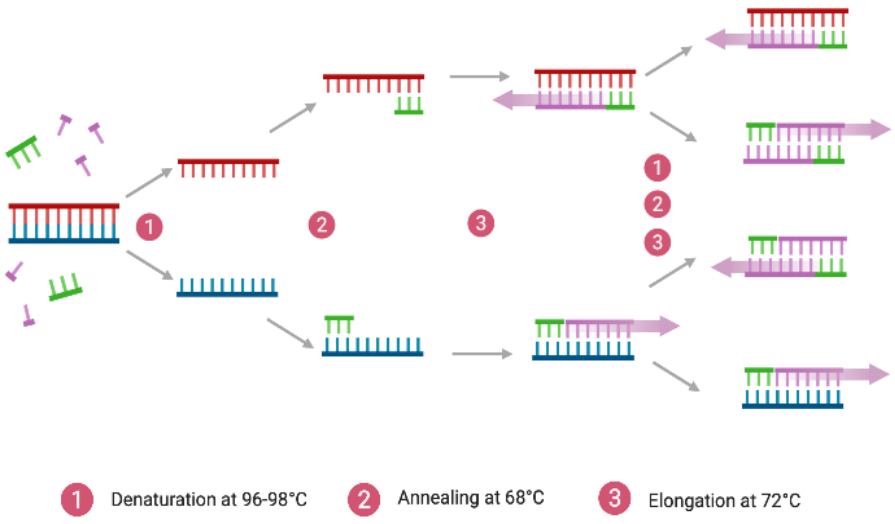


Figure 2.1. Schematic representation of a polymerase chain reaction.[51]

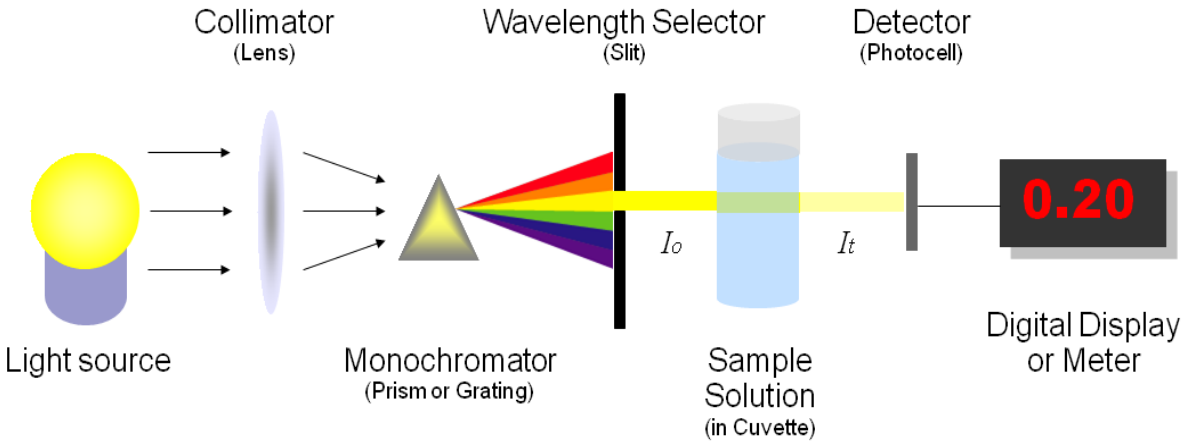


Figure 2.2. Simplified structure of a spectrophotometer.[52]

coworkers in 2009, was used in this work because of its robustness in assembling many DNA fragments with long overlap ends in a single isothermal reaction[53] (Figure 2.3).

Centrifugation

Centrifugation is a process that separates mixtures with the use of centrifugal force. Centrifugation is commonly used to recover cells from culture media and separate cell lysates from disrupted cell membranes and inclusion bodies during extraction and purification of macromolecules. This work used centrifugation to collect cells from liquid media and separate cell lysates from cellular debris.

Liquid Chromatography

Liquid chromatography is an analytical technique that is used to separate a mixture into its components. The separation occurs based on the interactions of the individual components of the mixture with the mobile and stationary phases. Fast Protein Liquid Chromatography (FPLC) is a high-performance liquid chromatography developed to purify protein macromolecules. It is compatible with common modes of chromatography, such as affinity, ion exchange, and gel filtration. The recombinant proteins biosynthesized in this work contained polyhistidine-tag (6xHis-tag) to their N-terminal and were purified by the FPLC technique on AKTA start systems using immobilized metal ion affinity chromatography (IMAC). An AKTA start system contains two buffer valves, a mixer, a pump, a pressure sensor, a wash valve, an injection valve, an outlet valve, a conductivity monitor, a UV detector, and a fraction collector. The system uses the UNICORN start software, allowing automation of the purification process.

Electroporation

Electroporation uses electrical pulses to create pores on cell membranes and transfect cells with genetic particles in a process called transformation. This method is an efficient way of introducing exogenous DNA into electrocompetent cells for heterologous expression of the DNA. Applying an electric field to cells is believed to increase the permeability of the cell membranes by creating tiny dimensional pores in the membranes, allowing the introduction of genetic materials into the cells. This work used electroporation to introduce plasmid DNA constructs into electrocompetent

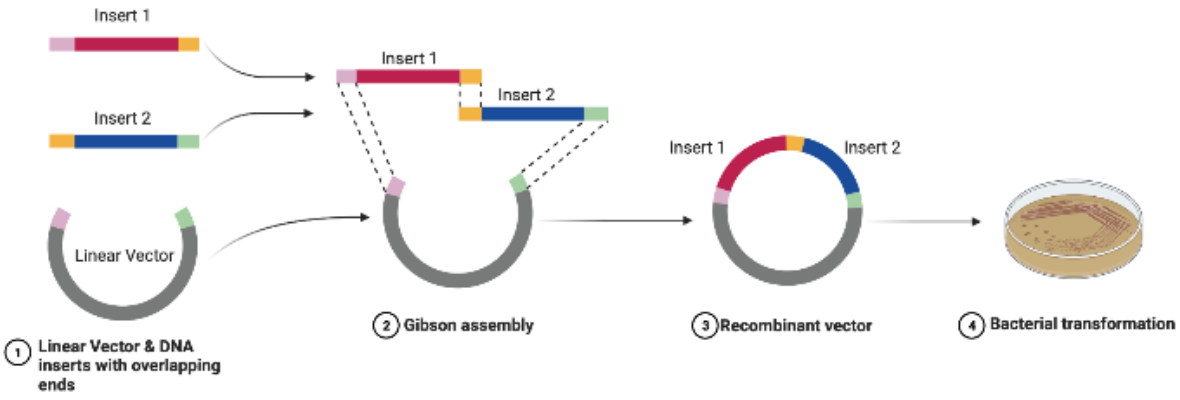


Figure 2.3. Schematic representation of DNA cloning (created in **BioRender.com**).

E. coli strains for DNA cloning and heterologous expression.

Sonication

Sonication is a technique that uses sound waves generated from electrical energy to agitate particles in a solution. This technique applies in various ways, such as phytochemical extraction, cell disruption, nanoparticle dispersion, liposome production, and homogenization. This method is used in chemical biology laboratories to disrupt cell membranes to release cellular contents. Although lysis buffers can disrupt cell membranes, this does not usually result in complete lysis of the cell membranes and often require further physical disruption, such as sonication for complete cell lysis. The sonication technique was used in this work to rupture cell membranes and release cell lysates for protein extraction and purification.

DNA Sequencing

A DNA molecule naturally consists of four nucleotides or bases, namely, thymine (T), guanine (G), adenine (A), and cytosine (C). DNA sequencing refers to the determination of the exact nucleotide sequence of a DNA macromolecule. One of the most used traditional methods of sequencing DNA molecules is Sanger sequencing (Figure 2.4), also called the chain termination method, developed by Fred Sanger and colleagues in 1977.[54] Sanger sequencing is a PCR-like reaction capable of sequencing up to 1000 base pairs of DNA fragments. More advanced technological tools (e.g., Oxford Nanopore Technologies) are also available for rapid DNA reading at low cost. This work used the Sanger sequencing at the UT Genomics Core and the Plasmidsaurus service to sequence DNA constructs.

Bradford Protein Assay

Bradford protein assay, which was developed by Marion Bradford in 1976,[55] is a colorimetric method for determining protein concentration in a sample. This work used the Bradford protein assay to quantify all the recombinant proteins biosynthesized. The technique is based on the binding of Coomassie Brilliant Blue G-250 to protein molecules under acidic conditions, leading to a change in the absorption maximum of the Coomassie Brilliant Blue G-250 from 465 nm to 595 nm. The Coomassie Brilliant Blue G-250 changes from its reddish/brown

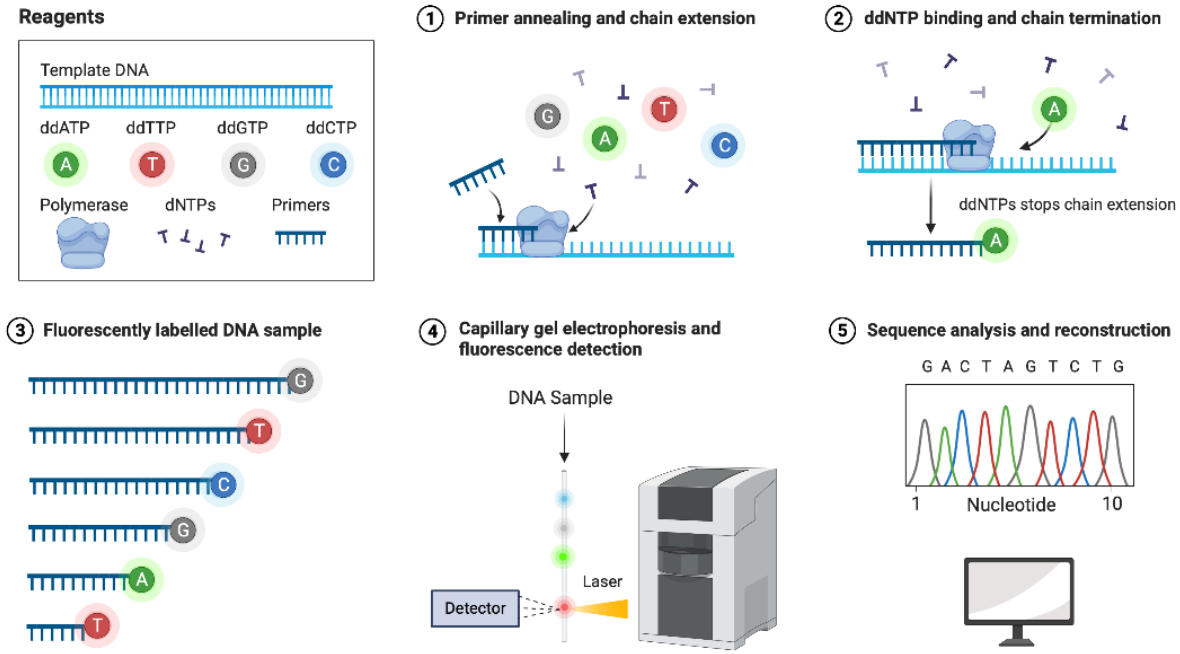


Figure 2.4. Diagrammatic representation of Sanger sequencing technique.[56]

color to blue upon binding to protein macromolecules, forming a protein-dye complex. The protein-dye complex formation can complete within 5 min, and the intensity of the resulting color is directly proportional to protein concentration. The little or no interference by reagents and ions and the high extinction coefficient of the protein-dye complex enable excellent sensitivity in determining protein concentration using this technique.

Software Analysis

Integrated DNA Technologies (IDT)[57] codon optimization tool was used to reduce the guanine-cytosine (GC) content of DNA nucleotide sequences for better protein expression in *E. coli* (a low GC content organism). DIVA DNA Construction program[58] was used to design recombinant plasmid DNA for DNA assembly of fragments. Clustal Omega program[59] was used to perform sequence analysis of designed and cloned constructs. Basic Local Alignment Search Tool (BLAST)[60] was used to confirm the nucleotide and protein sequences of constructs obtained from lab members. IntFOLD Integrated Protein Structure and Function Prediction Server (version 7.0)[61,62] was used for the prediction of domain boundaries.

Experimental Procedures

Cloning of DNA Constructs

Construction of pET15bAmpKR2ACP2 (1)

Circular pET15bAmpKR2 plasmid DNA was linearized and amplified by PCR using primers 1 and 2 (Table 2.1) in a 25 μL reaction (0.132 ng/μL pET15bAmpKR2, 0.5 μM each primer, 0.5 μL dimethyl sulfoxide (DMSO), and 12.5 μL NEBNext Q5 Hot Start HiFi PCR Master Mix). *E. coli* codon-optimized synthetic AmpACP2 DNA was amplified using primers 3 and 4. Analysis of PCR products was performed using DNA agarose gel electrophoresis, followed by gel DNA recovery using Zymoclean gel DNA recovery kit (Zymoresearch, USA). Amplified pET15bAmpKR2 and AmpACP2 fragments were assembled in a 20 μL Gibson reaction (10 μL 2x Gibson assembly master mix (New England Biolabs, MA, USA), 96 ng pET15bAmpKR2, and 288 ng AmpACP2, incubated at 50 ^{o}C for 2 hr). 50 μL DH5α electrocompetent *E. coli* strain was transformed (2000 V, Eppendorf eporator) with 2 μL of the assembly reaction product and incubated at 37 ^{o}C overnight on Luria broth (LB) agar plate with 100 μg/mL carbenicillin

Table 2.1. List of primers

Primer number	**Nucleotide sequence (**5′-3′)
1	CCGAGAATTCCGCCAATTAACTCGAGGTCTTTGAGGACCAGG
2	CCATCAGCGGGACCTTCATCGGACAGCGCACTTACGGC
3	GCTGTCCGATGAAGGTCCCGCTGATGGAGGCGC
4	CCTCAAAGACCTCGAGTTAATTGGCGGAATTCTCGGCAGC
5	CGTTGAGATCCGGCTGCTAACTCGAGGTCTTTGAGGACCAGGAAGTTTGAGG
6	GCAGCGACGCCAGACGGGCACGCAGTGGTGGT
7	ACTGCGTGCCCGTCTGGCGTCGCTGCCCGCG
8	TCAAAGACCTCGAGTTAGCAGCCGGATCTCAACGCCG
9	GCGGGTGGGCGCGCTTGACGGATTGCCTACCGCCGAAC
10	CGACCTCGGTGGAATGGCTGCCGCGCGGCAC
11	GCGCGGCAGCCATTCCACCGAGGTCGACGAGGTTTCC
12	GCAATCCGTCAAGCGCGCCCACCCGCGGTTC
13	CCGCTCCAGATGAGTAACATATGCTCGAGGATCCGGCTGCTAACAAAG
14	CACGCAGTGGTGGGCTGCCGCGCGGCACCAG
15	GCCGCGCGGCAGCCCACCACTGCGTGCCCGT
16	CCGGATCCTCGAGCATATGTTACTCATCTGGAGCGGTCCCGT
17	AACTCGGCGGTTAACATATGCTCGAGGATCCGGCTGCTAACAAAG
18	CGACGCCAGCGCGCTGCCGCGCGGCACCAG
19	TGCCGCGCGGCAGCGCGCTGGCGTCGCTGCCC
20	ATCCTCGAGCATATGTTAACCGCCGAGTTCGGCG

antibiotic. Two colonies were observed on the plate in less than 24 hr, and circular pET15bAmpKR2ACP2 plasmid DNA construct was extracted from the cells using ZymoPURE plasmid DNA miniprep kit (Zymoresearch, USA).

Construction of pET15bAmpKR2EryACP1 (2)

pET15bAmpKR2 was amplified from **1** using primers 5 and 6, while EryACP1 was amplified out of pET28aEryACP1 using primers 7 and 8. Amplified pET15bAmpKR2 (100 ng) and EryACP1 (500 ng) fragments were assembled and cloned as reported for **1**. One colony was observed in less than 24 hr, and DNA extraction was carried out as described for **1**.

Construction of pET15bEryKR1AmpACP2 (3)

pET15bAmpACP2 was amplified from **1** using primers 9 and 10, while EryKR1 was amplified from pET28aEryKR1ACP1 using primers 11 and 12. Amplified pET15bAmpACP2 (98 ng) and EryKR1 (400 ng) fragments were assembled and cloned as reported for **1**. Few colonies were observed in less than 24 hr, and DNA extraction was carried out as described for **1**.

Construction of pET15bAmpACP2 (4)

pET15b was amplified out of pET15bSfp using primers 13 and 14, while AmpACP2 was amplified from *E. coli* codon-optimized AmpACP2 using primers 15 and 16. Amplified pET15b (100 ng) and AmpACP2 (600 ng) fragments were assembled and cloned as reported for **1**. Two colonies were observed in less than 24 hr, and DNA extraction was carried out as described for **1**.

Construction of pET15bEryACP1 (5)

pET15b was amplified from 4 using primers 17 and 18, while EryACP1 was amplified from pET28aEryACP1 using primers 19 and 20. Amplified fragments were assembled and cloned as reported for **1**. Two colonies were observed in less than 24 hr, and DNA extraction was carried out as described for **1**.

Protein Expression

Protein Expression of AmpKR2ACP2 (6)

50 µL BL21 (DE3) electrocompetent *E. coli* strain was transformed (2000 V, Eppendorf eporator) with **1** (2 µL) and incubated at 37 ºC overnight on LB agar plate with 100 µg/mL carbenicillin antibiotic. Few colonies from the plate were used to grow (37 ºC, 250 rpm) an overnight culture in 23 mL LB medium with 100 µg/mL carbenicillin antibiotic. 10 mL overnight culture was used to grow a 2 L culture (37 ºC, 250 rpm) to an OD_{600} ~ 0.6 (measured on SpectraMax ABS Plus spectrophotometer using quartz cuvette). The cells were allowed to cool to ~16 ºC and induced for protein expression with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) for ~20 hr with constant shaking (250 rpm). Cells were collected by centrifugation (4000 rpm, 40 min), resuspended in 10 mL wash buffer (50 mM Na_2HPO_4, 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM Imidazole, and 5% glycerol, pH 7.64) and 5 mL HEPES buffer (50 mM HEPES, 5% glycerol, pH 7.64), and sonicated (20 sec on, 30 sec off, 5 times, OMNI Sonic Ruptor 400 Ultrasonic Homogenizer) on ice, followed by removal of cellular debris by centrifugation (10,000 rpm, 40 min). Cell lysate was loaded on a 5 mL Ni Sepharose High-Performance affinity resin column for purification of AmpKR2ACP2. The column was washed with 28 column volume (cv) of wash buffer after which AmpKR2ACP2 was eluted with 3 column volume of elution buffer (50 mM Na_2HPO_4, 50 mM NaH_2PO_4, 300 mM NaCl, 200 mM Imidazole, and 5% glycerol, pH 7.64) in gradient elution.

Protein Expression of AmpKR2EryACP1 (7)

50 µL BL21 Star (DE3) electrocompetent *E. coli* strain was transformed with **2** (2 µL) as reported for **6**. An overnight culture was grown (37 ºC, 250 rpm) in LB medium with 100 µg/mL carbenicillin antibiotic. 2 L LB medium containing 100 µg/mL carbenicillin antibiotic was inoculated with the overnight culture and grown (37 ºC, 250 rpm) to an OD_{600} ~ 0.65 (measured on SpectraMax ABS Plus spectrophotometer using quartz cuvette). The cells were cooled to ~16 ºC and induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) for ~20 hr with constant shaking at 250 rpm. Cells were collected by centrifugation (4000 rpm, 40 min), resuspended in 15 mL wash buffer (same composition as in 6), sonicated and purified as described for **6**.

Protein Expression of EryKR1AmpACP2 (8)

50 μL BL21 Star (DE3) electrocompetent *E. coli* strain was transformed with **3** (2 μL) as described for **6**. An overnight culture was grown (37 ºC, 250 rpm) in LB medium with 100 μg/mL carbenicillin antibiotic and used to inoculate 2 L LB medium, grown (37 ºC, 250 rpm) to an OD_{600} ~ 0.73 (measured on SpectraMax ABS Plus spectrophotometer using quartz cuvette). The cells were allowed to cool to ~ 16 ºC and induced with 0.2 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) for ~20 hr, constantly shaking at 250 rpm. The remaining procedure was as described for **7**.

Protein Expression of AmpKR2 (9)

50 μL BL21 (DE3) electrocompetent *E. coli* strain was transformed with 2 μL pET15bAmpKR2 DNA construct. The remaining procedure was as described for **6**.

Protein Expression of Sfp (10)

50 μL BL21 (DE3) electrocompetent *E. coli* strain was transformed with 2 μL pET15bSfp DNA construct. The remaining procedure was as described for **6** except that the cell resuspension was in 15 mL HEPES buffer (50 mM HEPES, 5% glycerol, pH 7.64), and the column was washed with 25 column volume of wash buffer.

Protein Expression of AmpACP2 (11)

50 μL BL21 (DE3) electrocompetent *E. coli* strain was transformed with **4** (2 μL) as described for **6**. An overnight culture was grown (37 ºC, 250 rpm) in LB medium with 100 μg/mL carbenicillin antibiotic and used to inoculate 4 L LB medium, grown (37 ºC, 250 rpm) in batches of 1 L to an OD_{600} ~ 0.65 (measured on SpectraMax ABS Plus spectrophotometer using quartz cuvette). The cells were cooled to ~16 ºC and induced with 0.2 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) for ~20 hr with constant shaking (250 rpm). Cells were collected by centrifugation (4000 rpm, 40 min), resuspended in 10 mL (per liter of culture) wash buffer (same as above), and sonicated (15 sec on, 30 sec off, 5 times/L) on ice, followed by removal of cellular debris by centrifugation (10,000 rpm, 40 min). Cell lysate/supernatant loaded on a 5 mL Ni Sepharose High-Performance affinity resin column. 15 column volume (cv) of wash buffer was

used to wash the column, after which AmpACP2 was eluted with 3 column volume of elution buffer in gradient elution. Eluted fractions were further purified, washing the column with 6 column volume (cv) of wash buffer.

Protein Expression of EryKR1 (12)

50 µL BL21 (DE3) electrocompetent *E. coli* strain was transformed with 2 µL pET28aEryKR1 DNA construct and grown on LB agar plate with 50 µg/mL kanamycin antibiotic at 37 ºC overnight. An overnight culture was grown (37 ºC, 250 rpm) in LB medium with 50 µg/mL kanamycin antibiotic and used to grow 2 L culture (37 ºC, 250 rpm) to an OD_{600} ~ 0.6 (measured on SpectraMax ABS Plus spectrophotometer using quartz cuvette). The remaining procedure was as described for **6**.

Protein Expression of EryKR1ACP1 (13)

50 µL BL21 Star (DE3) electrocompetent *E. coli* strain was transformed with 2 µL pET28aEryKR1ACP1 DNA construct and grown on LB agar plate with 50 µg/mL kanamycin antibiotic at 37 ºC overnight. An overnight culture was grown (37 ºC, 250 rpm) in LB medium with 50 µg/mL kanamycin antibiotic and used to inoculate 2 L LB medium, grown (37 ºC, 250 rpm) in batches of 1 L to an OD_{600} ~ 0.8 (measured on SpectraMax QuickDrop spectrophotometer using quartz cuvette). The cells were cooled to ~16 ºC and induced with 0.2 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) for ~20 hr with constant shaking (250 rpm). The remaining procedure was as described for **6**.

Protein Expression of EryACP1 (14)

Expression of EryACP1 was carried out as reported for ***11*** but not yet purified.

In vitro Conversion of _apo_-ACPs to Acetoacetyl-_holo_ACPs

A 50 µL reaction mixture of *apo*-ACP (20 µM), acetoacetyl-CoA (60 µM), Sfp (2 µM), $MgCl_2$ (10 mM), and Na_2HPO_4 (100 mM, pH 7.2) was incubated at 37 ºC for 1 hr. The finished reaction mixture was dialyzed in 1:1 methanol-water mixture plus 0.1% acetic acid (Slide-A-Lyzer MINI Dialysis Devices, ThermoFisher Scientific, USA) for mass spectrometry analysis.

CHAPTER THREE
RESULTS AND DISCUSSION

Constructs Design Strategy

Strategy for Constructing Chimeric Didomains

AmpKR2EryACP1 and EryKR1AmpACP2 didomains were designed by using sequence alignment to locate regions with maximum amino acids (aa) sequence similarity at the KR-ACP linker regions in AmpKR2ACP2 and EryKR1ACP1. AmpKR2EryACP1 required adding few amino acid residues from the N-terminal of AmpACP2 to the C-terminal of AmpKR2 before adding the EryACP1 sequence (Figure 3.1A). EryKR1AmpACP2 required removing few amino acids from the N-terminal of AmpACP2, while maintaining the entire EryKR1 sequence (Figure 3.1B). The EryACP1 and AmpACP2 used in designing these chimeric didomains are the 187 aa AmpACP2 and 182 aa EryACP1 present in the native AmpKR2ACP2 and EryKR1ACP1 didomain constructs, respectively (see appendix for DNA and protein sequences).

Strategy for Constructing Standalone AmpACP2

Effort to express the 187 aa AmpACP2 sequence, present in AmpKR2ACP2 (**6),** as standalone domain was unsuccessful. The domain boundaries of this ACP domain were not correctly determined in the initial design, which later happened to be the reason for the failed expression. To locate the domain boundaries and express soluble AmpACP2 domain, three constructs of different lengths (81 aa, 91 aa, and 101 aa) were designed and cloned based on information from DoBISCUIT, an online database of biosynthetic gene clusters for secondary metabolites. However, none of these three constructs expressed soluble protein. Following this observation, three new constructs of AmpACP2 (91 aa, 101 aa, 111 aa) were designed based on the work of Alekseyev and coworkers[63] on EryACP2. The 187 aa AmpACP2 sequence was at the same time analyzed on ***IntFOLD Integrated Protein Structure and Function Prediction Server (version 7.0)*** for more accurate prediction of the domain boundaries. The results from the server predicted the 187 aa sequence to consist of AmpACP2 (~ 99 aa, ~10.4 kDa) and another polypeptide at the C-terminal, possibly the AmpACP2 C-terminal linker for the non-covalent interaction between module 2 and module 3 of amphotericin PKS (Figure 3.2). The newly designed

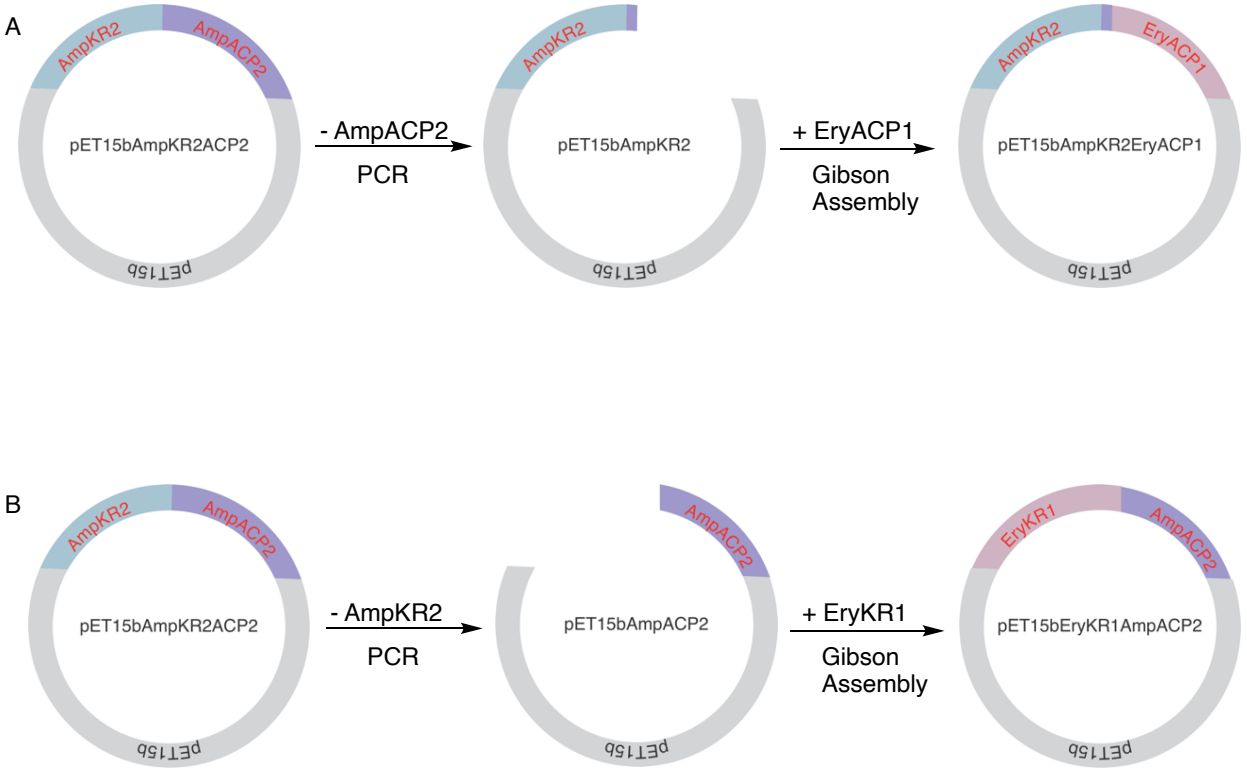


Figure 3.1. Schematic representation of the construction of chimeric didomains from native didomains.

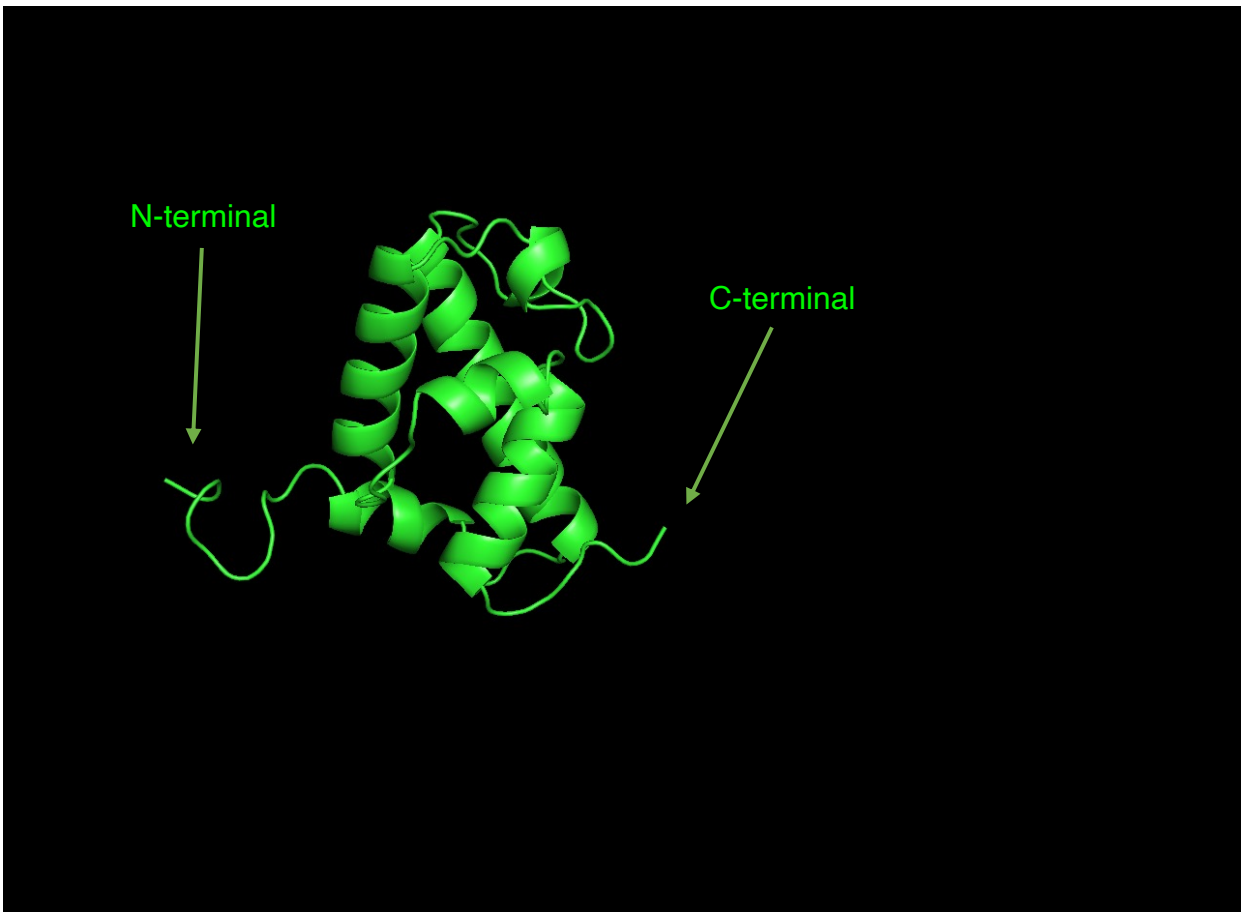


Figure 3.2. Model of AmpACP2 (attached polypeptide not shown here).

101 aa sequence only differed from the software-predicted sequence by two additional amino acids to the N-terminal. These four newly designed sequences were cloned into pET15b vector and confirmed by sequencing. Expression of 111 aa AmpACP2 (11.5 kDa) and the software-predicted 99 aa AmpACP2 (10.4 kDa) constructs was carried out in *E. coli* BL21 (DE3). The two constructs expressed AmpACP2 but with a larger percentage as insoluble protein in insoluble fraction as inclusion bodies (Figure 3.3). It can be concluded from these results that the extra amino acids chain in the initial 187 aa sequence negatively impacts the solubility of AmpACP2. This observation reveals the role that extra amino acid residues can play on solubility of protein macromolecules.

Strategy for Constructing Standalone EryACP1

The 182 aa EryACP1 sequence present in EryKR1ACP1 (**8**) was initially cloned into pET28a vector. The expression of this plasmid in *E. coli* BL21 (DE3) and *E. coli* BL21 Star (DE3) strains resulted in no observable soluble EryACP1. Following this observation and the observation from AmpACP2 construct design, domain boundaries of EryACP1 was investigated. The 182 aa EryACP1 was found to consist of EryACP1 (~ 89 aa) linked to the N-terminal of EryKS2 (Figure 3.4), in agreement with the work of Donadio and coworker.[64] EryKS2 is the ketosynthase domain in DEBS module 2, downstream of EryACP1. It receives the processed polyketide intermediate from EryACP1 to catalyze a new decarboxylative Claisen-like condensation reaction with a newly loaded extender unit on the downstream EryACP2. EryACP1 is covalently connected to the N-terminal of EryKS2 via an intermodular linker, shown in blue in the model, facilitating the transfer of the growing polyketide between module 1 and module 2. The predicted ~ 89 aa sequence was cloned into pET15b vector and expressed in *E. coli* BL21 (DE3). However, a large percentage of the expressed protein was in the cell pellet as insoluble protein, necessitating optimization of the expression conditions (Figure 3.5). Expression and purification of this protein will not be completed in this thesis, but future expression will involve optimization to obtain EryACP1 as soluble and pure protein in high yield.

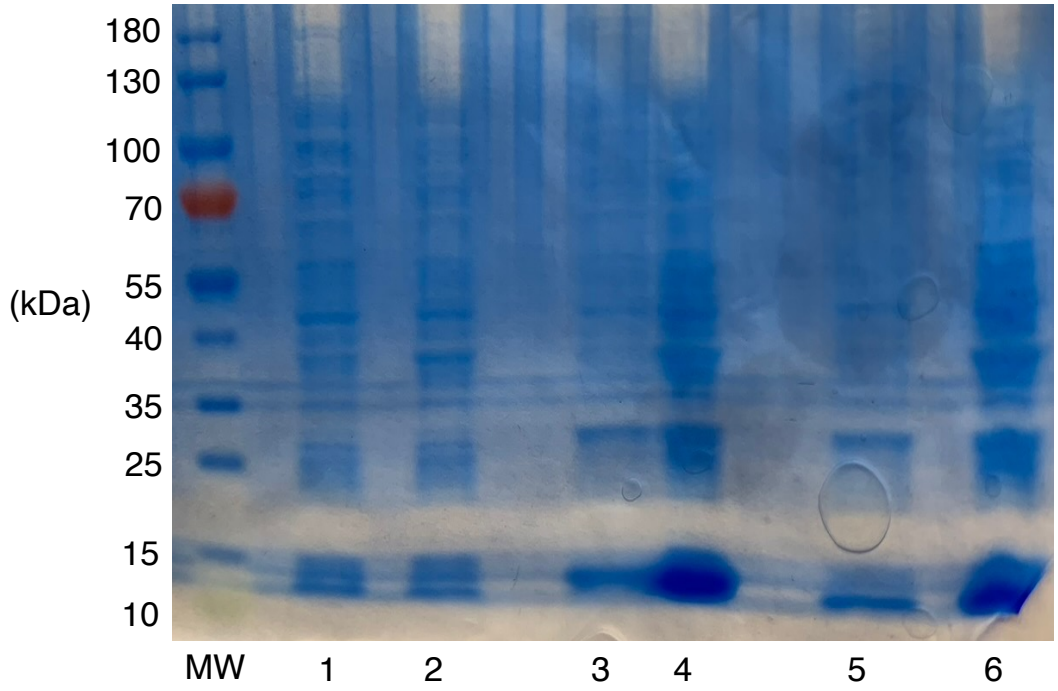


Figure 3.3. Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) of AmpACP2. Lane 3, lysate (11.5 KDa); lane 4, pellet (11.5 kDa); lane 5, lysate (10.4 kDa); and lane 6, pellet (10.4 kDa).

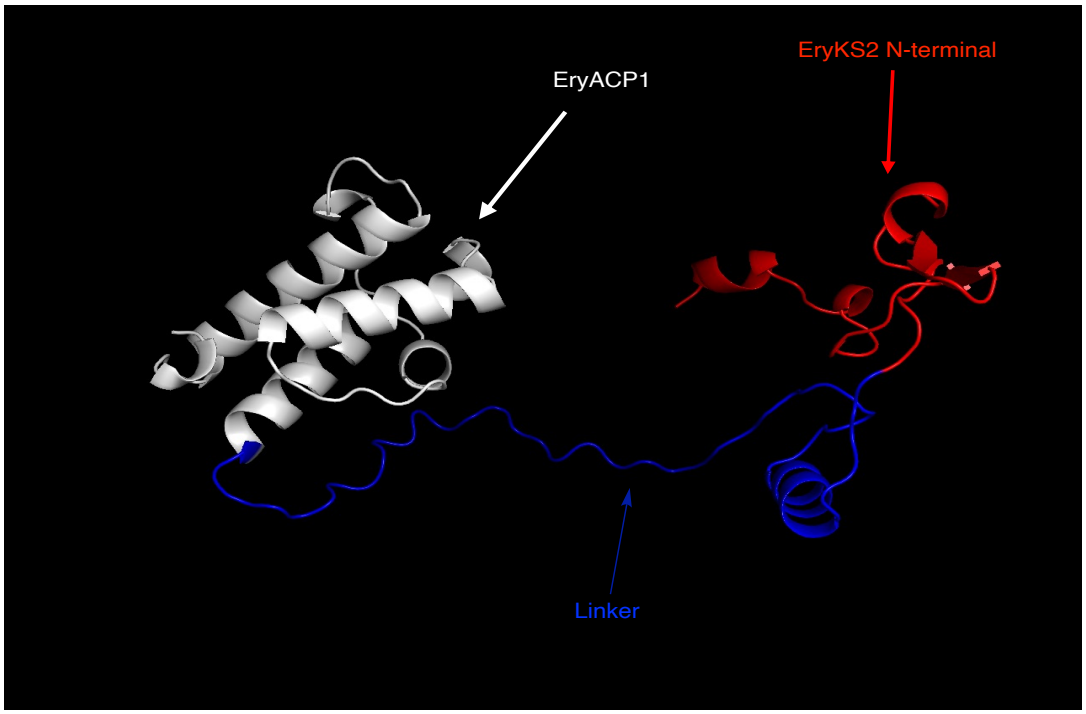


Figure 3.4. Model showing EryACP1 (white) linked to the N-terminal of EryKS2 (red) in the 182 aa sequence.

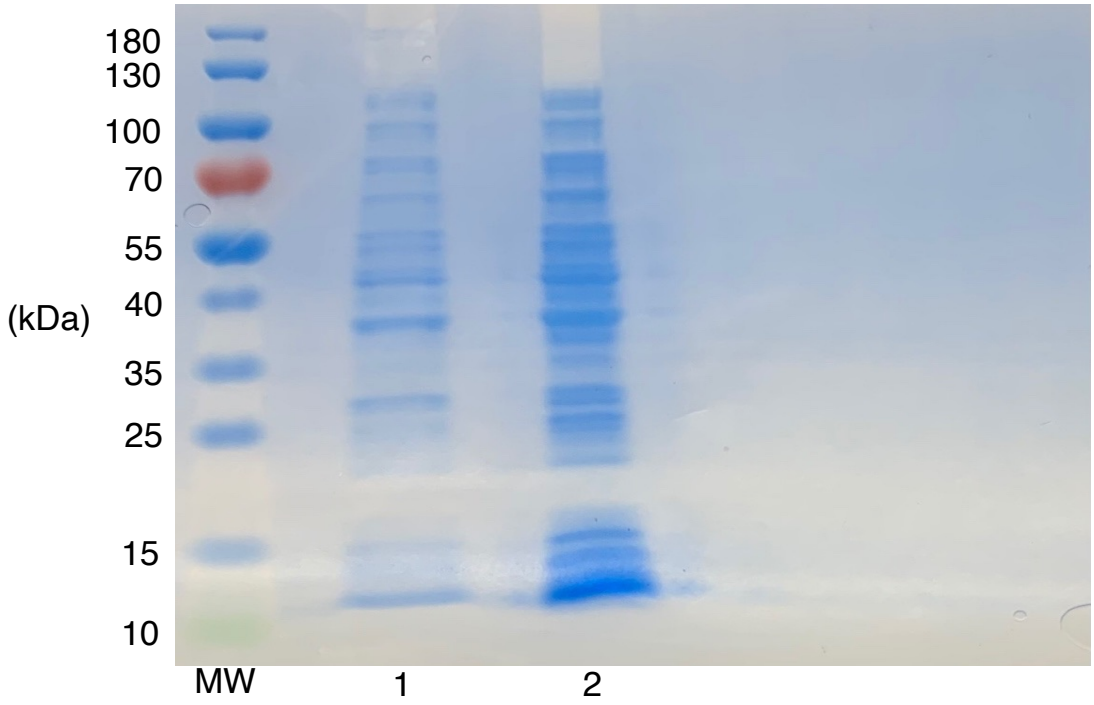


Figure 3.5. SDS-PAGE of EryACP1 (11.3 kDa). Lane 1, lysate; and lane 2, pellet.

Protein Expression and Purification

Purification of EryKR1AmpACP2

Expression of EryKR1AmpACP2 (**8**) in *E. coli* BL21 (DE3) strain (Figure 3.6) resulted in impure protein. Purification of **8** from cell lysate by affinity chromatography resulted in impure fractions of **8**, with an impurity appearing at ~ 57 kDa (Figure 3.7). The protein contaminant was thought to be a GroEL, an *E. coli* chaperonin that assists in protein folding. It is common to encounter the co-purification of chaperonin while purifying heterologously-expressed proteins. In such cases, the purification process may require optimization to obtain pure protein samples. As a result, effort was focused on removing the contaminant at the purification stage. Raji and coworker[65] developed a method that uses $MgCl_2$/ATP/KCl mixture to remove chaperones by incubating the mixture with the impure proteins. This method was not effective in the purification of EryKR1AmpACP2 from the protein contaminant despite several optimizations. Ion exchange chromatography using quaternary (Q) anion exchanger also failed to purify this protein of interest. It was later realized that the suspected GroEL chaperonin contaminant has a predicted isoelectric point (pI) of 4.85, while the protein of interest has a predicted pI of 4.84, making ion exchange chromatography unsuitable in this situation.

Efforts were later focused on optimizing protein expression conditions after several attempts to separate EryKR1AmpACP2 from the contaminant. A different strain of *E. coli*, BL21 Star (DE3), designed to decrease mRNA degradation, was used for the expression. This strain expressed soluble EryKR1AmpACP2, the purification of which yielded pure protein (Figure 3.8 and Table 3.1). These results suggested that the impurity observed when the protein was expressed in *E. coli* BL21 (DE3) is possibly a truncated form of EryKR1AmpACP2 and not a chaperone.

Purification of AmpKR2EryACP1

Expression of AmpKR2EryACP1 (71.6 kDa) in *E. coli* BL21 (DE3) and purification by affinity chromatography resulted in fractions of protein of interest with another protein of smaller molecular weight (Figure 3.9). The identity of this smaller protein could not be ascertained and was assumed to be a contaminant from the *E. coli* host proteins. Effort to purify AmpKR2EryACP1 from the cell contaminant was futile, including optimization of the $MgCl_2$/ATP/KCl method reported above for EryKR1AmpACP2 and ion-exchange chromatography. Expression of

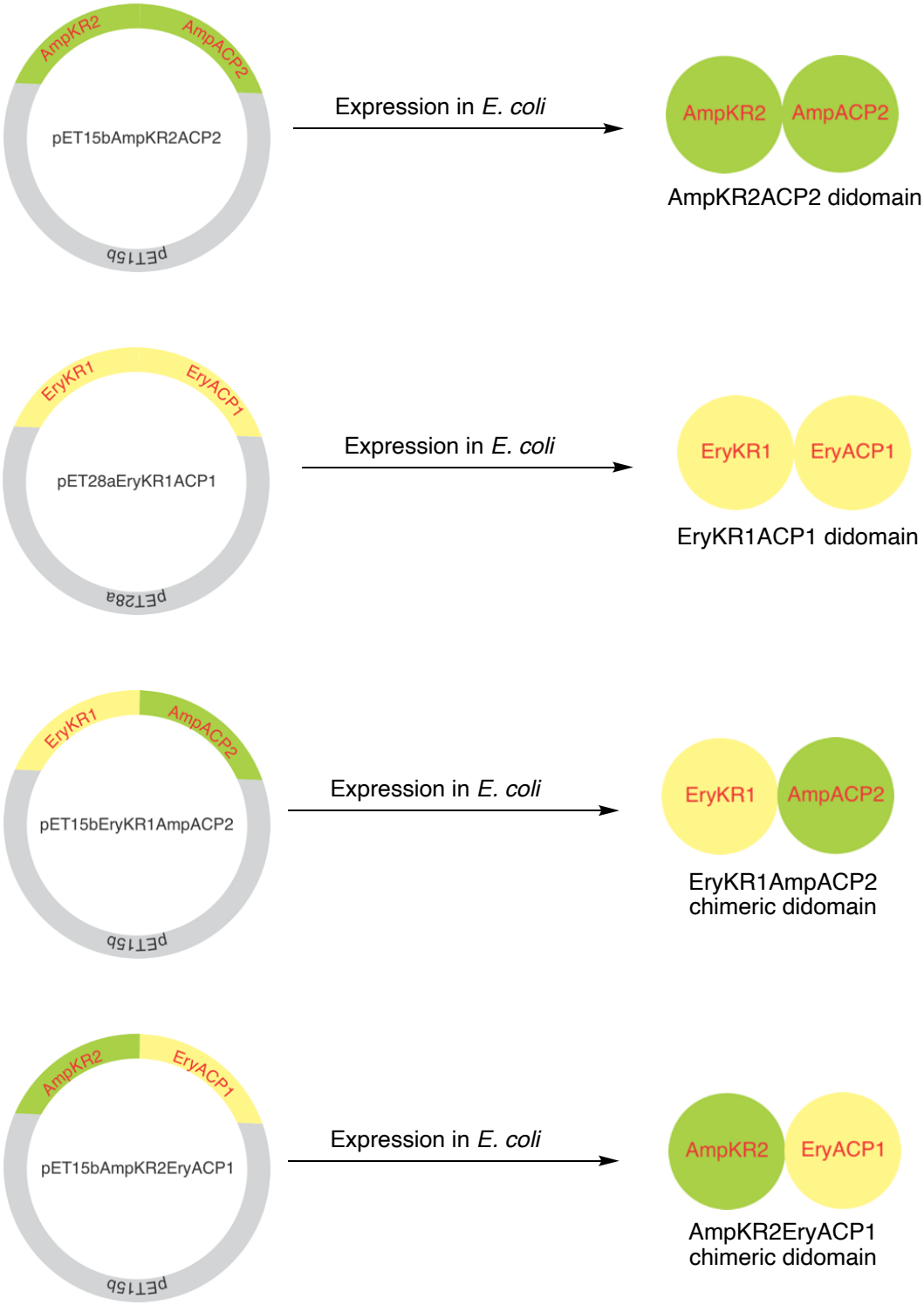


Figure 3.6. Schematic representation of the biosynthesis of enzymes from their corresponding constructs (color selection is random).

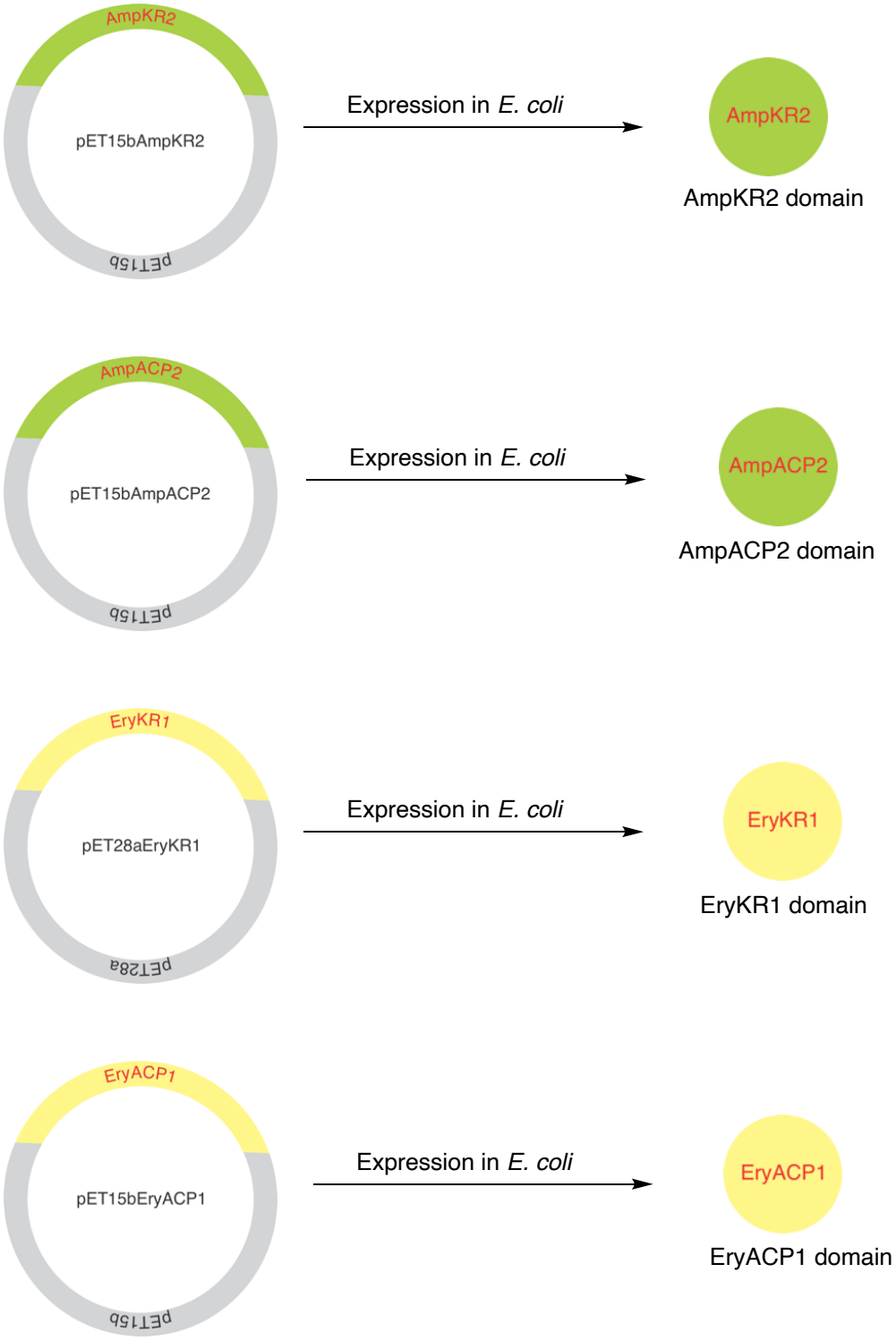


Figure 3.6. Continued

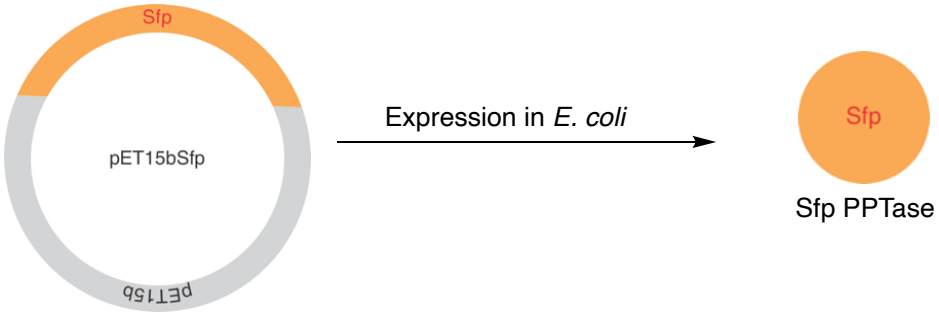


Figure 3.6. Continued

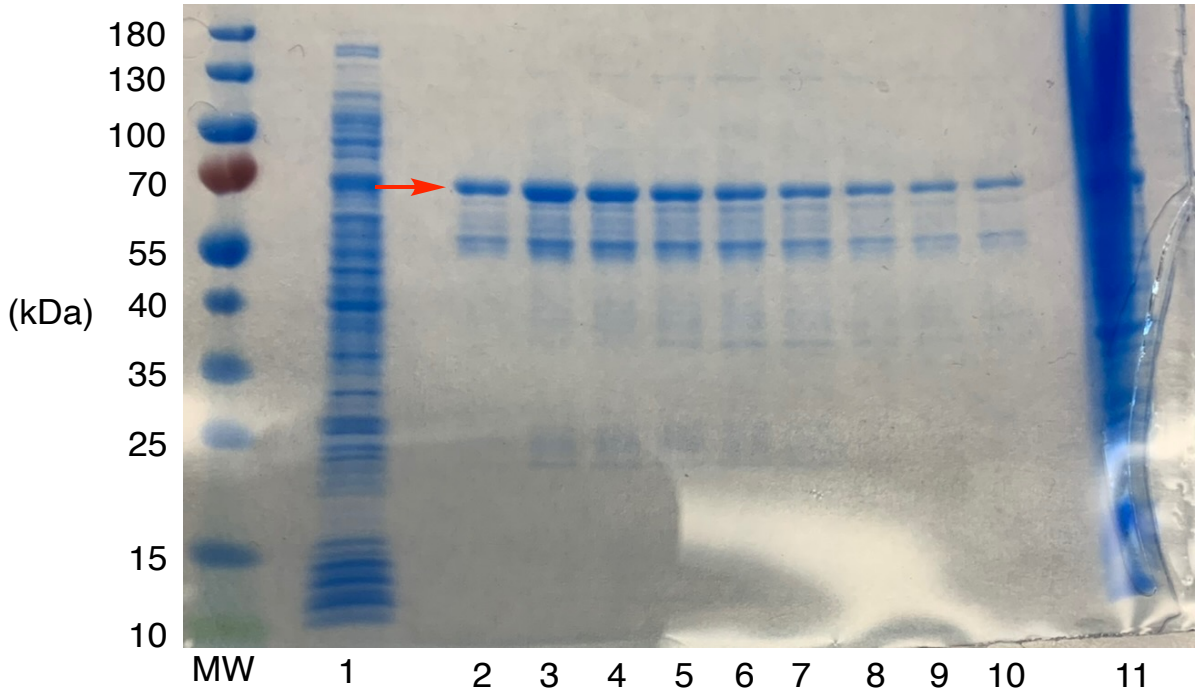


Figure 3.7. SDS-PAGE of impure EryKR1AmpACP2 (70.8 kDa). Lane 1, lysate; lanes 2-10, impure fractions; and lane 11, pellet.

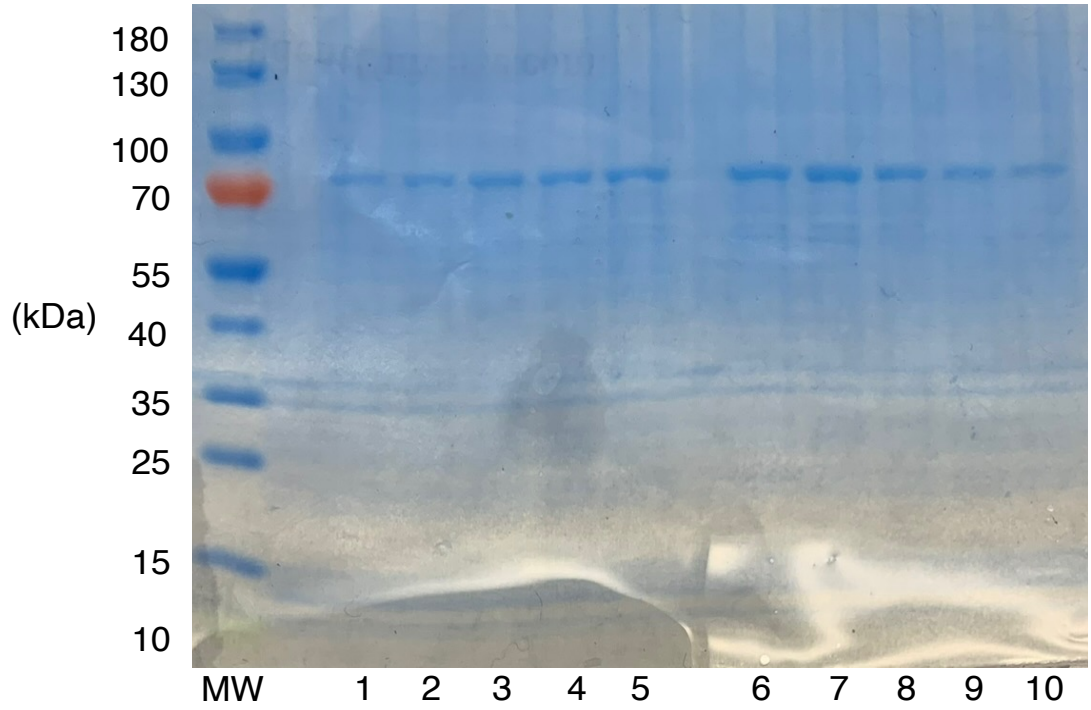


Figure 3.8. SDS-PAGE of EryKR1AmpACP2. Lanes 1-10, pure fractions of EryKR1AmpACP2.

Table 3.1. List of proteins biosynthesized and purified in this thesis.

Protein	**Yield (mg)**	**Approximate MW including 6xHis-tag (kDa)**
EryKR1ACP1	7.26	71.6
AmpKR2EryACP1	2.65	71.6
EryKR1AmpACP2	4.70	70.8
AmpKR2ACP2	2.50	70.7
AmpACP2	5.55	12.4
EryKR1	9.87	52.8
AmpKR2	10.00	51.3
Sfp	10.67	28.2

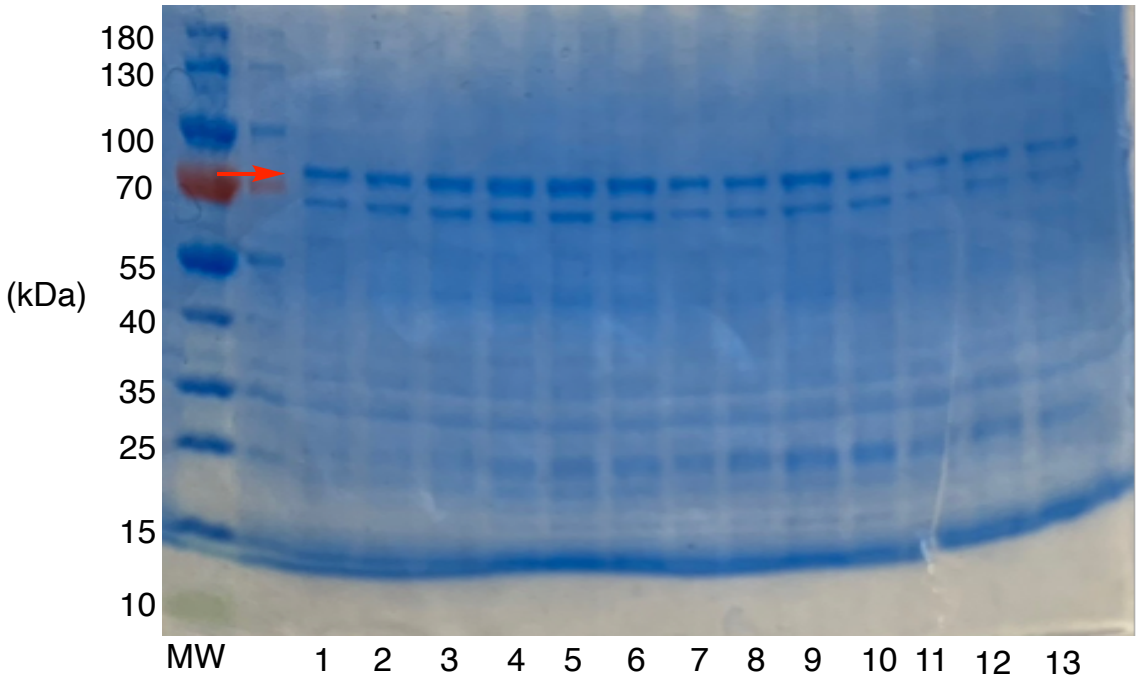


Figure 3.9. SDS-PAGE of impure AmpKR2EryACP1 (71.6 kDa). Lanes 1-13, impure fractions of AmpKR2EryACP1.

AmpKR2EryACP1 in *E. coli* BL21 Star (DE3) resulted in cleaner protein but with a trace amount of the contaminant (Figure 3.10). It is possible that the assumed contaminant is a truncated form of AmpKR2EryACP1 that shares the same chromatographic binding properties with the full length AmpKR2EryACP1.

Purification of AmpACP2

Expression of AmpACP2 in *E. coli* BL21 (DE3) resulted in the production of AmpACP2 with a larger percentage of the protein remaining insoluble (Figure 3.3). Purification of the small soluble AmpACP2 on 5 mL affinity column was unsuccessful. The expression condition was further optimized for the 99 aa AmpACP2 construct, resulting in more soluble AmpACP2 than previously obtained. Initial attempt to purify this protein yielded impure fractions, which necessitated a second round of purification. Partially pure fractions from the first purification were selectively collected and purified on the same column, resulting in pure fractions of AmpACP2 (Figure 3.11).

Purification of EryKR1ACP1

Expression of EryKR1ACP1 in *E. coli* BL21 Star (DE3) yielded a good amount of soluble protein of interest in the cell lysate. Purification of this protein by affinity chromatography resulted in fractions of EryKR1ACP1 with slight impurity (Figure 3.12). All efforts to completely remove this contaminant was futile, leading to the conclusion that it might be a truncated form of EryKR1ACP1.

Purification of AmpKR2ACP2

Expression and purification of AmpKR2ACP2 resulted in the production of soluble protein with significant amount of impurities. Further purification of fractions from the first purification also failed to yield pure protein of interest. Subsequent expression and purification of this protein involved collection of the partially pure fractions from the first purification, followed by a second round of purification on affinity column. This method resulted in production of AmpKR2ACP2 with trace amount of impurities (Figure 3.13).

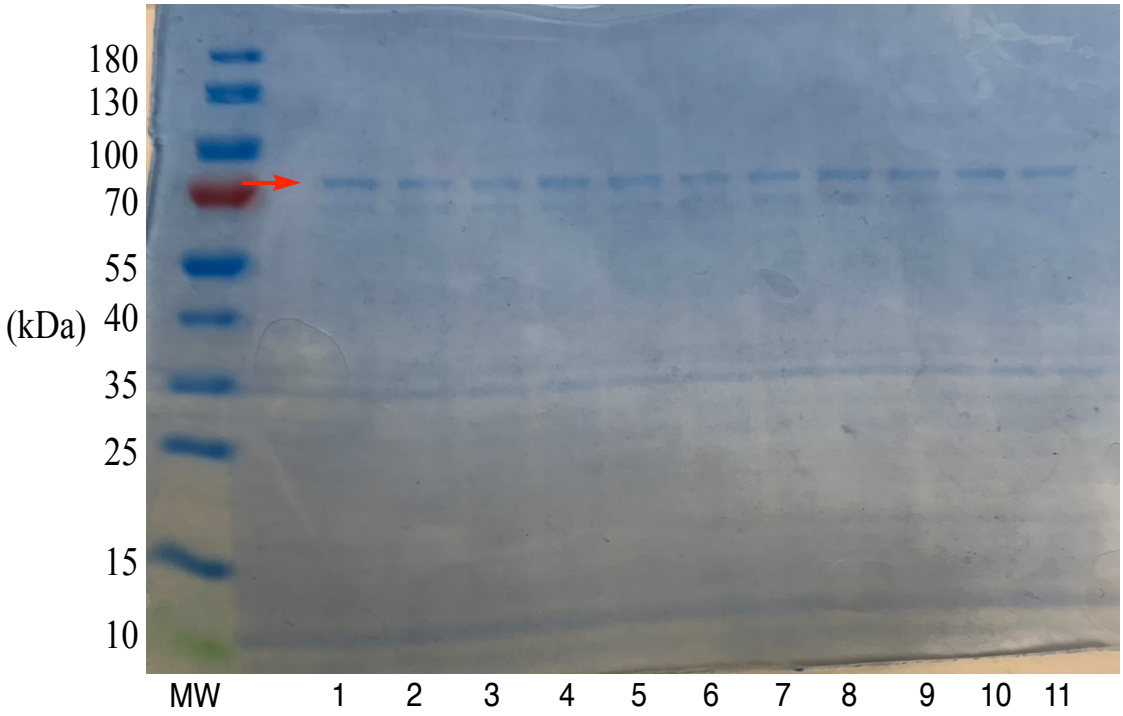


Figure 3.10. SDS-PAGE of AmpKR2EryACP1 (71.6 kDa). Lanes 1-11, pure fractions of AmpKR2EryACP1.

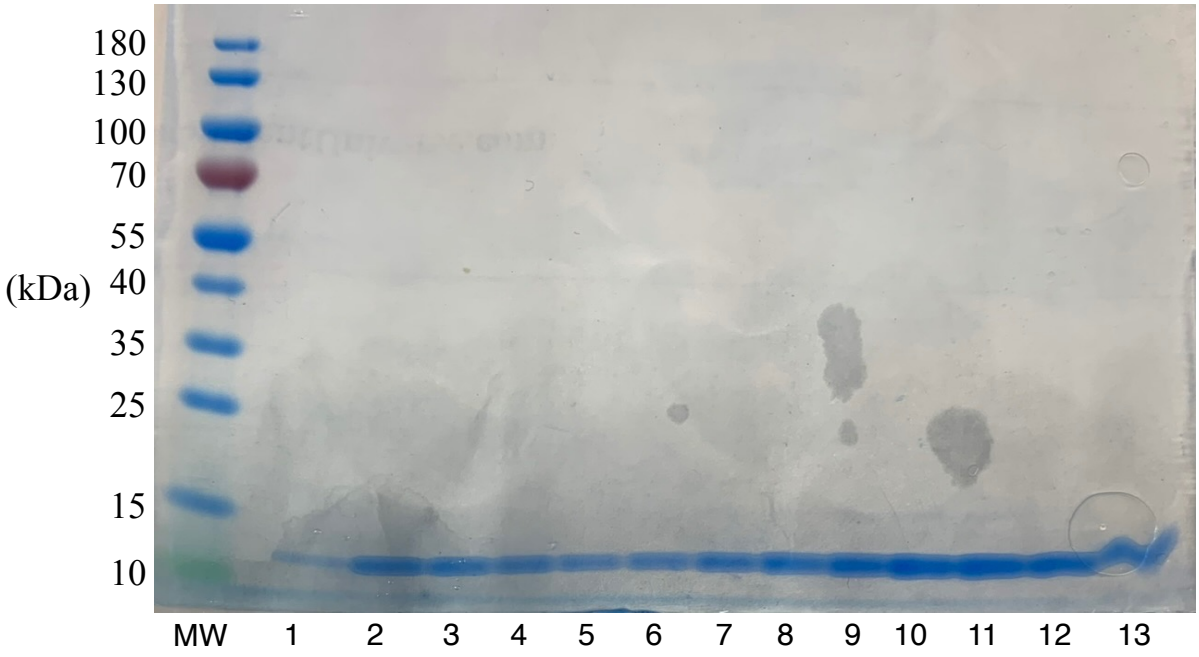


Figure 3.11. SDS-PAGE of AmpACP2 (12.4 kDa). Lanes 1-13, pure fractions of AmpACP2.

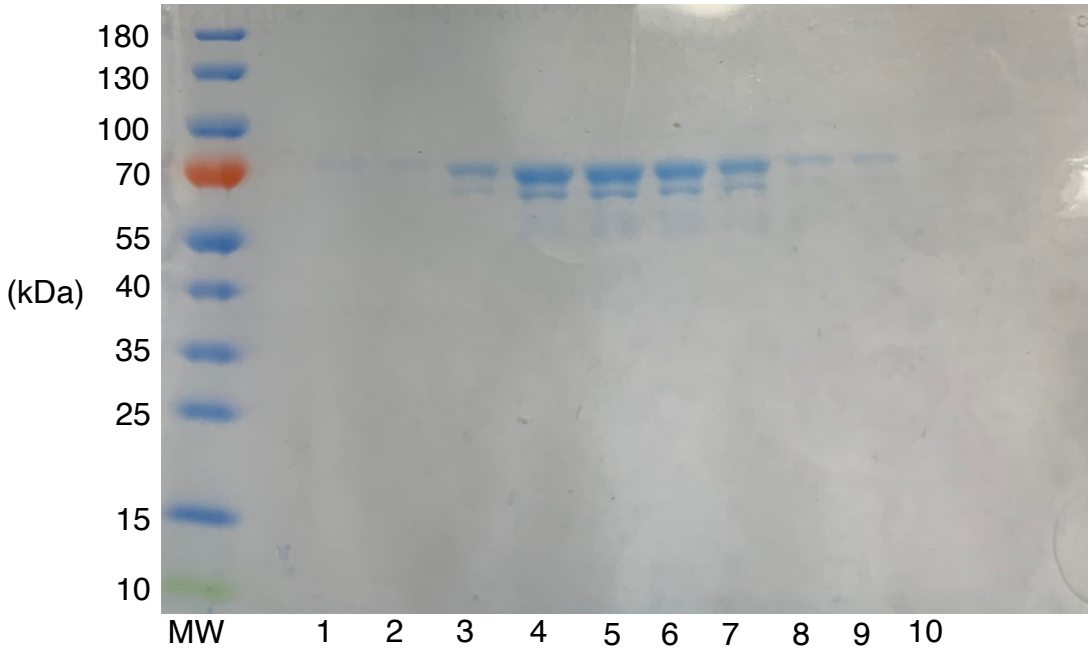


Figure 3.12. SDS-PAGE of EryKR1ACP1 (71.6 kDa). Lanes 1-10, pure fractions of EryKR1ACP1.

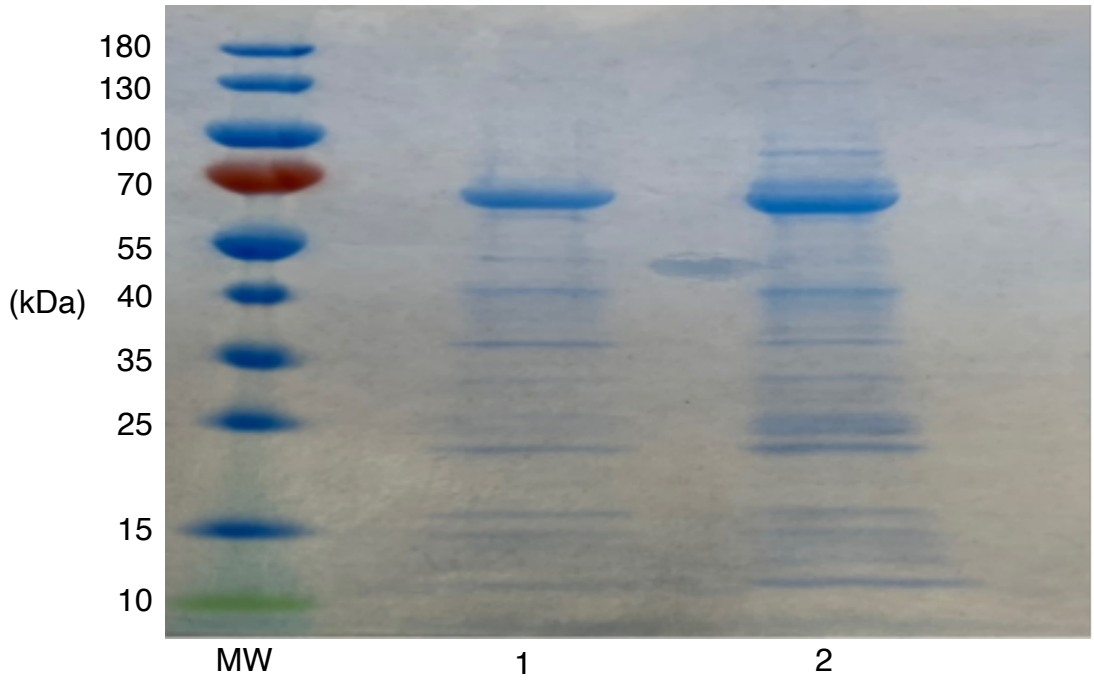


Figure 3.13. SDS-PAGE of AmpKR2ACP2 (70.7 kDa). Lanes 1and 2, purified fractions of AmpKR2ACP2.

Purification of AmpKR2, EryKR1, and Sfp

The expression of AmpKR2, EryKR1, and Sfp was carried out in *E. coli* BL21 (DE3), followed by purification on an affinity column, resulting in the isolation of pure proteins with molecular weights that were consistent with their expected values. (Figures 3.14 – 3.16).

Sfp-catalyzed Phosphopantetheinylation of *apo*-ACP Domains

The acyl carrier proteins (ACP) biosynthesized here were expressed in their inactive (*apo*) form and require post-translational modification with the PPant arm of a coenzyme-A substrate to become active (Figure 1.2). The enzymes were expressed inactive, so they can be easily phosphopantetheinylated *in vitro* to carry the substrate (acetoacetyl thioester) that will be reduced by the ketoreductase domains in the NADPH dependent kinetic assay. Sfp PPTase from *Bacillus subtilis* was used to convert the *apo*-ACPs to acetoacetyl-*holo*ACP in the presence of acetoacetyl-CoA substrate (Figure 3.17).

In order to obtain reliable data from the kinetic assay, it is essential to confirm that the acyl carrier proteins were phosphopantetheinylated with the phosphopantetheinyl (PPant) prosthetic group of acetoacetylCcoA before performing the kinetics assay. The first attempt to detect the intact *holo*-AmpKR2ACP2 and *holo*-EryKR1ACP1 resulted in spectra with poor signal to noise, preventing assignment of m/z value for the intact proteins. This observation was suggested to have resulted from the residual nonvolatile salts in the buffer. In another attempt, the Sfp-catalyzed phosphopantetheinylation reaction products of AmpKR2ACP2, EryKR1ACP1, AmpKR2EryACP1, EryKR1AmpACP2, and AmpACP2 were dialyzed for mass spectrometry analysis to detect the intact phosphopantetheinylated proteins (Table 3.2). Mass spectrometry analysis of these samples is currently under optimization.

Another method we can use to confirm the phosphopantetheinylated proteins is the MS^3 fragmentation method developed by Meluzzi and coworkers.[66] This method, called phosphopantetheine ejection assay, is used to detect the molecular mass of the substrate tethered to the PPant arm of an acyl carrier protein. This is achieved by ejecting the PPant arm from the *holo*-ACP domain in the MS^2, which leaves the substrate attached to the PPant fragment. The PPant fragment from the MS^2 is usually indicative of a phosphopantetheinylated ACP domain and its tethered substrate, if any. This fragment can be further fragmented to give nine unique MS^3

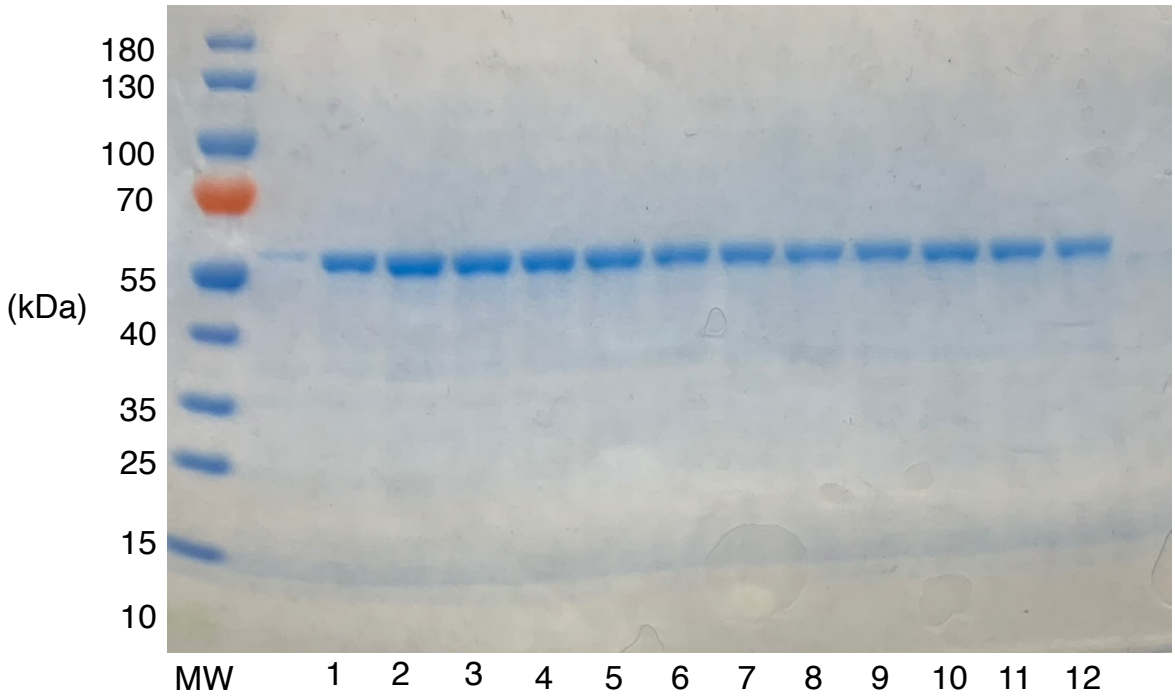


Figure 3.14. SDS-PAGE of EryKR1 (52.8 kDa). Lanes 1-12, pure fractions of EryKR1.

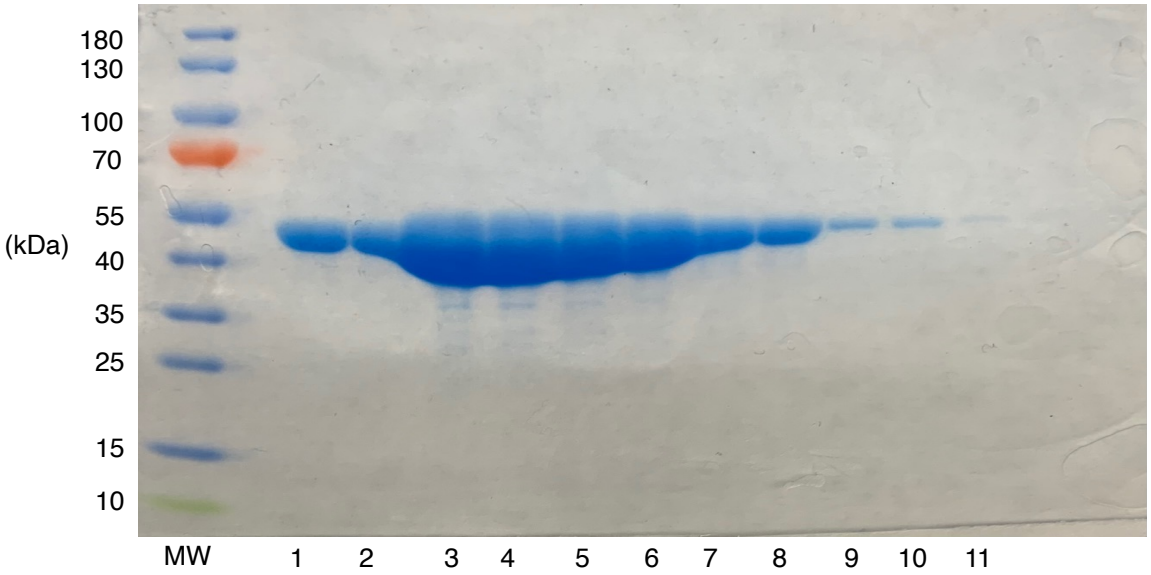


Figure 3.15. SDS-PAGE of AmpKR2 (51.3 kDa). Lanes 1-11, pure fractions of AmpKR2.

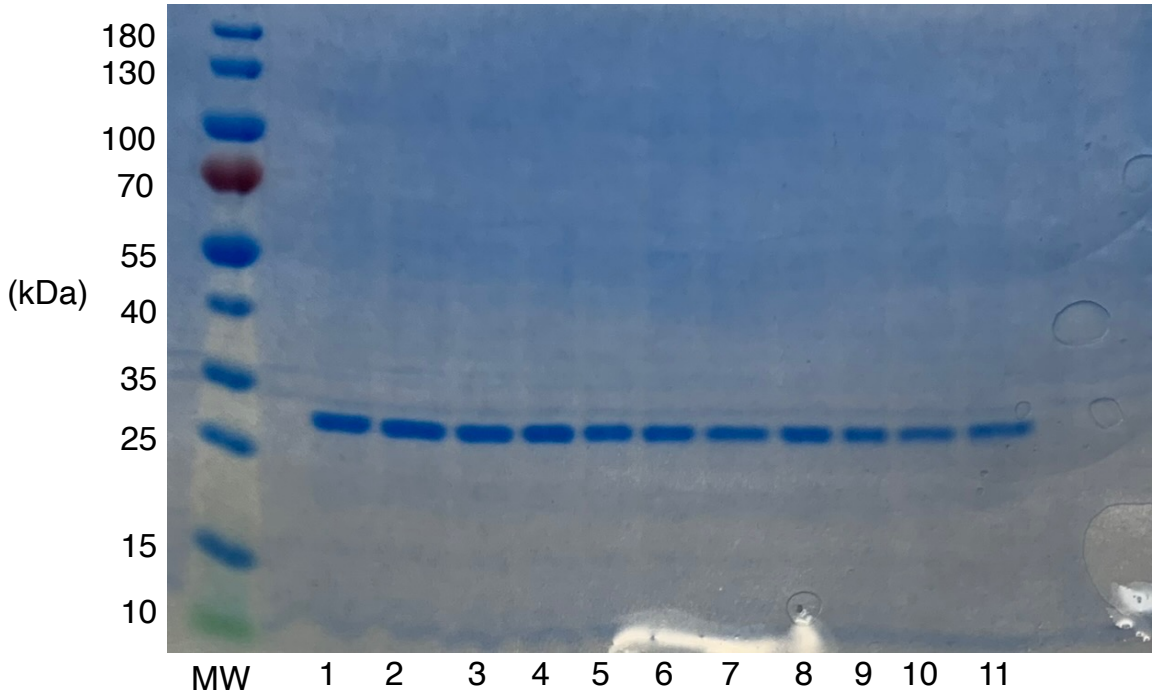


Figure 3.16. SDS-PAGE of Sfp (28.2 kDa). Lanes 1-11, pure fractions of Sfp.

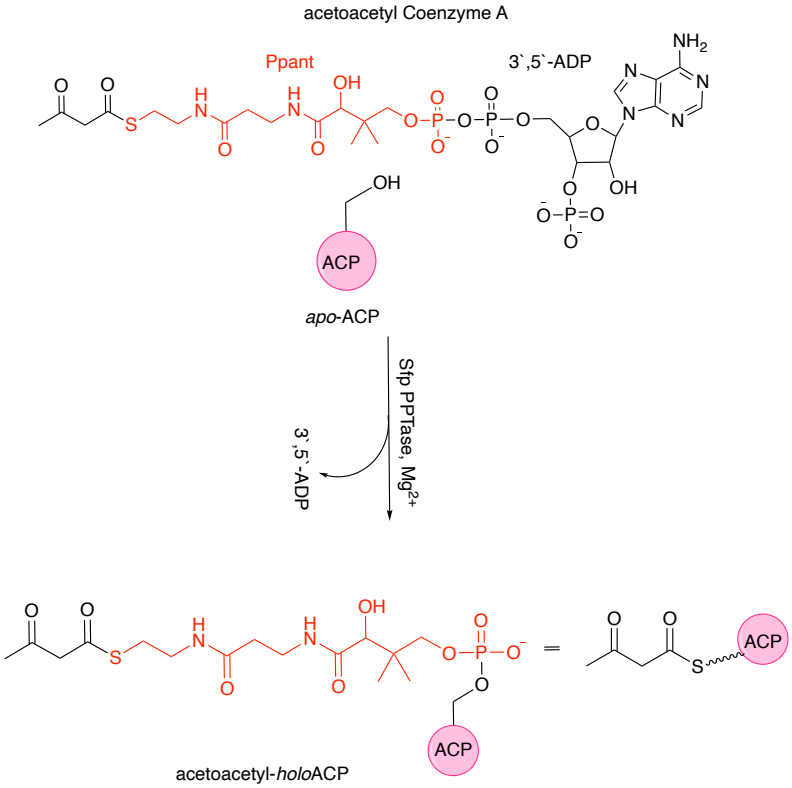


Figure 3.17. Conversion of *apo*-ACPs to acetoacetyl-*holo*ACPs.

Table 3.2. Phosphopantetheinylated ACP domains and KR-ACP didomains and their calculated molecular weight.

Modified Enzyme	**Calculated Molecular Weight (Da)**
Acetoacetyl-EryKR1ACP1	71,982.628
Acetoacetyl-AmpKR2EryACP1	72,017.908
Acetoacetyl-AmpKR2ACP2	71,092.738
Acetoacetyl-EryKR1AmpACP2	71,194.598
Acetoacetyl-AmpACP2	12,862.388

peaks with m/z values that are dependent on the mass of substrate and three other peaks with constant m/z values (142.1, 166.1, and 184.1) that are diagnostic of *holo*-ACP domains, irrespective of substrate mass (Figure 3.18). We will use this tandem MS technique to convincingly tell whether the phosphopantetheinylation was successful or not.

Proposed NADPH-Dependent Kinetic Assay

As mentioned earlier, NADPH is a co-factor in the catalytic reduction of β-keto group in the polyketide biosynthetic pathways, and this molecule is oxidized in the process, becoming $NADP^+$ (Figure 3.19). NADPH co-factor absorbs UV light at a wavelength of 340 nm, but $NADP^+$ does not absorb at this wavelength. After confirming phosphopantetheinylation of the ACP domains, the rate at which the β-keto group of the acetoacetyl-*holo*ACP is reduced by a standalone ketoreductase domain and a ketoreductase present in a didomain will be measured *in vitro*. The measurement will involve monitoring the absorbance at 340 nm and recording the decrease in absorbance as the reaction progresses. Data from the NADPH assay will be processed and fitted to the Michaelis-Menten kinetic equation (equation 4.1) to obtain relevant kinetic parameters (K_{cat} and K_m) used to quantify enzyme activities.

$$V_o = \frac{(V_{max}[S])}{(K_m + [S])} \qquad \text{Michaelis-Menten Equation.} \qquad (4.1)$$

V_o is the reaction rate at a particular substrate concentration

V_{max} is the reaction rate when an enzyme is completely saturated with substrates

[*S*] is the concentration of substrate

K_m is the substrate concentration an enzyme requires to reach half its V_{max}

K_{cat} is the turnover number of an enzyme, expressed as V_{max}/[*Enzyme*]

[*Enzyme*] is the enzyme concentration

To establish kinetic relationship between didomains and standalone domains, we will compare K_{cat} and K_m values from the reduction reactions catalyzed by native didomains and their corresponding standalone domains (Figure 3.20). We will also investigate protein-protein interactions by comparing K_{cat} and K_m values from reactions catalyzed by native domain partners and chimeric domains (Figure 3.21).

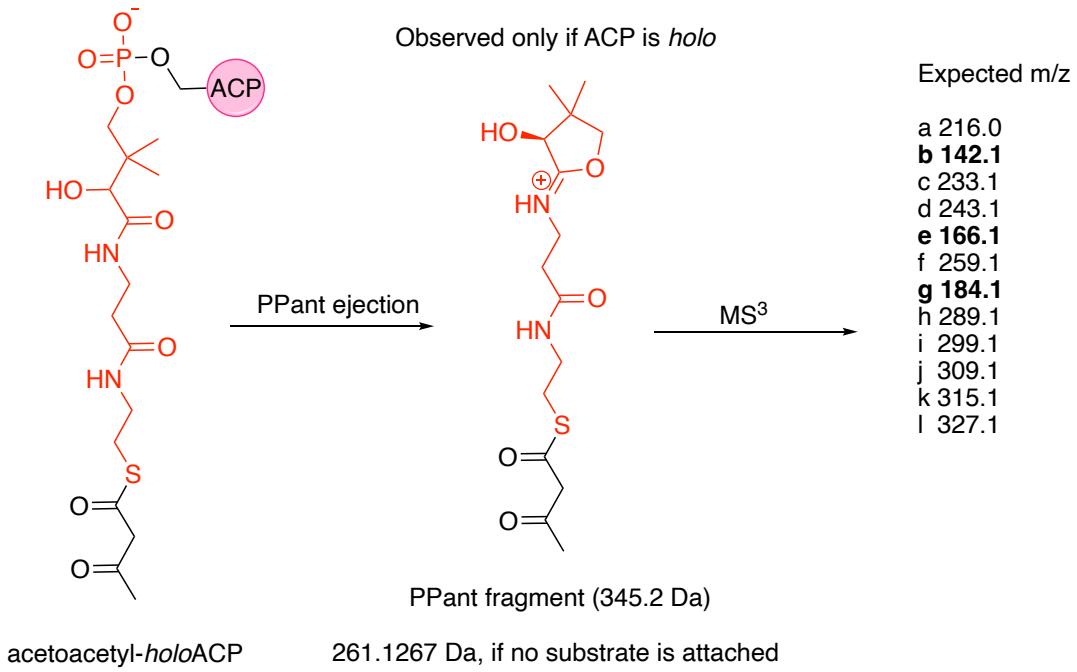


Figure 3.18. MS^3 fragmentation method to detect acetoacetyl-*holo*ACPs.

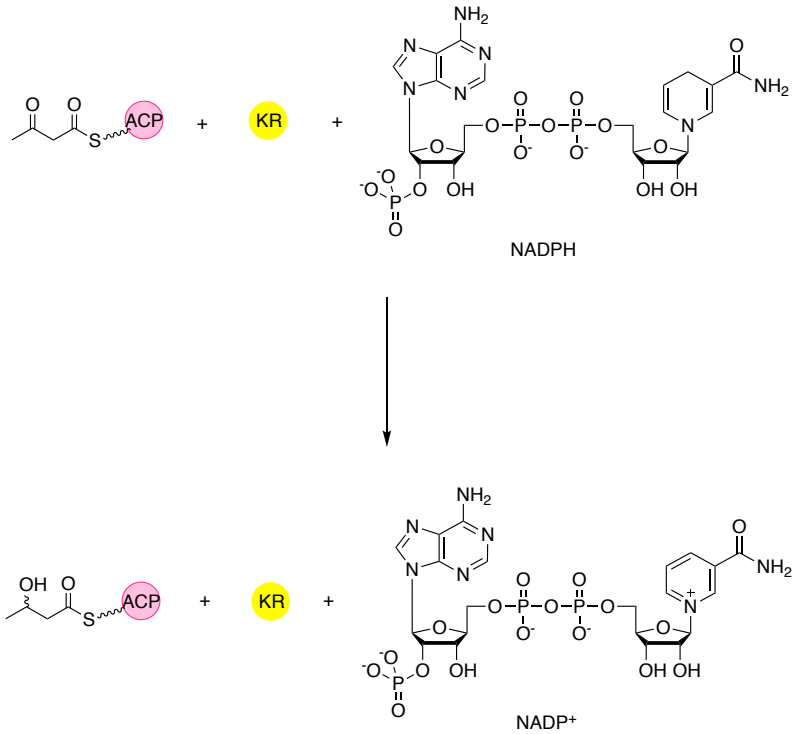


Figure 3.19. Schematic representation of KR-catalyzed reduction of acetoacetyl-*holo*ACP in the presence of NADPH.

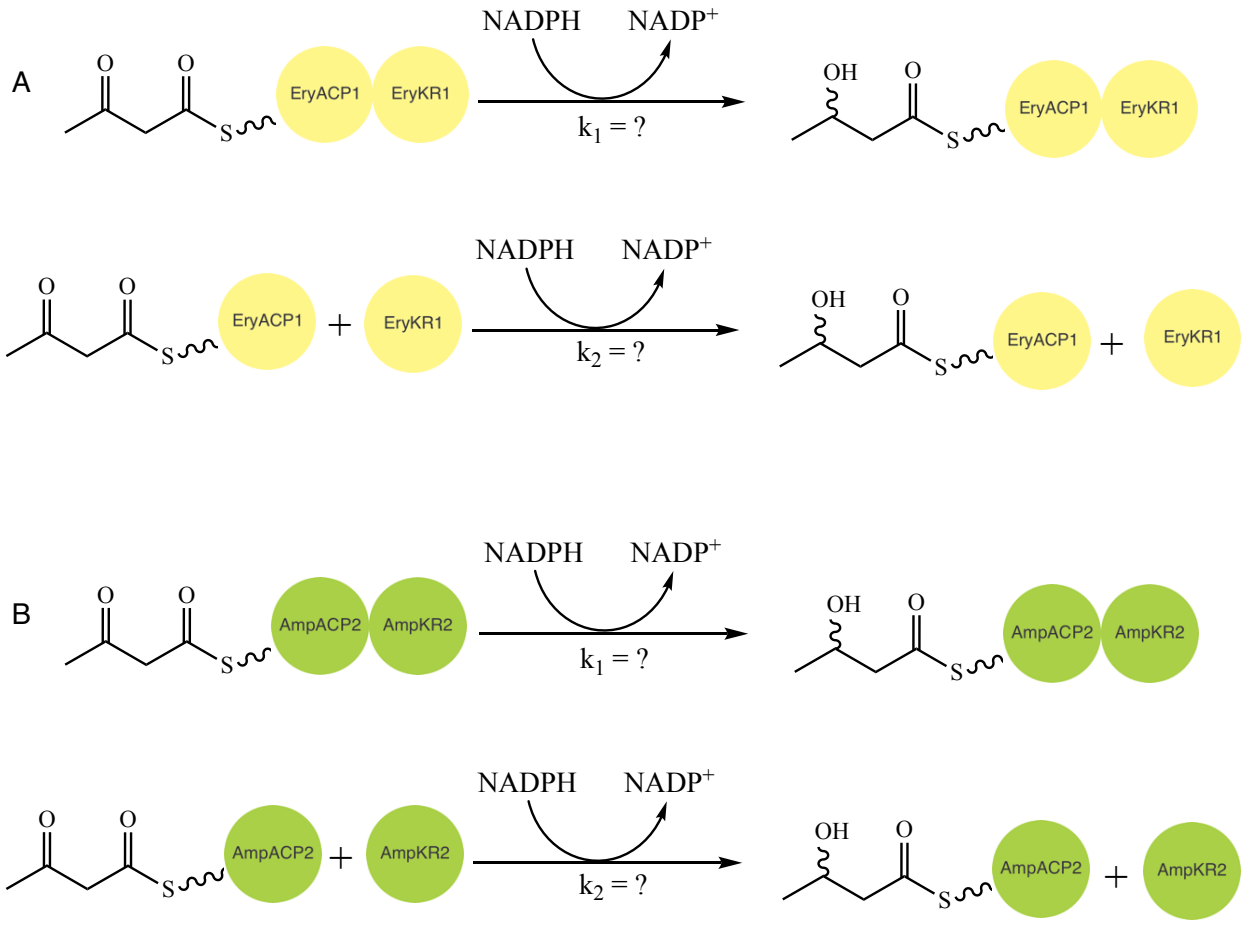


Figure 3.20. Schematic representation of type I PKS-like (didomains) and type II PKS-like (standalone domains) reactions catalyzed by native domains.

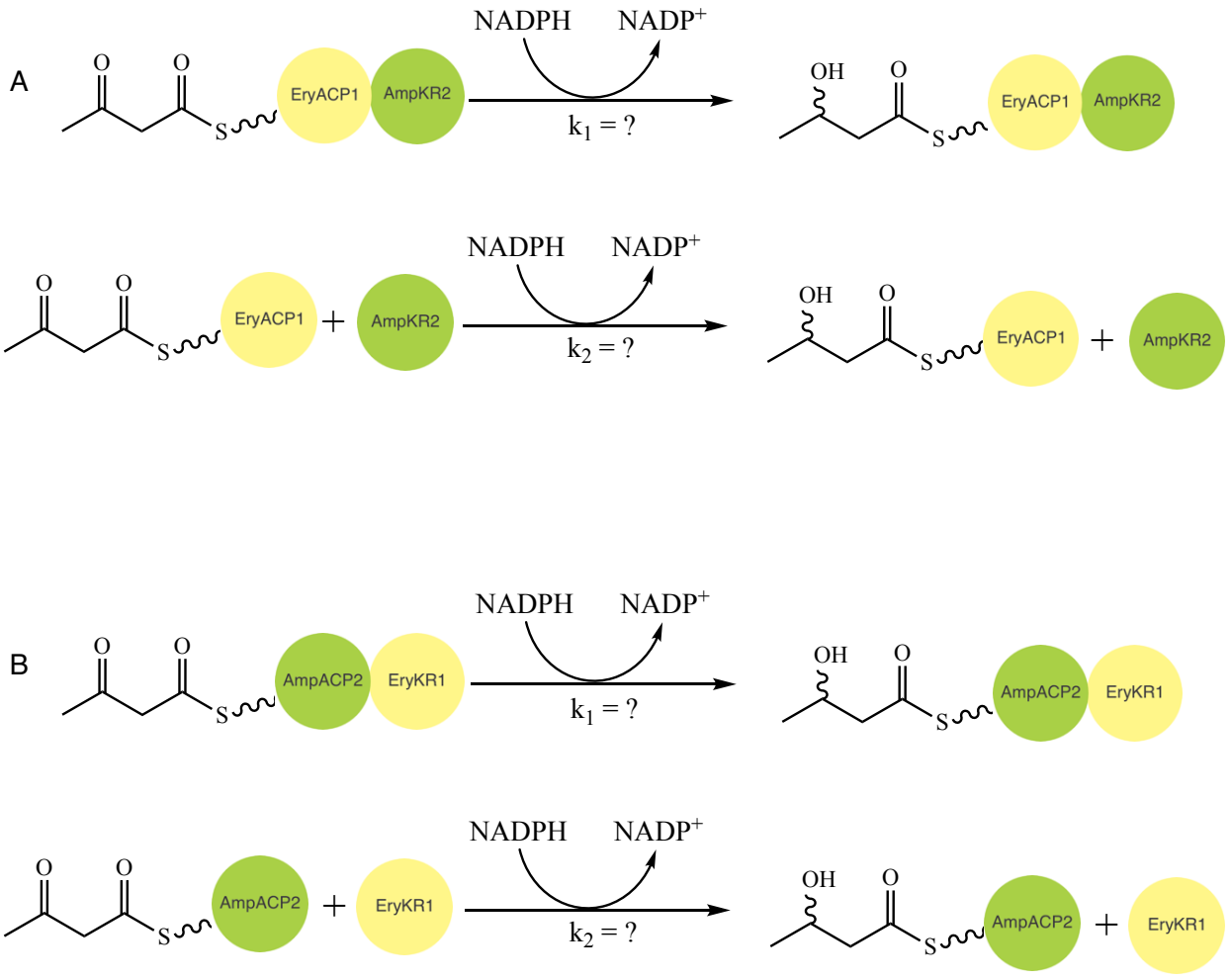


Figure 3.21. Schematic representation of type I PKS-like (didomains) and type II PKS-like (standalone domains) reactions catalyzed by chimeric domains.

CHAPTER FOUR
CONCLUSIONS

This thesis used the heterologous expression platform to biosynthesize ketoreductase-acyl carrier protein (KR-ACP) didomains and their corresponding standalone domains from the amphotericin polyketide synthase's second module and the 6-deoxyerythronolide B synthase's first module. Domain boundaries analysis was carried out to express the standalone acyl carrier proteins as soluble proteins, and it was discovered that extra amino acid residues at the C-terminal of these proteins can negatively impact their expression and solubility. The ability to avoid truncation of EryKR1AmpACP2 by expressing the chimeric didomain in *E. coli* BL21 Star (DE3) strain also indicates the continuous advancements of the heterologous expression platform and its potential in combinatorial biosynthesis.

The enzymes biosynthesized here will be used to establish kinetic relationships between didomains and standalone domains to gain more insight into the hypothesis that the type II polyketide synthases are the precursors to the type I polyketide synthases. The catalytic reactions occurring in type I PKSs are similar to intramolecular reactions in basic organic chemistry where both the substrate and the attacking species are parts of the same molecule. Whereas in type II PKSs, the enzymatic reactions are similar to intermolecular reactions where the substrate and the attacking species are separate entities. Establishing the relationship between the reaction rates in these two classes of PKSs can enlighten us on the importance of fusing the domains in type I PKSs, by knowing the possible kinetic penalty that is incurred in type II PKSs.

The enzymes will also be used to investigate protein-protein interactions between non-native acyl carrier proteins and ketoreductase domains. Investigation of domain-domain recognition between PKS domains is an important aspect of developing hybrid PKSs. Information from this investigation can guide rational selection of a set of domains that can be combined to increase the production of a natural product, modify the structure of an existing natural product, or biosynthesize a novel compound with predicted pharmacological properties. Data from this study will further supplement available data to aid rational engineering of chimeric modules for combinatorial biosynthesis.

REFERENCES

(1) Fischbach, M. A.; Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic Machinery, and Mechanisms. *Chemical Reviews*. **2006**, 3468–3496.

(2) Staunton, J.; Weissman, K. J. Polyketide Biosynthesis: A Millennium Review. *Natural Product Reports*. **2001**, 380–416.

(3) Helfrich, E. J. N.; Piel, J. Biosynthesis of Polyketides by Trans-AT Polyketide Synthases. *Natural Product Reports*. Royal Society of Chemistry. **2016**, 231–316.

(4) Motamedi, H.; Shafiee, A. The Biosynthetic Gene Cluster for the Macrolactone Ring of the Immunosuppressant FK506. *Euro. J. Biochem*. **1998**, *256* (3), 528-534.

(5) Tang, L.; Shah, S.; Chung, L.; Carney, J.; Katz, L.; Khosla, C.; Julien, B. Cloning and Heterologous Expression of the Epothilone Gene Cluster. *Science.* **2000**, *287* (5453), 640–642.

(6) Pelludat, C.; Rakin, A.; Jacobi, C. A.; Schubert, S.; Heesemann, J. The Yersiniabactin Biosynthetic Gene Cluster of Yersinia Enterocolitica: Organization and Siderophore-Dependent Regulation. *J. Bacteriology.* **1998**, *180* (3), 538-546.

(7) Wang, Z.; Bagde, S. R.; Zavala, G.; Matsui, T.; Chen, X.; Kim, C. Y. De Novo Design and Implementation of a Tandem Acyl Carrier Protein Domain in a Type I Modular Polyketide Synthase. *ACS Chem. Biol.* **2018**, *13* (11), 3072–3077.

(8) Schwecke, T.; Aparicio, J. F.; Molnar, I.; Konig, A.; Khaw, L. E.; Haydock, S. F.; Oliynyk, M.; Caffrey, P.; Cortes, S.; Lester, J. B.; Bohm, G. A.; Staunton, J.; Leadlay, P. F. The Biosynthetic Gene Cluster for the Polyketide Immunosuppressant Rapamycin. *Proc. Nat. Acad. Sci.* **1995**, *92* (17), 7839-7843.

(9) Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of E. Coli. *Science*, **2001**, *291* (5509) 1790-1792.

(10) Caffrey, P.; Lynch, S.; Flood, E.; Finnan, S.; Oliynyk, M. Amphotericin Biosynthesis in Streptomyces Nodosus: Deductions from Analysis of Polyketide Synthase and Late Genes. *Chem. & Biol.* **2003**, *10* (1), 93-94.

(11) Kapur, S.; Lowry, B.; Yuzawa, S.; Kenthirapalan, S.; Chen, A. Y.; Cane, D. E.; Khosla, C. Reprogramming a Module of the 6-Deoxyerythronolide B Synthase for Iterative Chain Elongation. *Proc. Nat. Acad. Sci.* **2012**, *109* (11), 4110-4115.

(12) Floss, H. G. Combinatorial Biosynthesis-Potential and Problems. *J. of Biotech*. **2006**, *124* (1), 242–257.

(13) Tao, H.; Mori, T.; Wei, X.; Matsuda, Y.; Abe, I. One Polyketide Synthase, Two Distinct Products: Trans-Acting Enzyme-Controlled Product Divergence in Calbistrin Biosynthesis. *Angewandte Chemie - International Edition* **2021**, *60* (16), 8851–8858.

(14) Khosla, C.; Tang, Y.; Chen, A. Y.; Schnarr, N. A.; Cane, D. E. Structure and Mechanism of the 6-Deoxyerythronolide b Synthase. *Annual Review of Biochemistry*. **2007**, *76* (1), 195–221.

(15) Weissman, K. J.; Hong, H.; Oliynyk, M.; Siskos, A. P.; Leadlay, P. F. Identification of a Phosphopantetheinyl Transferase for Erythromycin Biosynthesis in Saccharopolyspora Erythraea. *ChemBioChem* **2004**, *5* (1), 116–125.

(16) Perham, R. N. *SWINGING ARMS AND SWINGING DOMAINS IN MULTIFUNCTIONAL ENZYMES: Catalytic Machines for Multistep Reaction*. **2000**, *69* (1), 961-1004.

(17) Keatinge-Clay, A. T.; Stroud, R. M. The Structure of a Ketoreductase Determines the Organization of the β-Carbon Processing Enzymes of Modular Polyketide Synthases. *Structure* **2006**, *14* (4), 737–748.

(18) Chen, H.; Du, L. Iterative Polyketide Biosynthesis by Modular Polyketide Synthases in Bacteria. *Applied Microbiology and Biotechnology*. **2016**, *100* (2), 541–557.

(19) Cox, R. J. Polyketides, Proteins and Genes in Fungi: Programmed Nano-Machines Begin to Reveal Their Secrets. *Org. Biomol. Chem.* **2007**, *5* (13), 2010–2026.

(20) Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angewandte Chemie - International Edition*. **2009**, *48* (26), 4688–4716.

(21) Cortes, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F. An Unusually Large MuHifunctional Polypeptide in the Erythromycin-Producing Polyketide Synthase of Saccharopolyspora Erythraea. *Nature.* **1990**, *348* (6297), 176-178.

(22) Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Takahashi, C.; Shinose, M.; Takahashi, Y.; Horikawa, H.; Nakazawa, H.; Osonoe, T.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M. Genome Sequence of an Industrial Microorganism Streptomyces Avermitilis: Deducing the Ability of Producing Secondary Metabolites. *Proc. Nat. Acad. Sci.* **2001**, *98* (21), 12215-12220.

(23) Crawford, J. M.; Townsend, C. A. New Insights into the Formation of Fungal Aromatic Polyketides. *Nature Reviews Microbiology*. **2010**, *8* (12), 879–889.

(24) Fujii, I. Functional Analysis of Fungal Polyketide Biosynthesis Genes. *Journal of Antibiotics*. **2010**, *63* (5), 207–218.

(25) Austin, M. B.; Noel, J. P. The Chalcone Synthase Superfamily of Type III Polyketide Synthases. *Natural Product Reports*. **2003**, *20* (1), 79–110.

(26) Tseng, C. C.; McLoughlin, S. M.; Kelleher, N. L.; Walsh, C. T. Role of the Active Site Cysteine of DpgA, a Bacterial Type III Polyketide Synthase. *Biochemistry* **2004**, *43* (4), 970–980.

(27) Watanabe, K. Exploring the Biosynthesis of Natural Products and Their Inherent Suitability for the Rational Design of Desirable Compounds through Genetic Engineering. *Bioscience, Biotechnology and Biochemistry*. **2008**, *72* (10), 2491–2506.

(28) Shen, B.; Thorson, J. S. Expanding Nature's Chemical Repertoire through Metabolic Engineering and Biocatalysis. *Current Opinion in Chemical Biology*. **2012**, *16* (1-2), 99–100.

(29) Malico, A. A.; Nichols, L.; Williams, G. J. Synthetic Biology Enabling Access to Designer Polyketides. *Current Opinion in Chemical Biology*. **2020**, *58,* 45–53.

(30) Cane, D. E.; Kudo, F.; Kinoshita, K.; Khosla, C. Precursor-Directed Biosynthesis: Biochemical Basis of the Remarkable Selectivity of the Erythromycin Polyketide Synthase toward Unsaturated Triketides. *Chem. & Biol.* **2002**, 9 (1), 131-142.

(31) Wu, J.; Kinoshita, K.; Khosla, C.; Cane, D. E. Biochemical Analysis of the Substrate Specificity of the β-Ketoacyl-Acyl Carrier Protein Synthase Domain of Module 2 of the Erythromycin Polyketide Synthase. *Biochemistry* **2004**, *43* (51), 16301–16310.

(32) Kinoshita, K.; Khosla, C.; Cane, D. E. Precursor-Directed Biosynthesis: Stereospecificity for Branched-Chain Diketides of the β-Ketoacyl-ACP Synthase Domain 2 of 6-Deoxyerythronolide B Synthase. *Helvetica Chimica Acta.* **2003**, *86* (12), 3889-3907.

(33) Wu, N.; Kudo, F.; Cane, D. E.; Khosla, C. Analysis of the Molecular Recognition Features of Individual Modules Derived from the Erythromycin Polyketide Synthase. *J. Am. Chem. Soc.* **2000**, *122* (20), 4847–4852.

(34) Khosla, C. Structures and Mechanisms of Polyketide Synthases. *Journal of Organic Chemistry*. **2009**, *74* (17), 6416–6420.

(35) Tsuji, S. Y.; Cane, D. E.; Khosla, C. Selective Protein-Protein Interactions Direct Channeling of Intermediates between Polyketide Synthase Modules. *Biochemistry* **2001**, *40* (8), 2326–2331.

(36) Zheng, J.; Taylor, C. A.; Piasecki, S. K.; Keatinge-Clay, A. T. Structural and Functional Analysis of A-Type Ketoreductases from the Amphotericin Modular Polyketide Synthase. *Structure* **2010**, *18* (8), 913–922.

(37) Kao, C. M.; Guanglin Luo, la; Leonard Katz, lb; David Cane, lc E.; Khosla, C. Engineered Biosynthesis of a Triketide Lactone from an Incomplete Modular Polyketide Synthase. *J. Am. Chem. Soc.*, **1994**, 116 (25), 11612-11613.

(38) Hagen, A.; Poust, S.; de Rond, T.; Fortman, J. L.; Katz, L.; Petzold, C. J.; Keasling, J. D. Engineering a Polyketide Synthase for in Vitro Production of Adipic Acid. *ACS Synth Biol* **2016**, *5* (1), 21–27.

(39) Kim, C. Y.; Alekseyev, V. Y.; Chen, A. Y.; Tang, Y.; Cane, D. E.; Khosla, C. Reconstituting Modular Activity from Separated Domains of 6-Deoxyerythronolide B Synthase. *Biochemistry* **2004**, *43* (44), 13892–13898.

(40) D'Agostino, P. M.; Gulder, T. A. M. Direct Pathway Cloning Combined with Sequence- and Ligation-Independent Cloning for Fast Biosynthetic Gene Cluster Refactoring and Heterologous Expression. *ACS Synth. Biol.* **2018**, *7* (7), 1702–1708.

(41) Wang, W.; Li, S.; Li, Z.; Zhang, J.; Fan, K.; Tan, G.; Ai, G.; Lam, S. M.; Shui, G.; Yang, Z.; Lu, H.; Jin, P.; Li, Y.; Chen, X.; Xia, X.; Liu, X.; Dannelly, H. K.; Yang, C.; Yang, Y.; Zhang, S.; Alterovitz, G.; Xiang, W.; Zhang, L. Harnessing the Intracellular

Triacylglycerols for Titer Improvement of Polyketides in Streptomyces. *Nat Biotechnol* **2020**, *38* (1), 76–83.

(42) Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C. Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of E. Coli. *Science.* **2001**, 291 (5509), 1790-1792.

(43) Wong, F. T.; Khosla, C. Combinatorial Biosynthesis of Polyketides-a Perspective. *Current Opinion in Chemical Biology*. **2012**, 16 (1-2), 117–123.

(44) Nivina, A.; Yuet, K. P.; Hsu, J.; Khosla, C. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chemical Reviews*. *Am. Chem. Soc.* **2019**, 119 (24), 12524–12547.

(45) Keatinge-Clay, A. T., A Tylosin Ketoreductase Reveals How Chirality is Determined in Polyketides. *Chem. & Biol.* **2007**, 14 (8), 898-908.

(46) Østergaard, L. H.; Kellenberger, L.; Cortés, J.; Roddis, M. P.; Deacon, M.; Staunton, J.; Leadlay, P. F. Stereochemistry of Catalysis by the Ketoreductase Activity in the First Extension Module of the Erythromycin Polyketide Synthase. *Biochemistry* **2002**, *41* (8), 2719–2726.

(47) Siskos, A. P.; Baerga-Ortiz, A.; Bali, S.; Stein, V.; Mamdani, H.; Spiteller, D.; Popovic, B.; Spencer, J. B.; Staunton, J.; Weissman, K. J.; Leadlay, P. F. Molecular Basis of Celmer's Rules: Stereochemistry of Catalysis by Isolated Ketoreductase Domains from Modular Polyketide Synthases. *Chem. & Biol.* **2005**, *12* (10), 1145–1153.

(48) Kao, C. M.; McPherson, M.; McDaniel, R. N.; Fu, H.; Cane, D. E.; Khosla, C. Alcohol Stereochemistry in Polyketide Backbones Is Controlled by β-Ketoreductase Domains of Modular Polyketide Synthases. *J. Am. Chem. Soc.*, **1998**, 120 (10), 2478-2479.

(49) Baerga-Ortiz, A.; Popovic, B.; Siskos, A. P.; O'Hare, H. M.; Spiteller, D.; Williams, M. G.; Campillo, N.; Spencer, J. B.; Leadlay, P. F. Directed Mutagenesis Alters the Stereochemistry of Catalysis by Isolated Ketoreductase Domains from the Erythromycin Polyketide Synthase. *Chem Biol* **2006**, *13* (3), 277–285.

(50) Chen, A. Y.; Schnarr, N. A.; Kim, C. Y.; Cane, D. E.; Khosla, C. Extender Unit and Acyl Carrier Protein Specificity of Ketosynthase Domains of the 6-Deoxyerythronolide B Synthase. *J. Am. Chem. Soc.* **2006**, *128* (9), 3067–3074.

(51) BioRender. *https://app.biorender.com/biorender-templates/t-5cf004269cf1ff003377d8cd-polymerase-chain-reaction-pcr (accessed 2023-03-27)*.

(52) LibreTexts Chemistry. *https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Kinetics/02%3A_Reaction_Rates/2.01%3A_Experimental_Determination_of_Kinetics/2.1.05%3A_Spectrophotometry; https://creativecommons.org/licenses/by-nc-sa/4.0/(accessed 2023-03-27)*.

(53) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat Methods* **2009**, *6* (5), 343–345.

(54) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Nat. Acad. Sci.* **1977**, 74 (12), 5463-5467.

(55) Bradford, M. M. A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Analytical Biochemistry.* **1976**, 72 (1), 248-254.

(56) BioRender. *https://app.biorender.com/biorender-templates/t-5ef132f6c7bcd500b388a9c3-sanger-sequencing (2023-03-27)*.

(57) Integrated DNA Technologies. *https://www.idtdna.com/pages/tools/codon-optimization-tool*.

(58) Diva. *https://public-diva.jbei.org/*.

(59) Madeira, F.; Pearce, M.; Tivey, A. R. N.; Basutkar, P.; Lee, J.; Edbali, O.; Madhusoodanan, N.; Kolesnikov, A.; Lopez, R. Search and Sequence Analysis Tools Services from EMBL-EBI in 2022. *Nucleic Acids Res.* **2022**, *50* (W1), W276–W279.

(60) BLAST. National Library of Medicine. *https://blast.ncbi.nlm.nih.gov/Blast.cgi*.

(61) Mcguffin, L. J.; Adiyaman, R.; Maghrabi, A. H. A.; Shuid, A. N.; Brackenridge, D. A.; Nealon, J. O.; Philomina, L. S. IntFOLD: An Integrated Web Resource for High Performance Protein Structure and Function Prediction. *Nucleic Acids Res* **2019**, *47* (W1), W408–W413.

(62) McGuffin, L. J.; Shuid, A. N.; Kempster, R.; Maghrabi, A. H. A.; Nealon, J. O.; Salehe, B. R.; Atkins, J. D.; Roche, D. B. Accurate Template-Based Modeling in CASP12 Using the IntFOLD4-TS, ModFOLD6, and ReFOLD Methods. *Proteins: Structure, Function and Bioinformatics.* **2018**, *86* (S1), 335–344.

(63) Alekseyev, V. Y.; Liu, C. W.; Cane, D. E.; Puglisi, J. D.; Khosla, C. Solution Structure and Proposed Domain-Domain Recognition Interface of an Acyl Carrier Protein Domain from a Modular Polyketide Synthase. *Protein Science* **2007**, *16* (10), 2093–2107.

(64) Donadio, S.; Katz, L. Organization of the Enzymatic Domains in the Multifunctional Polyketide Synthase Involved in Erythro-Mycin Formation in Saccharopolyspora Erythraea. *Gene*, **1992,** 111 (1), 51-60.

(65) Joseph, R. E.; Andreotti, A. H. Bacterial Expression and Purification of Interleukin-2 Tyrosine Kinase: Single Step Separation of the Chaperonin Impurity. *Protein Expr Purif* **2008**, *60* (2), 194–197.

(66) Meluzzi, D.; Zheng, W. H.; Hensler, M.; Nizet, V.; Dorrestein, P. C. Top-down Mass Spectrometry on Low-Resolution Instruments: Characterization of Phosphopantetheinylated Carrier Domains in Polyketide and Non-Ribosomal Biosynthetic Pathways. *Bioorg. Med. Chem.* **2008**, *18* (10), 3107–3111.

APPENDIX

DNA sequence of AmpKR2ACP2

ATGGACGCCCTTCGCTACCATATCGAATGGAACCGCGTGGCAGAACCTGGGACTGC
GCGTCCCGCCGGACGCCTGCTGGCGGTTATTTCCCCCGATCACGCCGGGGCCCCCTG
GGTAACTGCGGTGTTAGACGCACTGGGGCCAGATACAGTACGTTTTGAAGCCAAGG
GCACAGACCGCGCCGCATGGGCGGCTCAACTGGCTCAATTGCGCGAAGACGAGGGA
GAGTTTCATGCAGTAGTTAGTTTATTAGCGGCGGCTGAGGCGCTTCATACCGACCAC
GGCTCAGTCCCGCTTGGGCTGGCGCAAACTTTATTGCTTGCACAAGCCCTTGGCGAT
GCTGGGCTTACGGCTCCGTTGTGGTGTCTTACCCGCGGTGGCGTAGCCGCGGGCCGC
GGTGACGTACTGTCCTCTCCGGTTCAGGGTGCTTTGTGGGGTCTTGGCCGTGTTATTG
GACTTGAACACCCAGATCGTTGGGGCGGTCTTATTGATTTACCAGAGACGGTAGACA
CACGTGCTGCAGCCCGTTTGACGGGCCTTTTAGCTGATGCGGGCGGTGAAGATCAAT
TAGCTATTCGTGGGAGCGGTGTGTTAGCGCGCCGTTTAGCCCATGCCGCCCCAGCCG
TACCTGGGTCCGGTAAACGTCCTCCAGTGCATGGGTCTGTCTTGGTCACAGGAGGGA
CCGGAGGCATTGGCGGCCGTGTTGCGCGTCGCTTGGCAGAGCAAGGCGCAGCCCAT
CTTGTTCTGACTAGTCGCCGCGGCGCAGACGCTCCAGGTGCTGCAGAGCTTCGTGCT
GAGTTGGAGCAATTGGGAGTTCGCGTCACCATCGCCGCGTGCGACGCAGCAGACCG
TGAAGCCCTTGCTGCGTTGCTGGCGGAGTTGCCGGAAGACGCCCCATTAACGGCGGT
CTTCCACAGCGCAGGAGTAGCACATGACGATGCACCGGTAGCAGACTTAACTCTTG
GACAGTTAGATGCATTGATGCGCGCGAAACTGACGGCTGCCCGTCATTTACACGAAT
TGACGGCGGACCTGGATCTTGATGCCTTTGTTCTTTTTTCCAGTGGAGCAGCGGTGTT
TGGATCCGGCGGGCAACCTGGTTATGCGGCAGCTAACGCGTATCTGGACGCACTTGC
GGAACACCGCCGCAGCCTTGGCCTGACTGCCTCTTCCGTAGCCTGGGGGACTTGGGG
TGAAGTCGGGATGGCGACAGATCCCGAGGTACACGACCGTCTTGTACGTCAGGGTG
TCCTGGCTATGGAACCCGAGCATGCTCTGGGCGCACTGGATCAGATGCTTGAGAACG
ACGACACTGCTGCTGCCATTACTTTAATGGATTGGGAAATGTTTGCGCCGGCCTTCA
CGGCTAACCGTCCCTCGGCTCTGCTTAGTACTGTTCCCGAAGCCGTAAGTGCGCTGT
CCGATGAAGGTCCCGCTGATGGAGGCGCGGATGCAGCAGTACCACCACTGCGTGCC
CGTCTTGACGGATTGCCTACCGCCGAACGCGCCCGTGCCCTTGTGGAAGCAGTTCGT
GCTGAAGCGTCAGCGACTCTTGGCCACGATTCACCTGACGCCGTTCCAGCCTCGCGC

GCCTTCCGTGATGTGGGATTCGACTCCGTCACTGCCGTAGAATTACGTAACCGTTTA
CGTGCTGCACTGGGACTGCCATTGCCTGCGGCGTTAGTCTTTGATTATCCCACGCCA
ACAGCTCTGGCGCAACATATCGGGAGTTTATTATACGGGACCGCTCCAGATGAGGC
AGACACGGGACGTGCAGATGACCCTGACGCGCGTATTCGTGAAGCTCTTGCAACGG
TCCCAATCGGGCGCCTGCGCAAGGCCGGTTTACTTGACATGGTTCTGAAACTGGCCG
AGAGTGACGATACCGAAGACTCCGCGACAGCGACAGATGATGCGGCAGATGCTCTG
GATTTAGACGACATGGACGCGGAAGCATTGTTGCGCCTTGCTGCCGAGAATTCCGCC
AAT

(Black, AmpKR2; red, AmpACP2).

Protein sequence of AmpKR2ACP2

MDALRYHIEWNRVAEPGTARPAGRLLAVISPDHAGAPWVTAVLDALGPDTVRFEAKGT
DRAAWAAQLAQLREDEGEFHAVVSLLAAAEALHTDHGSVPLGLAQTLLLAQALGDAG
LTAPLWCLTRGGVAAGRGDVLSSPVQGALWGLGRVIGLEHPDRWGGLIDLPETVDTRA
AARLTGLLADAGGEDQLAIRGSGVLARRLAHAAPAVPGSGKRPPVHGSVLVTGGTGGI
GGRVARRLAEQGAAHLVLTSRRGADAPGAAELRAELEQLGVRVTIAACDAADREALA
ALLAELPEDAPLTAVFHSAGVAHDDAPVADLTLGQLDALMRAKLTAARHLHELTADLD
LDAFVLFSSGAAVFGSGGQPGYAAANAYLDALAEHRRSLGLTASSVAWGTWGEVGMA
TDPEVHDRLVRQGVLAMEPEHALGALDQMLENDDTAAAITLMDWEMFAPAFTANRPS
ALLSTVPEAVSALSDEGPADGGADAAVPPLRARLDGLPTAERARALVEAVRAEASATL
GHDSPDAVPASRAFRDVGFDSVTAVELRNRLRAALGLPLPAALVFDYPTPTALAQHIGS
LLYGTAPDEADTGRADDPDARIREALATVPIGRLRKAGLLDMVLKLAESDDTEDSATAT
DDAADALDLDDMDAEALLRLAAENSAN

(Black, AmpKR2; red, AmpACP2).

DNA sequence of EryKR1ACP1

TCCACCGAGGTCGACGAGGTTTCCGCGCTGCGCTACCGCATCGAGTGGCGGCCCACC
GGTGCCGGTGAACCCGCCCGGCTCGACGGCACCTGGCTGGTGGCGAAGTACGCCGG
AACCGCGGACGAGACGAGCACCGCGGCTCGGGAGGCCCTGGAGTCGGCCGGGGCG
CGGGTCCGCGAACTGGTCGTGGACGCCCGCTGCGGTCGCGACGAACTCGCGGAGCG

GCTTCGTTCGGTCGGCGAGGTGGCAGGAGTGCTGTCCCTGCTCGCGGTGGACGAAGC
GGAGCCGGAGGAGGCGCCGCTCGCGCTGGCTTCGCTGGCGGACACGCTCAGCCTCG
TGCAGGCGATGGTGTCGGCCGAACTCGGATGTCCGCTGTGGACGGTGACGGAAAGC
GCCGTCGCGACGGGGCCGTTCGAACGCGTCCGCAACGCCGCCCACGGCGCCCTGTG
GGGCGTCGGGCGGGTCATCGCGCTGGAGAACCCCGCCGTGTGGGGCGGCCTGGTCG
ACGTGCCCGCGGGGTCGGTCGCCGAGCTGGCCCGGCACCTCGCGGCGGTCGTGTCC
GGCGGCGCCGGTGAGGACCAGCTCGCGCTGCGCGCCGACGGGGTGTACGGACGCCG
GTGGGTGCGCGCGGCGGCCCCGGCGACCGATGACGAGTGGAAACCCACCGGAACCG
TGCTGGTCACCGGTGGCACGGGCGGTGTCGGCGGGCAGATCGCGCGCTGGCTCGCC
CGGCGGGGCGCGCCCCACCTGCTGCTGGTGAGCCGCAGCGGGCCGGACGCGGACGG
CGCCGGCGAACTGGTCGCCGAGCTCGAGGCGCTGGGCGCCCGGACGACCGTCGCGG
CCTGCGACGTGACCGACCGCGAGTCGGTTCGCGAGCTGCTCGGCGGCATCGGTGAC
GACGTCCCGCTCTCGGCGGTGTTCCACGCCGCCGCCACGCTCGACGACGGCACCGTG
GACACCCTCACCGGCGAGCGCATCGAGCGGGCAAGTCGCGCCAAGGTGCTCGGCGC
GCGCAACCTGCACGAGCTGACGCGCGAGCTGGACCTGACCGCCTTCGTGCTGTTCTC
GTCCTTCGCCTCGGCCTTCGGCGCCCCCGGGCTCGGCGGCTACGCGCCCGGCAACGC
CTACCTCGACGGCCTCGCCCAGCAGCGGCGGAGCGACGGACTCCCCGCGACCGCCG
TGGCGTGGGGGACGTGGGCGGGCAGCGGGATGGCCGAAGGCCCGGTGGCCGACCG
CTTCCGCAGGCACGGCGTCATCGAGATGCCTCCCGAGACGGCCTGCCGGGCGTTGC
AGAACGCGCTGGACCGCGCCGAGGTCTGCCCGATCGTCATCGACGTCAGGTGGGAC
CGGTTCCTGCTCGCCTACACCGCGCAGCGGCCGACCAGGCTCTTCGACGAGATCGAC
GACGCGCGGCGGGCTGCGCCGCAGGCGGCGGCCGAACCGCGGGTGGGCGCGCTGG
CGTCGCTGCCCGCGCCGGAGCGCGAGAAAGCGCTGTTCGAGCTCGTGCGCTCGCAC
GCGGCCGCCGTCCTCGGCCACGCCTCGGCCGAGCGGGTGCCCGCCGACCAGGCCTT
CGCGGAACTCGGCGTCGACTCGCTGTCGGCGCTTGAGCTGCGCAACCGGCTCGGCG
CCGCGACCGGTGTCCGCCTGCCGACGACGACCGTCTTCGACCACCCCGACGTGCGG
ACGCTGGCGGCGCACCTGGCCGCCGAACTCGGCGGTGCGACCGGAGCCGAGCAGGC
GGCACCGGCGACCACGGCCCCCGTCGACGAGCCGATCGCGATCGTCGGCATGGCGT
GCCGGCTGCCCGGGGAGGTCGACTCCCCGGAGCGGCTGTGGGAGCTGATCACCTCG
GGACGCGACTCCGCGGCGGAGGTCCCCGATGACCGGGGCTGGGTCCCCGACGAGCT

GATGGCCTCCGACGCGGCGGGAACCCGCCGCGCCCACGGCAACTTCATGGCGGGCG
CCGGTGACTTCGACGCGGCGTTGAGATCCGGCTGC

(Black, EryKR1; red, EryACP1).

Protein sequence of EryKR1ACP1

STEVDEVSALRYRIEWRPTGAGEPARLDGTWLVAKYAGTADETSTAAREALESAGARV
RELVVDARCGRDELAERLRSVGEVAGVLSLLAVDEAEPEEAPLALASLADTLSLVQAM
VSAELGCPLWTVTESAVATGPFERVRNAAHGALWGVGRVIALENPAVWGGLVDVPAG
SVAELARHLAAVVSGGAGEDQLALRADGVYGRRWVRAAAPATDDEWKPTGTVLVTG
GTGGVGGQIARWLARRGAPHLLLVSRSGPDADGAGELVAELEALGARTTVAACDVTD
RESVRELLGGIGDDVPLSAVFHAAATLDDGTVDTLTGERIERASRAKVLGARNLHELTR
ELDLTAFVLFSSFASAFGAPGLGGYAPGNAYLDGLAQQRRSDGLPATAVAWGTWAGS
GMAEGPVADRFRRHGVIEMPPETACRALQNALDRAEVCPIVIDVRWDRFLLAYTAQRP
TRLFDEIDDARRAAPQAAAEPRVGALASLPAPEREKALFELVRSHAAAVLGHASAERVP
ADQAFAELGVDSLSALELRNRLGAATGVRLPTTTVFDHPDVRTLAAHLAAELGGATGA
EQAAPATTAPVDEPIAIVGMACRLPGEVDSPERLWELITSGRDSAAEVPDDRGWVPDEL
MASDAAGTRRAHGNFMAGAGDFDAALRSGC

(Black, EryKR1; red, EryACP1).

DNA sequence of EryKR1AmpACP2

TCCACCGAGGTCGACGAGGTTTCCGCGCTGCGCTACCGCATCGAGTGGCGGCCCACC
GGTGCCGGTGAACCCGCCCGGCTCGACGGCACCTGGCTGGTGGCGAAGTACGCCGG
AACCGCGGACGAGACGAGCACCGCGGCTCGGGAGGCCCTGGAGTCGGCCGGGGCG
CGGGTCCGCGAACTGGTCGTGGACGCCCGCTGCGGTCGCGACGAACTCGCGGAGCG
GCTTCGTTCGGTCGGCGAGGTGGCAGGAGTGCTGTCCCTGCTCGCGGTGGACGAAGC
GGAGCCGGAGGAGGCGCCGCTCGCGCTGGCTTCGCTGGCGGACACGCTCAGCCTCG
TGCAGGCGATGGTGTCGGCCGAACTCGGATGTCCGCTGTGGACGGTGACGGAAAGC
GCCGTCGCGACGGGGCCGTTCGAACGCGTCCGCAACGCCGCCCACGGCGCCCTGTG
GGGCGTCGGGCGGGTCATCGCGCTGGAGAACCCCGCCGTGTGGGGCGGCCTGGTCG
ACGTGCCCGCGGGGTCGGTCGCCGAGCTGGCCCGGCACCTCGCGGCGGTCGTGTCC

GGCGGCGCCGGTGAGGACCAGCTCGCGCTGCGCGCCGACGGGGTGTACGGACGCCG
GTGGGTGCGCGCGGCGGCCCCGGCGACCGATGACGAGTGGAAACCCACCGGAACCG
TGCTGGTCACCGGTGGCACGGGCGGTGTCGGCGGGCAGATCGCGCGCTGGCTCGCC
CGGCGGGGCGCGCCCCACCTGCTGCTGGTGAGCCGCAGCGGGCCGGACGCGGACGG
CGCCGGCGAACTGGTCGCCGAGCTCGAGGCGCTGGGCGCCCGGACGACCGTCGCGG
CCTGCGACGTGACCGACCGCGAGTCGGTTCGCGAGCTGCTCGGCGGCATCGGTGAC
GACGTCCCGCTCTCGGCGGTGTTCCACGCCGCCGCCACGCTCGACGACGGCACCGTG
GACACCCTCACCGGCGAGCGCATCGAGCGGGCAAGTCGCGCCAAGGTGCTCGGCGC
GCGCAACCTGCACGAGCTGACGCGCGAGCTGGACCTGACCGCCTTCGTGCTGTTCTC
GTCCTTCGCCTCGGCCTTCGGCGCCCCCGGGCTCGGCGGCTACGCGCCCGGCAACGC
CTACCTCGACGGCCTCGCCCAGCAGCGGCGGAGCGACGGACTCCCCGCGACCGCCG
TGGCGTGGGGGACGTGGGCGGGCAGCGGGATGGCCGAAGGCCCGGTGGCCGACCG
CTTCCGCAGGCACGGCGTCATCGAGATGCCTCCCGAGACGGCCTGCCGGGCGTTGC
AGAACGCGCTGGACCGCGCCGAGGTCTGCCCGATCGTCATCGACGTCAGGTGGGAC
CGGTTCCTGCTCGCCTACACCGCGCAGCGGCCGACCAGGCTCTTCGACGAGATCGAC
GACGCGCGGCGGGCTGCGCCGCAGGCGGCGGCCGAACCGCGGGTGGGCGCGCTTGA
CGGATTGCCTACCGCCGAACGCGCCCGTGCCCTTGTGGAAGCAGTTCGTGCTGAAGC
GTCAGCGACTCTTGGCCACGATTCACCTGACGCCGTTCCAGCCTCGCGCGCCTTCCG
TGATGTGGGATTCGACTCCGTCACTGCCGTAGAATTACGTAACCGTTTACGTGCTGC
ACTGGGACTGCCATTGCCTGCGGCGTTAGTCTTTGATTATCCCACGCCAACAGCTCT
GGCGCAACATATCGGGAGTTTATTATACGGGACCGCTCCAGATGAGGCAGACACGG
GACGTGCAGATGACCCTGACGCGCGTATTCGTGAAGCTCTTGCAACGGTCCCAATCG
GGCGCCTGCGCAAGGCCGGTTTACTTGACATGGTTCTGAAACTGGCCGAGAGTGAC
GATACCGAAGACTCCGCGACAGCGACAGATGATGCGGCAGATGCTCTGGATTTAGA
CGACATGGACGCGGAAGCATTGTTGCGCCTTGCTGCCGAGAATTCCGCCAAT
(Black, EryKR1; red, AmpACP2).

Protein sequence of EryKR1AmpACP2

STEVDEVSALRYRIEWRPTGAGEPARLDGTWLVAKYAGTADETSTAAREALESAGARV
RELVVDARCGRDELAERLRSVGEVAGVLSLLAVDEAEPEEAPLALASLADTLSLVQAM

VSAELGCPLWTVTESAVATGPFERVRNAAHGALWGVGRVIALENPAVWGGLVDVPAG
SVAELARHLAAVVSGGAGEDQLALRADGVYGRRWVRAAAPATDDEWKPTGTVLVTG
GTGGVGGQIARWLARRGAPHLLLVSRSGPDADGAGELVAELEALGARTTVAACDVTD
RESVRELLGGIGDDVPLSAVFHAAATLDDGTVDTLTGERIERASRAKVLGARNLHELTR
ELDLTAFVLFSSFASAFGAPGLGGYAPGNAYLDGLAQQRRSDGLPATAVAWGTWAGS
GMAEGPVADRFRRHGVIEMPPETACRALQNALDRAEVCPIVIDVRWDRFLLAYTAQRP
TRLFDEIDDARRAAPQAAAEPRVGALDGLPTAERARALVEAVRAEASATLGHDSPDAV
PASRAFRDVGFDSVTAVELRNRLRAALGLPLPAALVFDYPTPTALAQHIGSLLYGTAPD
EADTGRADDPDARIREALATVPIGRLRKAGLLDMVLKLAESDDTEDSATATDDAADAL
DLDDMDAEALLRLAAENSAN

(Black, EryKR1; red, AmpACP2).

DNA sequence of AmpKR2EryACP1

ATGGACGCCCTTCGCTACCATATCGAATGGAACCGCGTGGCAGAACCTGGGACTGC
GCGTCCCGCCGGACGCCTGCTGGCGGTTATTTCCCCCGATCACGCCGGGGCCCCCTG
GGTAACTGCGGTGTTAGACGCACTGGGGCCAGATACAGTACGTTTTGAAGCCAAGG
GCACAGACCGCGCCGCATGGGCGGCTCAACTGGCTCAATTGCGCGAAGACGAGGGA
GAGTTTCATGCAGTAGTTAGTTTATTAGCGGCGGCTGAGGCGCTTCATACCGACCAC
GGCTCAGTCCCGCTTGGGCTGGCGCAAACTTTATTGCTTGCACAAGCCCTTGGCGAT
GCTGGGCTTACGGCTCCGTTGTGGTGTCTTACCCGCGGTGGCGTAGCCGCGGGCCGC
GGTGACGTACTGTCCTCTCCGGTTCAGGGTGCTTTGTGGGGTCTTGGCCGTGTTATTG
GACTTGAACACCCAGATCGTTGGGGCGGTCTTATTGATTTACCAGAGACGGTAGACA
CACGTGCTGCAGCCCGTTTGACGGGCCTTTTAGCTGATGCGGGCGGTGAAGATCAAT
TAGCTATTCGTGGGAGCGGTGTGTTAGCGCGCCGTTTAGCCCATGCCGCCCCAGCCG
TACCTGGGTCCGGTAAACGTCCTCCAGTGCATGGGTCTGTCTTGGTCACAGGAGGGA
CCGGAGGCATTGGCGGCCGTGTTGCGCGTCGCTTGGCAGAGCAAGGCGCAGCCCAT
CTTGTTCTGACTAGTCGCCGCGGCGCAGACGCTCCAGGTGCTGCAGAGCTTCGTGCT
GAGTTGGAGCAATTGGGAGTTCGCGTCACCATCGCCGCGTGCGACGCAGCAGACCG
TGAAGCCCTTGCTGCGTTGCTGGCGGAGTTGCCGGAAGACGCCCCATTAACGGCGGT
CTTCCACAGCGCAGGAGTAGCACATGACGATGCACCGGTAGCAGACTTAACTCTTG

GACAGTTAGATGCATTGATGCGCGCGAAACTGACGGCTGCCCGTCATTTACACGAAT
TGACGGCGGACCTGGATCTTGATGCCTTTGTTCTTTTTTCCAGTGGAGCAGCGGTGTT
TGGATCCGGCGGGCAACCTGGTTATGCGGCAGCTAACGCGTATCTGGACGCACTTGC
GGAACACCGCCGCAGCCTTGGCCTGACTGCCTCTTCCGTAGCCTGGGGGACTTGGGG
TGAAGTCGGGATGGCGACAGATCCCGAGGTACACGACCGTCTTGTACGTCAGGGTG
TCCTGGCTATGGAACCCGAGCATGCTCTGGGCGCACTGGATCAGATGCTTGAGAACG
ACGACACTGCTGCTGCCATTACTTTAATGGATTGGGAAATGTTTGCGCCGGCCTTCA
CGGCTAACCGTCCCTCGGCTCTGCTTAGTACTGTTCCCGAAGCCGTAAGTGCGCTGT
CCGATGAAGGTCCCGCTGATGGAGGCGCGGATGCAGCAGTACCACCACTGCGTGCC
CGTCTGGCGTCGCTGCCCGCGCCGGAGCGCGAGAAAGCGCTGTTCGAGCTCGTGCG
CTCGCACGCGGCCGCCGTCCTCGGCCACGCCTCGGCCGAGCGGGTGCCCGCCGACC
AGGCCTTCGCGGAACTCGGCGTCGACTCGCTGTCGGCGCTTGAGCTGCGCAACCGGC
TCGGCGCCGCGACCGGTGTCCGCCTGCCGACGACGACCGTCTTCGACCACCCCGACG
TGCGGACGCTGGCGGCGCACCTGGCCGCCGAACTCGGCGGTGCGACCGGAGCCGAG
CAGGCGGCACCGGCGACCACGGCCCCCGTCGACGAGCCGATCGCGATCGTCGGCAT
GGCGTGCCGGCTGCCCGGGGAGGTCGACTCCCCGGAGCGGCTGTGGGAGCTGATCA
CCTCGGGACGCGACTCCGCGGCGGAGGTCCCCGATGACCGGGGCTGGGTCCCCGAC
GAGCTGATGGCCTCCGACGCGGCGGGAACCCGCCGCGCCCACGGCAACTTCATGGC
GGGCGCCGGTGACTTCGACGCGGCGTTGAGATCCGGCTGC

(Black, AmpKR2; red, EryACP1).

Protein sequence of AmpKR2EryACP1

MDALRYHIEWNRVAEPGTARPAGRLLAVISPDHAGAPWVTAVLDALGPDTVRFEAKGT
DRAAWAAQLAQLREDEGEFHAVVSLLAAAEALHTDHGSVPLGLAQTLLLAQALGDAG
LTAPLWCLTRGGVAAGRGDVLSSPVQGALWGLGRVIGLEHPDRWGGLIDLPETVDTRA
AARLTGLLADAGGEDQLAIRGSGVLARRLAHAAPAVPGSGKRPPVHGSVLVTGGTGGI
GGRVARRLAEQGAAHLVLTSRRGADAPGAAELRAELEQLGVRVTIAACDAADREALA
ALLAELPEDAPLTAVFHSAGVAHDDAPVADLTLGQLDALMRAKLTAARHLHELTADLD
LDAFVLFSSGAAVFGSGGQPGYAAANAYLDALAEHRRSLGLTASSVAWGTWGEVGMA
TDPEVHDRLVRQGVLAMEPEHALGALDQMLENDDTAAAITLMDWEMFAPAFTANRPS

ALLSTVPEAVSALSDEGPADGGADAAVPPLRARLASLPAPEREKALFELVRSHAAAVLG
HASAERVPADQAFAELGVDSLSALELRNRLGAATGVRLPTTTVFDHPDVRTLAAHLAA
ELGGATGAEQAAPATTAPVDEPIAIVGMACRLPGEVDSPERLWELITSGRDSAAEVPDD
RGWVPDELMASDAAGTRRAHGNFMAGAGDFDAALRSGC

Black, AmpKR2; red, EryACP1).

DNA sequence of EryKR1

TCCACCGAGGTCGACGAGGTTTCCGCGCTGCGCTACCGCATCGAGTGGCGGCCCACC
GGTGCCGGTGAACCCGCCCGGCTCGACGGCACCTGGCTGGTGGCGAAGTACGCCGG
AACCGCGGACGAGACGAGCACCGCGGCTCGGGAGGCCCTGGAGTCGGCCGGGGCG
CGGGTCCGCGAACTGGTCGTGGACGCCCGCTGCGGTCGCGACGAACTCGCGGAGCG
GCTTCGTTCGGTCGGCGAGGTGGCAGGAGTGCTGTCCCTGCTCGCGGTGGACGAAGC
GGAGCCGGAGGAGGCGCCGCTCGCGCTGGCTTCGCTGGCGGACACGCTCAGCCTCG
TGCAGGCGATGGTGTCGGCCGAACTCGGATGTCCGCTGTGGACGGTGACGGAAAGC
GCCGTCGCGACGGGGCCGTTCGAACGCGTCCGCAACGCCGCCCACGGCGCCCTGTG
GGGCGTCGGGCGGGTCATCGCGCTGGAGAACCCCGCCGTGTGGGGCGGCCTGGTCG
ACGTGCCCGCGGGGTCGGTCGCCGAGCTGGCCCGGCACCTCGCGGCGGTCGTGTCC
GGCGGCGCCGGTGAGGACCAGCTCGCGCTGCGCGCCGACGGGGTGTACGGACGCCG
GTGGGTGCGCGCGGCGGCCCCGGCGACCGATGACGAGTGGAAACCCACCGGAACCG
TGCTGGTCACCGGTGGCACGGGCGGTGTCGGCGGGCAGATCGCGCGCTGGCTCGCC
CGGCGGGGCGCGCCCCACCTGCTGCTGGTGAGCCGCAGCGGGCCGGACGCGGACGG
CGCCGGCGAACTGGTCGCCGAGCTCGAGGCGCTGGGCGCCCGGACGACCGTCGCGG
CCTGCGACGTGACCGACCGCGAGTCGGTTCGCGAGCTGCTCGGCGGCATCGGTGAC
GACGTCCCGCTCTCGGCGGTGTTCCACGCCGCCGCCACGCTCGACGACGGCACCGTG
GACACCCTCACCGGCGAGCGCATCGAGCGGGCAAGTCGCGCCAAGGTGCTCGGCGC
GCGCAACCTGCACGAGCTGACGCGCGAGCTGGACCTGACCGCCTTCGTGCTGTTCTC
GTCCTTCGCCTCGGCCTTCGGCGCCCCCGGGCTCGGCGGCTACGCGCCCGGCAACGC
CTACCTCGACGGCCTCGCCCAGCAGCGGCGGAGCGACGGACTCCCCGCGACCGCCG
TGGCGTGGGGGACGTGGGCGGGCAGCGGGATGGCCGAAGGCCCGGTGGCCGACCG
CTTCCGCAGGCACGGCGTCATCGAGATGCCTCCCGAGACGGCCTGCCGGGCGTTGC

AGAACGCGCTGGACCGCGCCGAGGTCTGCCCGATCGTCATCGACGTCAGGTGGGAC
CGGTTCCTGCTCGCCTACACCGCGCAGCGGCCGACCAGGCTCTTCGACGAGATCGAC
GACGCGCGGCGGGCTGCGCCGCAGGCGGCGGCCGAACCGCGGGTGGGCGCG

Protein sequence of EryKR1

STEVDEVSALRYRIEWRPTGAGEPARLDGTWLVAKYAGTADETSTAAREALESAGARV
RELVVDARCGRDELAERLRSVGEVAGVLSLLAVDEAEPEEAPLALASLADTLSLVQAM
VSAELGCPLWTVTESAVATGPFERVRNAAHGALWGVGRVIALENPAVWGGLVDVPAG
SVAELARHLAAVVSGGAGEDQLALRADGVYGRRWVRAAAPATDDEWKPTGTVLVTG
GTGGVGGQIARWLARRGAPHLLLVSRSGPDADGAGELVAELEALGARTTVAACDVTD
RESVRELLGGIGDDVPLSAVFHAAATLDDGTVDTLTGERIERASRAKVLGARNLHELTR
ELDLTAFVLFSSFASAFGAPGLGGYAPGNAYLDGLAQQRRSDGLPATAVAWGTWAGS
GMAEGPVADRFRRHGVIEMPPETACRALQNALDRAEVCPIVIDVRWDRFLLAYTAQRP
TRLFDEIDDARRAAPQAAAEPRVGA

DNA sequence of AmpKR2

ATGGACGCCCTTCGCTACCATATCGAATGGAACCGCGTGGCAGAACCTGGGACTGC
GCGTCCCGCCGGACGCCTGCTGGCGGTTATTTCCCCCGATCACGCCGGGGCCCCCTG
GGTAACTGCGGTGTTAGACGCACTGGGGCCAGATACAGTACGTTTTGAAGCCAAGG
GCACAGACCGCGCCGCATGGGCGGCTCAACTGGCTCAATTGCGCGAAGACGAGGGA
GAGTTTCATGCAGTAGTTAGTTTATTAGCGGCGGCTGAGGCGCTTCATACCGACCAC
GGCTCAGTCCCGCTTGGGCTGGCGCAAACTTTATTGCTTGCACAAGCCCTTGGCGAT
GCTGGGCTTACGGCTCCGTTGTGGTGTCTTACCCGCGGTGGCGTAGCCGCGGGCCGC
GGTGACGTACTGTCCTCTCCGGTTCAGGGTGCTTTGTGGGGTCTTGGCCGTGTTATTG
GACTTGAACACCCAGATCGTTGGGGCGGTCTTATTGATTTACCAGAGACGGTAGACA
CACGTGCTGCAGCCCGTTTGACGGGCCTTTTAGCTGATGCGGGCGGTGAAGATCAAT
TAGCTATTCGTGGGAGCGGTGTGTTAGCGCGCCGTTTAGCCCATGCCGCCCCAGCCG
TACCTGGGTCCGGTAAACGTCCTCCAGTGCATGGGTCTGTCTTGGTCACAGGAGGGA
CCGGAGGCATTGGCGGCCGTGTTGCGCGTCGCTTGGCAGAGCAAGGCGCAGCCCAT
CTTGTTCTGACTAGTCGCCGCGGCGCAGACGCTCCAGGTGCTGCAGAGCTTCGTGCT

GAGTTGGAGCAATTGGGAGTTCGCGTCACCATCGCCGCGTGCGACGCAGCAGACCG
TGAAGCCCTTGCTGCGTTGCTGGCGGAGTTGCCGGAAGACGCCCCATTAACGGCGGT
CTTCCACAGCGCAGGAGTAGCACATGACGATGCACCGGTAGCAGACTTAACTCTTG
GACAGTTAGATGCATTGATGCGCGCGAAACTGACGGCTGCCCGTCATTTACACGAAT
TGACGGCGGACCTGGATCTTGATGCCTTTGTTCTTTTTTCCAGTGGAGCAGCGGTGTT
TGGATCCGGCGGGCAACCTGGTTATGCGGCAGCTAACGCGTATCTGGACGCACTTGC
GGAACACCGCCGCAGCCTTGGCCTGACTGCCTCTTCCGTAGCCTGGGGGACTTGGGG
TGAAGTCGGGATGGCGACAGATCCCGAGGTACACGACCGTCTTGTACGTCAGGGTG
TCCTGGCTATGGAACCCGAGCATGCTCTGGGCGCACTGGATCAGATGCTTGAGAACG
ACGACACTGCTGCTGCCATTACTTTAATGGATTGGGAAATGTTTGCGCCGGCCTTCA
CGGCTAACCGTCCCTCGGCTCTGCTTAGTACTGTTCCCGAAGCCGTAAGTGCGCTGT
CCGATGAA

Protein sequence of AmpKR2

MDALRYHIEWNRVAEPGTARPAGRLLAVISPDHAGAPWVTAVLDALGPDTVRFEAKGT
DRAAWAAQLAQLREDEGEFHAVVSLLAAAEALHTDHGSVPLGLAQTLLLAQALGDAG
LTAPLWCLTRGGVAAGRGDVLSSPVQGALWGLGRVIGLEHPDRWGGLIDLPETVDTRA
AARLTGLLADAGGEDQLAIRGSGVLARRLAHAAPAVPGSGKRPPVHGSVLVTGGTGGI
GGRVARRLAEQGAAHLVLTSRRGADAPGAAELRAELEQLGVRVTIAACDAADREALA
ALLAELPEDAPLTAVFHSAGVAHDDAPVADLTLGQLDALMRAKLTAARHLHELTADLD
LDAFVLFSSGAAVFGSGGQPGYAAANAYLDALAEHRRSLGLTASSVAWGTWGEVGMA
TDPEVHDRLVRQGVLAMEPEHALGALDQMLENDDTAAAITLMDWEMFAPAFTANRPS
ALLSTVPEAVSALSDE

DNA sequence of EryACP1

GCGCTGGCGTCGCTGCCCGCGCCGGAGCGCGAGAAAGCGCTGTTCGAGCTCGTGCG
CTCGCACGCGGCCGCCGTCCTCGGCCACGCCTCGGCCGAGCGGGTGCCCGCCGACC
AGGCCTTCGCGGAACTCGGCGTCGACTCGCTGTCGGCGCTTGAGCTGCGCAACCGGC
TCGGCGCCGCGACCGGTGTCCGCCTGCCGACGACGACCGTCTTCGACCACCCCGACG
TGCGGACGCTGGCGGCGCACCTGGCCGCCGAACTCGGCGGT

Protein sequence of EryACP1

ALASLPAPEREKALFELVRSHAAAVLGHASAERVPADQAFAELGVDSLSALELRNRLGA
ATGVRLPTTTVFDHPDVRTLAAHLAAELGG

DNA sequence of AmpACP2

CCACCACTGCGTGCCCGTCTTGACGGATTGCCTACCGCCGAACGCGCCCGTGCCCTT
GTGGAAGCAGTTCGTGCTGAAGCGTCAGCGACTCTTGGCCACGATTCACCTGACGCC
GTTCCAGCCTCGCGCGCCTTCCGTGATGTGGGATTCGACTCCGTCACTGCCGTAGAA
TTACGTAACCGTTTACGTGCTGCACTGGGACTGCCATTGCCTGCGGCGTTAGTCTTTG
ATTATCCCACGCCAACAGCTCTGGCGCAACATATCGGGAGTTTATTATACGGGACCG
CTCCAGATGAG

Protein sequence of AmpACP2

PPLRARLDGLPTAERARALVEAVRAEASATLGHDSPDAVPASRAFRDVGFDSVTAVELR
NRLRAALGLPLPAALVFDYPTPTALAQHIGSLLYGTAPDE

DNA sequence of Sfp

ATGAAGATTTACGGAATTTATATGGACCGCCCGCTTTCACAGGAAGAAAATGAACG
GTTCATGTCTTTCATATCACCTGAAAAACGGGAGAAATGCCGGAGATTTTATCATAA
AGAAGATGCTCACCGCACCCTGCTGGGAGATGTGCTCGTTCGCTCAGTCATAAGCAG
GCAGTATCAGTTGGACAAATCCGATATCCGCTTTAGCACGCAGGAATACGGGAAGC
CGTGCATCCCTGATCTTCCCGACGCTCATTTCAACATTTCTCACTCCGGACGCTGGGT
CATTTGCGCGTTTGATTCACAGCCGATCGGCATAGATATCGAAAAAACGAAACCGAT
CAGCCTTGAGATCGCCAAGCGCTTCTTTTCAAAAACAGAGTACAGCGACCTTTTAGC
AAAAGACAAGGACGAGCAGACAGACTATTTTTATCATCTATGGTCAATGAAAGAAA
GCTTTATCAAACAGGAAGGCAAAGGCTTATCGCTTCCGCTTGATTCCTTTTCAGTGC
GCCTGCACCAGGACGGACAAGTATCCATTGAGCTTCCGGACAGCCATTCCCCATGCT
ATATCAAAACGTATGAGGTCGATCCCGGCTACAAAATGGCTGTATGCGCCGTACACC
CTGATTTCCCCGAGGATATCACAATGGTCTCGTACGAAGAGCTTTTA

Protein sequence of Sfp

MKIYGIYMDRPLSQEENERFMSFISPEKREKCRRFYHKEDAHRTLLGDVLVRSVISRQYQ
LDKSDIRFSTQEYGKPCIPDLPDAHFNISHSGRWVICAFDSQPIGIDIEKTKPISLEIAKRFFS
KTEYSDLLAKDKDEQTDYFYHLWSMKESFIKQEGKGLSLPLDSFSVRLHQDGQVSIELP
DSHSPCYIKTYEVDPGYKMAVCAVHPDFPEDITMVSYEELL

VITA

Mujeeb Wakeel is a Nigerian and graduated with a Bachelor of Science in Chemistry from Obafemi Awolowo University, Ile-Ife, in 2019. He came to the United States in 2021 to attend the University of Tennessee, Knoxville, for his Doctor of Philosophy in Chemistry. He graduated with a Master of Science degree when his advisor resumed a faculty position outside the United States. Mujeeb plans to attend the Georgia Institute of Technology for his Doctor of Philosophy in Chemistry.