Structural and Functional Studies of Nonribosomal Peptide Synthases

by

Kyle Swain

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2022 by the
Graduate Supervisory Committee:

Brent Nannenga, Chair
David Nielsen
Abhinav Acharya
Jeremy Mills
S. Eileen Seo

ARIZONA STATE UNIVERSITY

December 2022

ABSTRACT

The world today needs novel solutions to address current challenges in areas spanning areas from sustainable manufacturing to healthcare, and biotechnology offers the potential to help address some of these issues. One tool that offers opportunities across multiple industries is the use of nonribosomal peptide synthases (NRPSs). These are modular biological factories with individualized subunits that function in concert to create novel peptides.

One element at the heart of environmental health debates today is plastics. Biodegradable alternatives for petroleum-based plastics is a necessity. One NRPS, cyanophycin synthetase (CphA), can produce cyanophycin grana protein (CGP), a polymer composed of a poly-aspartic acid backbone with arginine side chains. The aspartic backbone has the potential to replace synthetic polyacrylate, although current production costs are prohibitive. In Chapter 2, a CphA variant from *Tatumella morbirosei* is characterized, that produces up to 3x more CGP than other known variants, and shows high iCGP specificity in both flask and bioreactor trials. Another CphA variant, this one from *Acinetobacter baylyi*, underwent rational protein design to create novel mutants. One, G217K, is 34% more productive than the wild type, while G163K produces a CGP with shorter chain lengths. The current structure refined from 4.4Å to 3.5Å.

Another exciting application of NRPSs is in healthcare. They can be used to generate novel peptides such as complex antibiotics. A recently discovered iterative polyketide synthase (IPTK), dubbed AlnB, produces an antibiotic called allenomycin. One of the modular subunits, a dehydratase named AlnB_DH, was crystallized to 2.45Å. Several mutations were created in multiple active site residues to help understand the functional

mechanism of AlnB_DH. A preliminary holoenzyme AlnB structure at 3.8Å was generated although the large disorganized regions demonstrated an incomplete structure. It was found that chain length is the primary factor in driving dehydratase action within AlnB_DH, which helps lend understanding to this module.

DEDICATION

I'd like to dedicate this work to anyone who gives up on a dream, comes to regret it, and sacrifices comfort to pursue it a second time. Life is too short for regrets.

ACKNOWLEDGMENTS

I would first and foremost like to acknowledge and extend my gratitude to my family who have been incredibly supportive throughout this entire, lengthy process. I would never have made it this far without your love, encouragement, and at times gentle reminders to get back on task.

I would like to thank my wonderful advisor, Dr. Brent Nannega. He has gone above and beyond in his support throughout my time in his lab. His willingness to help at any time, his passion for mentorship and growth in his students, and his love for knowledge have directly contributed to finishing this process. I would not be writing this dissertation without him.

As well, I would like to thank all the lab members for their help in this process: Dr. Guanhong Bu, Dr. Amar Thakar, Sagnik Sen, Alison Haymaker, Wyatt Blackson, and many others.

I would like to thank my committee for their advice and mentorship: Dr. Brent Nannenga, Dr. David Nielsen, Dr. Jeremy Mills, Dr. Abhinav Acharya, and Dr. Eileen Seo.

TABLE OF CONTENTS

Page

APPENDIX Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1: INTRODUCTION

1.1 Nonribosomal Peptide Synthases

1.1.1 Enzyme basics

In today's world, many dynamic issues such as environmental challenges or human health require modern, equally dynamic solutions. The recent explosion of biotechnology as a provider of both products and services has allowed the industry to offer novel solutions to some of these problems. In many cases, biological molecules called enzymes play a pivotal role in every step of the solution workflow. Enzymes are bioactive catalysts that regulate and accelerate chemical reactions in natural processes [1]. Within this broad category, nonribosomal peptides synthases (NRPSs) are of particular interest due to their ability to function independently of any nucleic acid 'blueprints' as long as the necessary substrates are available [2]. They can be removed from their host cell and system without any detriment to their production capabilities, making them more desirable as individual products or as part of a biomanufacturing assembly system. NRPSs are modular in nature, with separate subdomains such as an adenylation domain or a condensation domain that perform individualized functions before passing on the peptide chain to the next section of the enzyme [3]. They are excellent for producing complex small molecules such as antibiotics, tumor drugs, or specific compositions and conformations of peptide chains [4]. However, due to their size and complexity, determining the structures and functional mechanisms of these megaenzymes has proved a distant target for much of the history of protein science [2].

This is further complicated by the tendency of many NRPSs to exhibit conformational flexibility as they rearrange structural elements to accommodate the large, multi-unit

peptide chains that they produce. Several recent scientific advances have helped us understand more about the nature of the NRPSs, like the work of Reimer et. al in 2019 where they successfully crystallized a dimodular piece of an NRPS with intermediates to show that different or even similar sized intermediates caused major conformational shifts in the enzyme [5]. New and improved structural determination methods, such as cryo-Electron Microscopy (cryo-EM) are robust enough to handle these conformational changes and can allow us to piece together the structures of various NRPSs. In these works, both cryo-EM and more traditional X-ray crystallography will be used to help determine structural details about several NRPSs. In the long term, the appeal of working with modular NRPS type systems would be to design de novo megaenzymes that produce custom-ordered peptides of interest, tailored to specific applications. Recent work done by Bozhüyük et al has demonstrated a tunable assembly structure that produces a novel peptide, one of the first of its kind, shown in figure 1.1 [6].

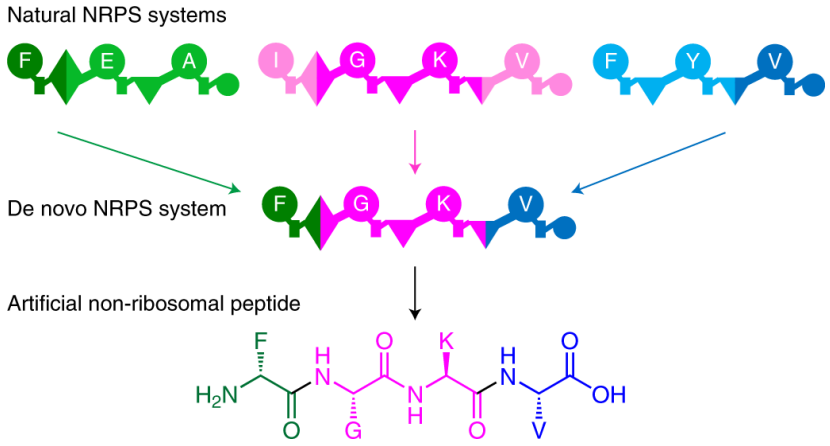


Figure 1.1: Conceptual illustration of a functional NRPS-produced, de novo peptide with modules from three different enzymes combined into a single NRPS [6]. Adapted from Bozhüyük et al.

In the following works, NRPSs from two different families will be introduced and examined, which respectively produce antibiotics and bioplastics. Deeper understandings

of how these enzymes are structurally and functionally characterized will help us learn how to repurpose their modular subunits to make new and exciting products with broad-scale applications. NRPSs represent a significant chunk of the future of biotechnology and healthcare, lending significant importance to their study and deeper knowledge of their capabilities.

1.1.2 CphA and CGP

One potential area where non ribosomal peptide synthases can make an impact is the field of plastics. Petroleum-based plastics present significant challenge to the environment due to their prolonged degradation time under native conditions [7]. One potential solution is to produce biodegradable, bio-based plastics from native sources such as cyanobacteria. One enzyme in particular, Cyanophycin synthetase (CphA), produces a poly-aspartic acid based chain which can have applications in industries from plastics to pharmaceuticals to nutritional supplements [8]. CphA is a 400kDa, tetrameric enzyme that assembles polymers composed of poly-L-aspartic acid backbones with arginine side chains, denoted as cyanophycin grana protein (CGP) [9]. The structures of three CphA variants were solved by Sharon et. al, as recently as 2021, which demonstrated that CphA actually forms tetramers composed of two dimeric subunits rather than the dimer being the dominant, functional form. An example of one variant from *Tatumella morbirosei* (called tmCphA) is displayed in figure 1.2, which Sharon et. al solved using protein crystallography to 3.1Å [9].

Figure 1.2: TmCphA dimeric structure solved to 3.1Å

CGP is expressed in multiple species, although many of the identified orthologs appear in cyanobacteria [10]. In native sources, it functions as a nutrient storage mechanism, sequestering nitrogen and carbon in the form of amino acids to be hydrolyzed later by cyanophycinase (CphB) as needed [11-13]. In figure 1.3 below, there is a proposed mechanism for CGP production and composition, as determined by analytical HPLC in the lab of Berg et. al [14].

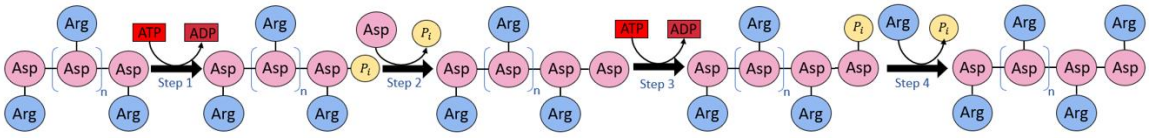


Figure 1.3: CGP production mechanism [14]

The average polymer size ranges from 25-100 kDa in cyanobacteria, but recombinant sources produce consistent 25-30 kDa sized material [15]. CphA has been found to incorporate a variety of amino acids into the side chain position, depending on the ortholog and bioavailability of substrate [16]. At times, one can find small amounts of lysine, canavanine, or citrulline in polymers, which can offer a wider variety of potential chemistry using the resulting CGP [17,18]. It was recently determined that CGP exists in two forms, denoted as soluble and insoluble, based on its behavior in water [19]. Higher lysine content in the side chains—1:4 or higher ratio of lysine to arginine—will cause the polymer to solubilize in water, while high arginine content will cause it to remain insoluble [20].

There are two classes of Cyanophycin synthetase, CphA1 and CphA2. CphA1 synthesizes CGP *de novo*, combining asp-arg together and building polymer chains from individual blocks using two distinct active sites [14]. CphA2 combines prebuilt dipeptide blocks of β-asp-arg together using a single active site [21,22]. For the purposes of this paper, we will refer to CphA1 as CphA, and CphA2 as CphA2. The works herein will solely focus on CphA1 (CphA).

Of note, neither CphA version has the ability to synthesize CGP in vitro without a starting primer [23]. It has not been determined why CphA is able to synthesize CGP *in vivo* without a starting primer but it is theorized that with enough aspartic acid and arginine present, eventually they will coincide and begin a runaway chain-reaction type process [14]. There has been one successful demonstration of a primer-independent CphA variant derived from *Thermosynechococcus elongatus* BP-1, but we have been unable to replicate this in our lab [24].

Cyanophycin displays potential to be an industrially relevant material due to its poly-aspartic acid backbone and the possibility of varying side chains, allowing for flexible platform development [25]. It also presents the potential to be a biodegradable substitute for synthetic polyacrylate, if the arginine content can be reduced enough to make pure poly-aspartic acid [26,27]. The limiting factor lies in the native species' low yields, which requires significant financial investment beyond the predicted profit margins [28]. Further work to create a more attractive option can be done by using other, more efficient expression platforms or developing new, higher value CGP derivatives. In Chapter 2, we will characterize a new, highly productive CphA variant that was first discussed by Sharon et. al, tmCphA. In Chapter 3, we will demonstrate the results of protein engineering on a CphA from *Acinetobacter baylyi* (abCphA) that results in a significant functional improvement, both in vitro and *in vivo*. Strategies like these may help bridge the cost gap needed to help bring future bioplastic producers to market.

1.1.3 AlnB and AlnB_DH

The constant evolution of health challenges today, from bacterial to viral threats, drives a parallel demand for new and improved solutions to these issues. Antibiotic resistance is on the rise and viruses are constantly evolving [29]. One potential avenue for producing new antibiotics is the use of microbial factories such as iterative polyketide synthases (iPKSs) [30,31]. These are commonly found in fungi, molds, and some bacterial strains, and can be repurposed to produce specific molecules of therapeutic interest. One such example, derived from the bacterial genus *Streptomyces*, was recently discovered by Wang et. al and found to produce allenomycin, a complex antibiotic with several potential applications [32]. The gene, sgriM25, was denoted as allenomycin biosynthetic

gene cluster, or AlnB [32]. Specifically, this is considered a multifunctional type I polyketide synthase (T1PKS), which works in either an iterative mode, or as an assembly line that stacks carbon chains [33]. These enzymes have specific subunits that perform certain tasks, such as an acyltransferase (AT), acyl carrier protein (ACP), a ketosynthase (KS), and other domains as necessary to produce the molecule such as an enoylreductase (ER) or a dehydratase (DH) [34]. As with any enzyme of particular interest, the next step after discovery is a full characterization to understand how the protein works. This would involve structural analysis to determine the folding and composition, mechanistic analysis to understand how the enzyme drives each step in the overall reaction, and physical characteristic analysis to determine long term stability and kinetic rates. For these works, our group focused on the structural determination of the overall enzyme by cryo_EM (see 1.2.2) and the novel AlnB dehydratase (AlnB_DH) by crystallography. As well, we helped develop and evaluate several mutants as part of a multi-group effort to tease out the mechanism of this dehydratase. More information and a full paper on the subject are presented in Chapter 4.

1.2 Techniques Used to Determine Enzyme Structures in these Works

1.2.1 Crystallography

One of the most common techniques for determining a protein's structure is x-ray crystallography. It was initially discovered in the early 1900s and has been used extensively to determine protein structures for molecules of various compositions and applications [35]. A general workflow is demonstrated in figure 1.4 below.

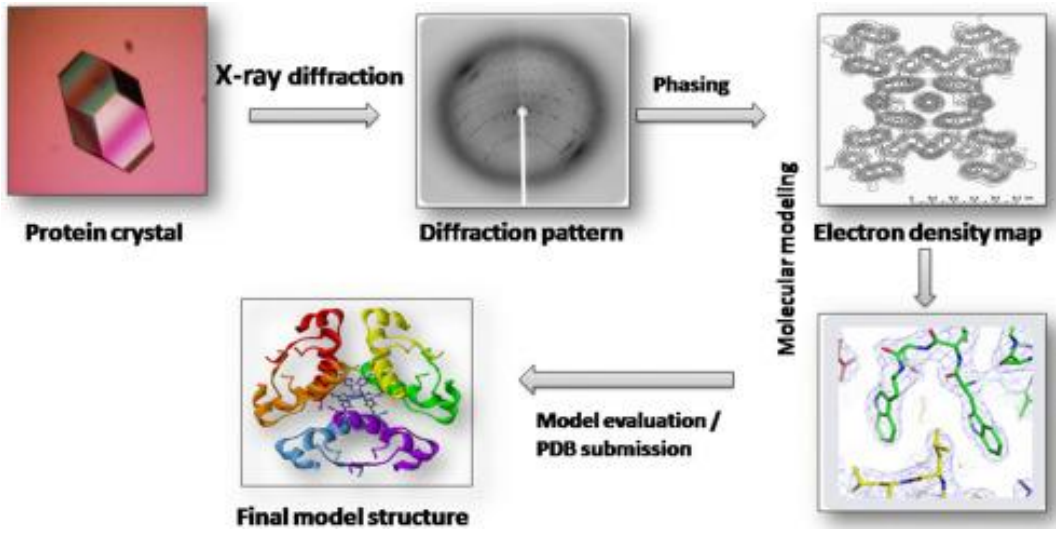


Figure 1.4: General structural determination workflow by x-ray crystallography [36]. Adapted from Gawas et. al.

Although it is only represented by a single step in the above diagram, generating high quality individual protein crystals can be a difficult, often rate-limiting step in this process. A specific combination of pH, temperature, salt ions, cofactors, or the presence of various weights and sizes of polyethylene glycol (PEG) can be contributing factors that allow the protein to nucleate. In figure 1.5 there are representations of the various zones a protein solution can inhabit.

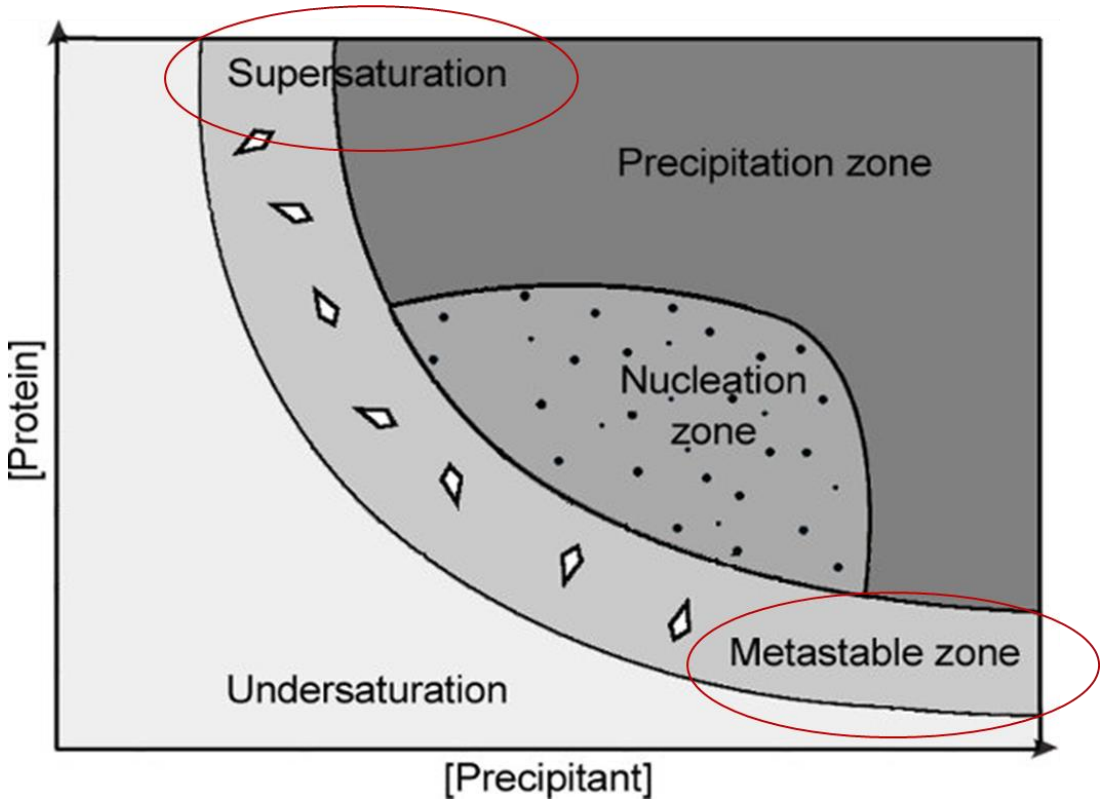


Figure 1.5: Phase diagram of protein crystallization zones[37]. Adapted from Pichlo et. al.

For the purposes of crystallization, variations in concentration and precipitant can help shift a protein solution into the ideal nucleation zone which can begin the crystallization process. As the solution progresses, the reduced available protein plus any potential buffer condition changes can help move the droplet into the metastable, supersaturated zone for crystal growth after nucleation has begun. Too much of any particular factor, though, can move a solution into the precipitation zone where the protein will aggregate rather than nucleate, and will be rendered useless for this process. Recent advances in technology, including high throughput screening techniques, has allowed labs to utilize machines to test a variety of conditions and concentrations quickly and efficiently, reducing the amount of time needed to identify crystallization conditions [38]. Often times, the hanging drop or sitting drop methods are utilized to help gradual buffer exchanges

with the crystallization solution and shift the solution from the nucleation zone to the metastable zone for crystal growth. This is shown in figure 1.6 below.

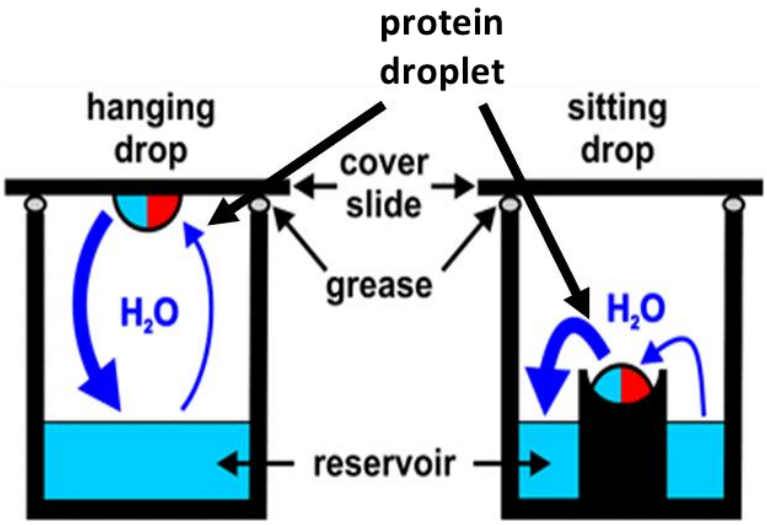


Figure 1.6: Hanging drop and sitting drop visual representation[39]. Adapted from Bijelic et. al.

In these experiments, we used the hanging drop method due to its ability to screen multiple proteins concentrations in a single well by placing drops of varying concentrations on the same coverslip.

Once multiple quality crystals have been developed, their diffraction patterns can be analyzed in order to determine the structure of the crystal. To reduce the amount of potential radiation damage during the diffraction process, crystals are often cryo-frozen and examined at subzero temperatures[40]. This also reduces the risk of temperature fluctuations which can change the crystal solution conditions, and thus destabilize some more delicate crystals and cause them to precipitate. After freezing, they are loaded into an x-ray beam. The most common method uses synchrotron radiation (SR) as a source of X-rays[41]. By rotating the crystal while it is suspended in the X-ray beam, we can generate a series of diffraction patterns which imply specific electron densities that refract the

beam of light. When combining these electron density maps with the electron densities of the amino acids in the protein sequence, we can generate structural maps of where each amino acid fits in the protein structure. The aid of computer software like Coot has dramatically increased our computational speed and accuracy in this part of the process [42,43].

1.2.2 Cryo-EM

Cryo-Electron Microscopy (cryo-EM), particularly single particle cryo-EM, has become a popular protein structural determination technique in recent years. It offers several advantages over traditional x-ray crystallography: 1) It negates the crystallization step and any necessary optimization therein, 2) It allows for the determination of larger, complex multimers whereas often times these would not crystallize well, 3) It allows for simultaneous classification of multiple orientations or reaction states from the same initial sample preparation, 4) It often requires significantly less sample and less complex storage buffers than crystallography, reducing production and screening costs [44,45]. The major drawback of this technique is that certain small (<100kDa) or highly disordered proteins are difficult to image and therefore may not be suitable for analysis with Cryo-EM. Sample preparation follows the standard expression (see Chapter 1.3.1) and purification steps (see Chapter 1.3.2). After this, protein solution is applied at various concentrations to TEM grids with holey film, blotted, and plunged into liquid ethane to create thin vitrified sheets of ice in the holes that contain individual protein molecules. A TEM is then used to collect several million images of these individual molecules. These images, called electron micrographs, are computationally processed to produce what are called particles—visual representations of a single protein molecule. They are classified

into different 2-dimensional orientation groups and those are used to create a 3D structural map. This is refined into a final structure which, when matched with the protein sequence and other homologous structures, can be used as a definitive protein structure. Often, much of this image processing and reconstruction is done in programs like CryoSPARC, which is our preferred program[46]. A sample cryo-EM workflow is described in figure 1.7.

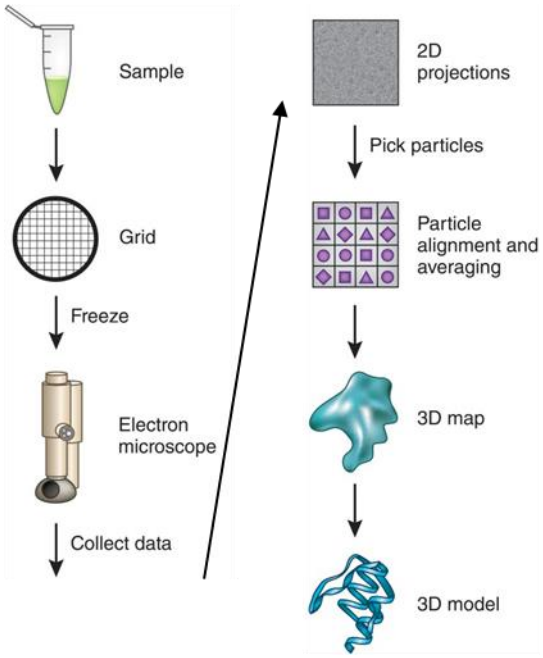


Figure 1.7: Sample workflow of a standard Cryo-EM procedure [47]. Adapted from Doerr et. al.

Consistent updates and refinement techniques have made this method a quick and powerful system for determining certain protein structures. As technology improves in this field, it will continue to overtake X-ray crystallography as a standard method and will provide insight into previously difficult-to-solve protein structures [48].

1.3 Enzyme Expression, Purification, and Characterization Techniques Utilized in these Works

1.3.1 Expression

Enzymes can be expressed in a variety of different cell lines including mammalian, insect, or bacterial cells. In many cases, it is more convenient to remove an enzyme from the native organism and express it in a more commonly-used system. This can increase protein yields, standardize growth conditions across the production of different enzymes, and reduce batch variability [49]. Usually, enzymes are transferred from one cell line to another through the use of plasmids, which contain the genetic code for a specific protein of interest. Through techniques like Gibson assembly and subcloning, we can create plasmids with our target enzyme and transform that into our production cell line. In these studies, we have selected *E. coli* as our production strain of choice for all AlnB (Chapter 2), AlnB_DH (Chapter 2), and CphA orthologs (Chapters 3 and 4) studied due to the well-established and low cost nature of this expression system [50]. For the Fenna-Matthews-Olson complex study in Chapter 5, we chose to express the protein in its native cyanobacterial strain due to the unique cellular machinery necessary to produce functional, soluble enzyme.

E. coli-based protein expression revolves around several important factors: 1) The media of choice, 2) the cell density and growth rate, and 3) the expression system [51]. In many cases, there are standard medium choices like Lysogeny Broth (LB) or Terrific Broth (TB) which offer strong nutritional profiles for many standard protein production methods. The specific medium may need to be customized depending on the anticipated results, such as selecting a limited nutritional profile when forcing uptake of radioactively-labeled amino acids for protein synthesis. In our case, we selected media based off of anticipated growth time and high nutrient availability, which lead to the use

of LB or TB throughout these studies. When growing *E. coli* for protein expression, one will often target the mid to late exponential growth phase to induce, which necessitates specific monitoring of optical density (OD) at 600nm, often called OD600. This metric is a direct representation of the cell density and can be used to determine the growth rate. For *E. coli*, the optimal OD600 for inducing expression is between 0.4-0.8 [49]. Induction after this OD600 measurement range can have consequences like reduced yields due to less nutrient availability and reduced cell replication rate. As well, it is common practice to measure OD600 at the end of the expression procedure to quantify the overall growth rate of the cells after they have produced the enzyme of interest, giving insight into whether the enzyme is toxic or well-tolerated by the strain. Once the basic media and expression rate have been determined, it is important to note the expression system of choice. To active a plasmid and induce protein production, one has to add an inducer to the cell culture. This is a chemical signaling agent which, when uptaken by the cell, removes the repressor from the plasmid and redirects cellular machinery to produce the gene of interest. In our case, we utilized the T7 expression system, the most common system for *E. coli* today, where production is induced by using a lactose-mimic sugar called Isopropyl β- d-1-thiogalactopyranoside (IPTG) [52]. Adding IPTG to the cell culture at the ideal OD600, between 0.4-0.8, then incubating at an appropriate temperature, resulted in high levels of the desired enzyme expressions.

1.3.2 Purification

Once the protein of interest has been expressed, it needs to be extracted from the cell and purified to homogeneity before it can be further studied. This eliminates several

experimental concerns including protein misidentification, yield accuracy, and potential contamination issues.

The first step in this process is to lyse the cells, which can be achieved through chemical or physical means. Often times, chemical lysis involves the use of a detergent or acid-based solution which will disrupt cell membranes and release the internal materials. In the case of CGP extraction, due to the polymer's solubility in strongly acidic solutions, this technique can be used to cheaply and efficiently remove CGP from cells [53]. In many cases though, the same value that the detergents and acids contribute to lysing cells can also destabilize and denature the enzyme of interest. For this reason, most purification methods involve physical lysis, such as homogenization or sonication [54]. Homogenization is an older technique and uses a spinning blade to shear open cell membranes by force. While effective, this method can involve significant damage to the protein of interest and also often requires significant hands-on time investment and clean-up. Sonication is a more modern version that uses ultrasonic sound waves to lyse cells through cavitation of the air in solution. The collapse of these air bubbles generates enough energy to disrupt cell membranes but will often leave internal enzymes intact and functional. For our purposes, we exclusively used sonication to extract and characterize AlnB, AlnB_DH, CphA, and FMO. Detergent-based lysis was tested with CphA but was found to interact with the enzyme and cause precipitation so that was discarded.

After lysis, the enzyme must be purified to homogeneity before conducting experiments. Often, this is done by a variety of chromatographic methods utilizing specific features of the protein of interest, such as charge (called ion-exchange chromatography), size (size-exclusion chromatography), or attaching an affinity tag (affinity chromatography) [54]. We

utilized both affinity chromatography and then size-exclusion chromatography. Our affinity tag of choice was a series of six or more consecutive histidines (6x-His) which bind with nickel-nitrilotriacetic acid resin (Ni-NTA) and allow the rest of the cell debris to pass through a chromatography column, as illustrated by figure 1.8 [55].

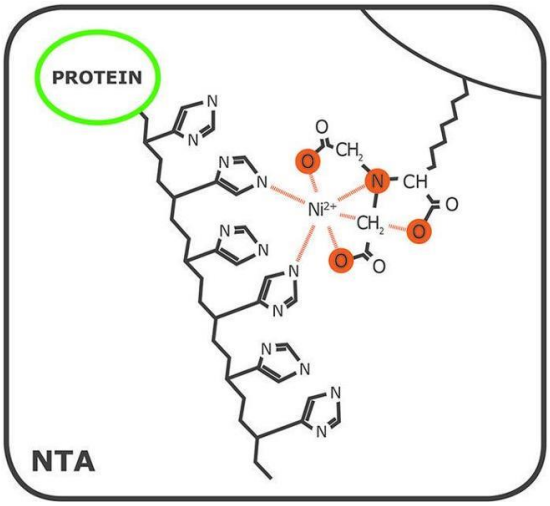


Figure 1.8: Example protein with 6x-His tag binding to nickel resin in Ni-NTA procedure [55]. Adapted from Trabbic-Carlson et. al.

Once the enzyme has been bound to the column, it can be eluted as desired through addition of imidazole or a pH differential. This yields a crudely purified enzyme. The next step, a polishing step, involves using the enzyme's relative size as compared to other proteins to help purify it further. Size-exclusion chromatography, as seen in figure 1.9, uses a mesh of resin beads with variously-sized pores to "slow down" smaller molecules and allow larger molecules to pass through the column at a faster rate and thus elute first.

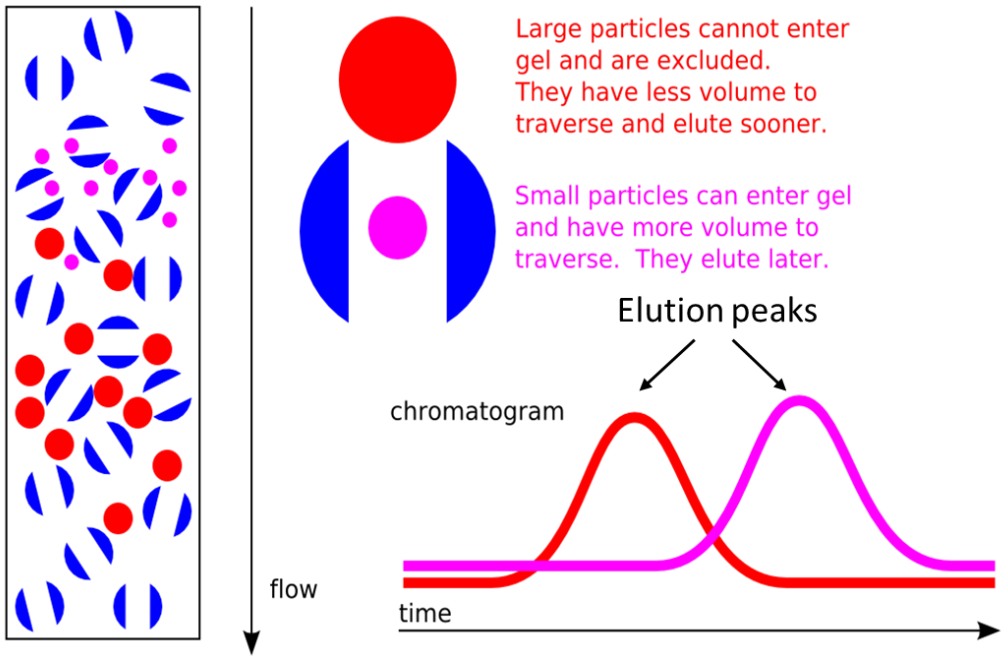


Figure 1.9: Example of an in-progress SEC run with various size molecules. Includes a chromatographic example with the red peak eluting first because the red molecule is larger than the pink molecule [56]. Adapted from Proteomics.

When combined with an in-line UV280 spectrophotometer, this provides a powerful tool for quickly and efficiently separating molecules. In summary, the combination of sonication, affinity chromatography, and size-exclusion chromatography provides a quick and effective way to purify our enzymes of interest [54].

1.3.3 Characterization

When characterizing an enzyme, one can look at both the structural and functional properties of the molecule. This can be done through a variety of different assays, although we will only discuss a few that are pertinent to the following works. Structural characterization can be divided into subcategories such as size or stability. Size-based characterization can be done through methods like size-exclusion chromatography, as discussed in 1.3.2. As well, one common method utilized is sodium

dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). This involves the use of an electric current to distribute proteins across a polyacrylamide gel based on their molecular mass, which can be checked against an internal standard to verify the relative size [57].

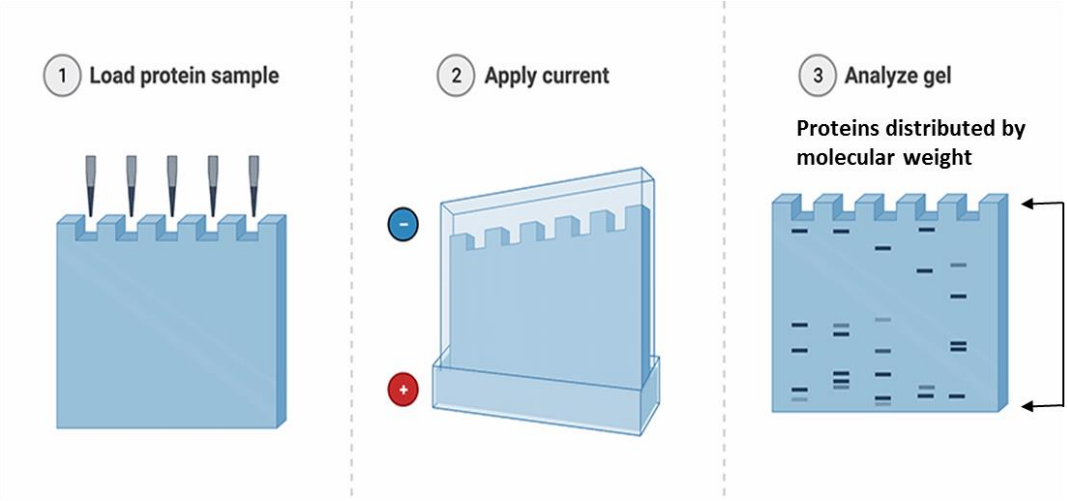


Figure 1.10: Workflow of an SDS-PAGE gel process [58]. Adapted from Immunotherapies, R.

Generally, this is used as a final confirmation step to validate that the correct enzyme has been purified from cell extract. As well, stability-based characterization will examine how long the enzyme remains folded and functional when exposed to stresses like heat, extreme pH, or time in solution. This is often useful because the longer an enzyme remains stable, the more substrate it can convert into product, and thus one can produce more material for the same cost. Techniques such as circular dichroism, thermal stability over time (quantified by SDS-PAGE), and time-based activity assays can all play roles in generating useful conclusions about the enzyme structure.

At the same time, functional assays are also useful tools. Quantifying the amount and rate of product generation that an enzyme can produce gives us useful data for understanding

how effective modifications can be. Enzyme coefficients such as the maximum enzyme velocity or substrate turnover rate, in combination with mathematical equivalences like the Michaelis-Menten equation, provide the baseline numbers for comparative analyses [59]. We can generate these empirical values through experiments that either examine the amount of product or the rate. In the case of AlnB_DH, we can ascertain the amount and type of product generated in an *in vitro* reaction to evaluate the effectiveness of specific site mutations using High Performance Liquid Chromatography (HPLC). On the other hand, with CphA, we can go two routes: 1) Use the CGP's property of light-scattering at OD600 to measure the production of polymer by the enzyme in realtime using a microwell plate reader, or 2) measure a byproduct of the reaction [9]. In this case, we can use the liberated phosphate released during the polymer production, in concert with an external dye like Malachite Green, to generate an OD reading that can be measured in a spectrophotometer. For convenience and cost reasons, OD600 was selected as the main functional assay choice. This can give us insight into the amount of polymer produced from a given starting substrate concentration, as well as information about the rate of production when OD600 is measured at regular intervals.

Overall, enzyme characterization is an important follow-up to expression and purification. This both validates that the correct molecule has been selected, and ensures that it functions as intended for future experimental work.

CHAPTER 2: HETEROLOGOUS PRODUCTION OF CYANOPHYCIN WITH *TATUMELLA MORBIROSEI* CYANOPHYCIN SYNTHETASE

Authors: Kyle Swain, Itai Sharon, Wyatt Blackson, Stefan Tekel, T. Martin Schmeing, David R. Nielsen, Brent L. Nannenga

Abstract

Microbial production of biopolymers represents a promising, sustainable alternative to current approaches for plastic production. Cyanophycin synthetase 1 (CphA), a non-ribosomal peptide synthetase, produces cyanophycin - an attractive biopolymer consisting of a poly-L-aspartic acid backbone decorated with L-arginines. In this work, a series of CphA homologues were screened in *Escherichia coli* and it was ultimately found that CphA from *Tatumella morbirosei* (TmCphA) was especially productive. In particular, we found that TmCphA was capable of supporting nearly two-fold greater yields of insoluble cyanophycin relative to other tested CphAs, whereas production in a bench-scale bioreactor yielded up to 1.8 g/L cyanophycin.

2.1 Introduction

Petroleum-based plastics present significant challenges to the environment due to their synthesis from non-renewable sources and prolonged degradation time under native conditions [7]. One solution to this problem is to produce biodegradable and renewable polymers from biological sources. Cyanophycin synthetase 1 (CphA), is a nonribosomal peptide synthetase that synthesizes cyanophycin, a biopolymer composed of poly-L-aspartic acid and poly-L-arginine subunits. A similar enzyme, cyanophycin synthetase 2, ligates prefabricated aspartic acid-arginine subunits together to form cyanophycin, although this is not as well studied [22]. Natively, cyanophycin is produced in cyanobacteria and several bacterial species as a storage mechanism for both nitrogen and carbon, and can later be hydrolyzed by a second enzyme, cyanophycinase (CphB) [11-13]. When isolated, cyanophycin can be separated into both insoluble and soluble fractions, which refers to the solubility of the polymer at neutral pH. Depending on the specific CphA homolog and expression conditions used, lysine can also be incorporated into cyanophycin in place of arginine, and the soluble fraction generally contains a higher ratio of lysine relative to the insoluble fraction[19,20]. Cyanophycin has attracted interest as a renewable biopolymer[60], however insoluble cyanophycin is particularly attractive for downstream applications due to the significantly reduced processing requirements for purification and higher levels of compositional homogeneity[10].

A prominent challenge for bioplastics production remains the high cost and low yield of most current biopolymers [28]. Therefore, it is of great interest to identify, engineer, and optimize new platforms with enhanced productivity. In this work, we have characterized cyanophycin production in engineered *E. coli* using CphA from *T.*

morbirosei DSM23827 (TmCphA), a Gram-negative rod-shaped bacterium [9,61], which we found to have enhanced yields relative to other, more commonly studied CphA homologues. TmCphA is a 400 kDa tetramer (Fig 2.1) that previously has been shown to be functionally expressed in *E. coli* and whose structure was recently determined to 3.1Å by X-ray crystallography[9].

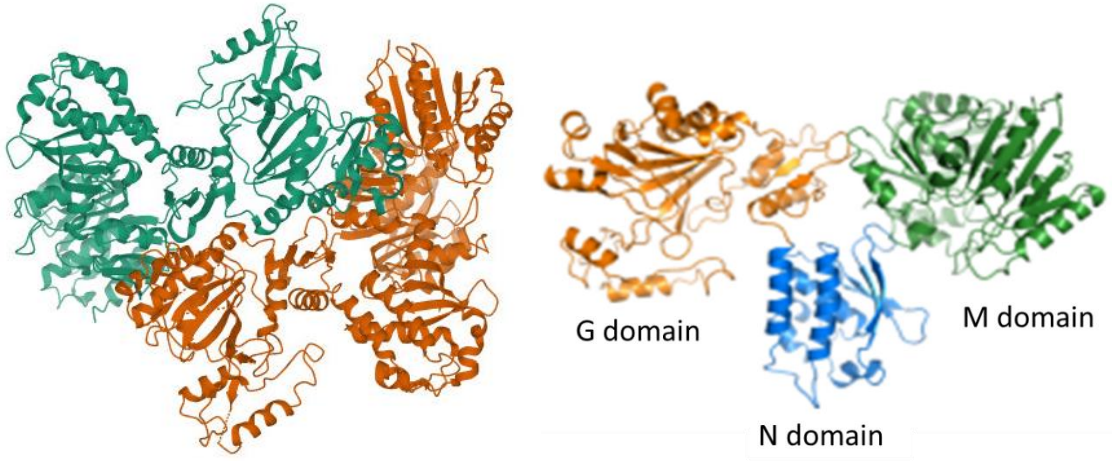


Figure 2.1: TmCphA dimer solved to 3.1A, with monomer demonstrating the separate functional domains [9]. Adapted from Sharon et. al.

Like other CphA homologues, TmCphA has three functional domains that each play an important role in the production of cyanophycin. The G domain uses ATP to catalyze the addition of aspartic acid to the end of the growing cyanophycin chain[62,63]. The role of the M domain, often compared with Mur ligases, is to add arginine, again with the use of ATP[64]. The N domain is thought to use salt bridges and hydrogen bonding to loosely anchor cyanophycin molecule in place and allow for a "flip-flop" mechanism where the polymer rebounds from the G domain to the M domain and back to build and attach individual subunits to the growing polymer chain[8,9]. The N domain has also recently been shown to be important for primer generation, which initiates cyanophycin synthesis *in*

vivo[65]. Cyanophycin has a natural size range from 25-100 kDa, with heterologous production in *E. coli* usually generating a range closer to 25-40 kDa [15].

2.2 Results and Discussion

Production of cyanophycin in *E. coli* has been demonstrated for several CphA homologues [10,28,66]; however, heterologous expression and purification of the stable enzyme in *E. coli* notable remains difficult for many CphAs [9]. This suggests that, although the synthetase is functional, it is not particularly stable when expressed in *E. coli*. Recently, CphA homologues that do express and purify as soluble and stable samples were identified and used for structure determination and functional analysis [9]. To investigate whether CphA homologues with robust and stable heterologous expression would produce enhanced levels of cyanophycin, production of cyanophycin using three different stable homologues - CphA from *Acinetobacter baylyi* DSM587 (AbCphA), *Synechocystis* sp. UTEX2470 (SuCphA), and TmCphA - was characterized and compared with that of CphA from *Synechocystis* sp. PCC6308 (6308CphA). Since *E. coli* harboring 6308CphA has been used in several past studies for heterologous production of cyanophycin [18,23,67], this strain served as a benchmark for identifying potential improvements in cyanophycin production relative to the three more stable homologues of interest. While cyanophycin production from 6308CphA is well-characterized, the 6308CphA enzyme itself does not express and purify readily, unlike the other CphA homologues studied here.

Flask scale (25 mL) cultures were first carried out for *E. coli* expressing each CphA homolog and levels of soluble, insoluble, and total cyanophycin produced were compared (Fig. 1). While the overall cyanophycin yields of 6308CphA, AbCphA, and SuCphA were similar, the more stably expressed AbCphA and SuCphA produced 3.8- and 3.1-times more insoluble CphA, respectively, relative to CphA6308. However, the greatest cyanophycin production was achieved using TmCphA which, relative to 6308CphA, enabled the production of 2.6-times more total cyanophycin and 10.8-times more insoluble cyanophycin. Even when compared to AbCphA, the next strongest producer of insoluble cyanophycin, TmCphA produced 2.9-times more insoluble cyanophycin.

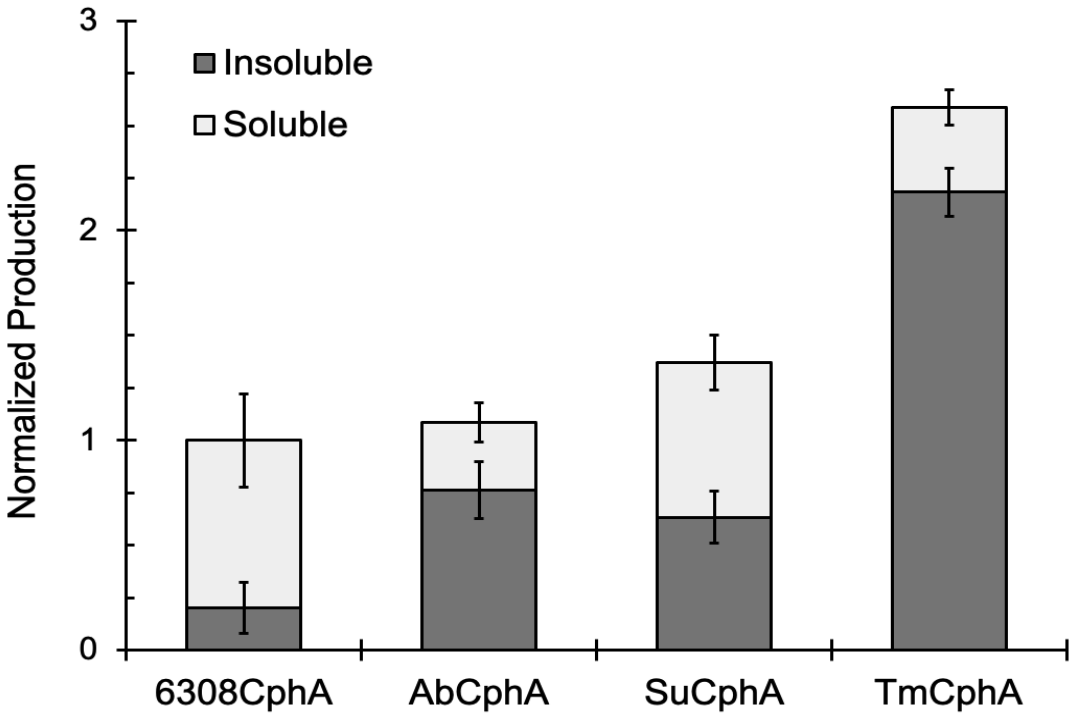


Figure 2.2: Comparison of cyanophycin production by *E. coli* strains expressing different CphA homologues. TmCphA produced more total cyanophycin, especially insoluble cyanophycin, relative to all other homologues tested. Experiments were

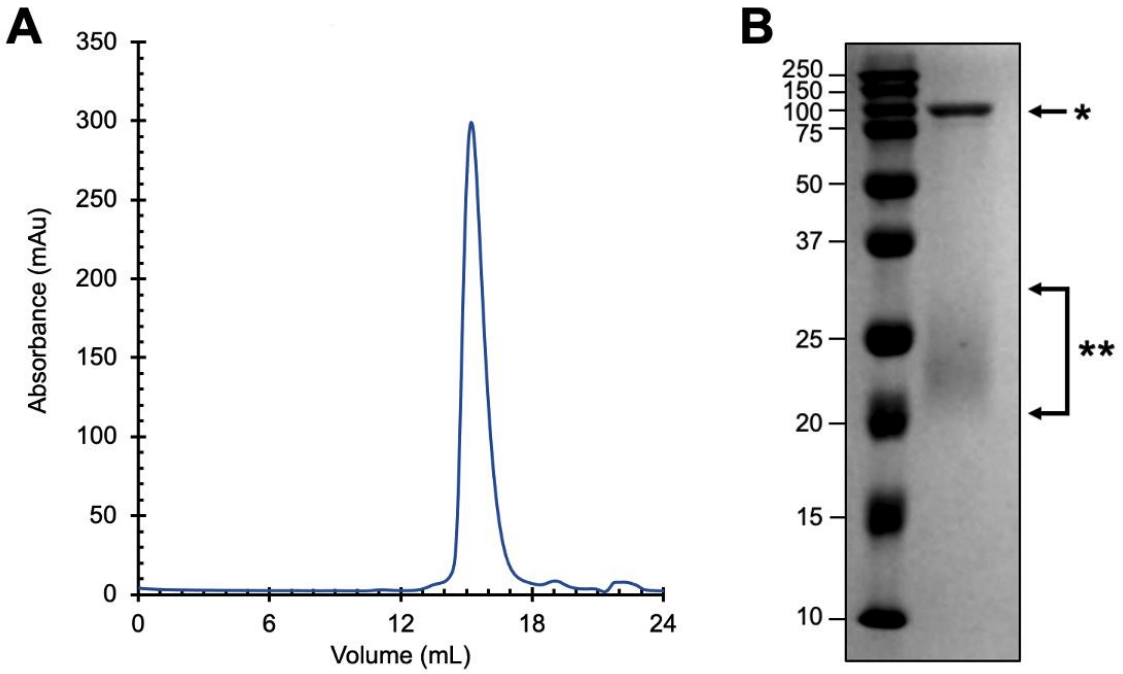
conducted in 25mL shake flasks in triplicate and normalized to total 6308CphA. Error bars represent the standard deviation.

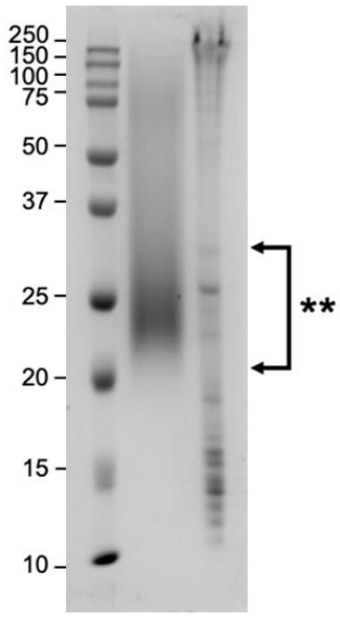
Because of its high level of production, further studies on cyanophycin production by TmCphA were conducted. TmCphA expressed in this system was confirmed by SEC to be a tetramer, as previously shown [9,65] (Fig 2A), and the functionality of purified TmCphA was confirmed by the presence of polymer in SDS-PAGE analysis of an *in vitro* assay (Fig 2B). Cyanophycin produced *in vitro* showed a size distribution of between 20 – 30 kDa, similar to what has been shown previously for TmCphA [9] and other CphA homologues [8,17,68-70].

The composition of soluble and insoluble cyanophycin produced by *E. coli* expressing TmCphA was next characterized. Here, CphB was first used to digest both soluble and insoluble cyanophycin into dipeptides subunits, which were then solubilized and analyzed by mass spectrometry (MS) (Table 1). For insoluble cyanophycin, MS confirmed the high ratio of arginine relative to lysine, with 93.9% β-Asp-Arg dipeptides and 6.1% β-Asp-Lys dipeptides. This level of arginine in the insoluble fraction produced by TmCphA is similar to what has been found in previous studies of other CphA homologues in *E. coli*, which generally ranges from 90-97% of the total side chain population [20,25]. In contrast, the soluble fraction produced by TmCphA showed a large increase in lysine content, with the β-Asp-Lys dipeptide comprising 35.4% of the total sample. An increase in lysine content was expected as previous studies have shown that higher lysine content enhances the solubility of cyanophycin near neutral pH [20].

Table 2.1. Relative composition of TmCphA produced cyanophycin determined by mass spectrometry

	β-Asp-Arg	β-Asp-Lys
Insoluble	93.9%	6.1%
Soluble	64.6%	35.4%





C:

Figure 2.3: Functional analysis of purified TmCphA. (A) The SEC chromatogram confirms that TmCphA is expressed as a stable homotetramer in *E. coli* with an approximate molecular weight of 400kDa. (B) SDS-PAGE analysis of purified TmCphA (*) shows that the enzyme is active and produces cyanophycin (**) in vitro in the presence of ATP, L-aspartic acid, L-arginine, and primer. (C) SDS-PAGE analysis of purified cyanophycin (**) from *in vivo* bioreactor production with insoluble and then soluble polymer samples.

Finally, to test the effects of improved pH control and aeration (the latter of which would help avoid potential ATP limitations at the flask scale), cyanophycin production by *E. coli* expressing TmCphA was investigated in a 1L bioreactor. In this case, dissolved oxygen levels never dropped below 80% and, after 48 hours, cells were harvested to recover both the soluble and insoluble fractions of cyanophycin. In this case, a total of 1.80 g of insoluble cyanophycin and 0.12 g of soluble cyanophycin were recovered per liter of culture. This represented a significant improvement over the 1.14g of insoluble cyanophycin and 0.18g of soluble cyanophycin produced by 1L flask trials.

2.3 Conclusion

TmCphA is a promising CphA variant supporting high level recombinant biosynthesis of cyanophycin, and which specifically enables the production of both insoluble and total cyanophycin exceeding that which has been reported for other, more commonly studied variants. In particular, using engineered *E. coli*, cyanophycin production reached as high as 1.92 g/L, with 1.80 g/L being insoluble cyanophycin. High levels of insoluble cyanophycin production are of particular interest as it is easier to process and purify at large scale relative to soluble cyanophycin. Additionally, the insoluble cyanophycin produced by TmCphA showed a high level of compositional homogeneity, with an arginine to lysine incorporation ration of 93.9% to 6.1%. Overall, TmCphA's superior solubility and high rate of expression demonstrate the strong potential for improved cyanophycin production in *E. coli*, which can be combined with future scale-up and strain engineering experiments to create a more robust source of biologically-derived polymer.

2.4 Materials and Methods

Heterologous cyanophycin production in *E. coli*

CphA from *Tatumella morbirosei* DSM 23827 (TmCphA), *Acinetobacter baylyi* DSM587 (AbCphA), and *Synechocystis* sp. UTEX2470 (SuCphA) are in pJ411-derived expression vectors [9], and CphA from *Synechocystis* sp. strain PCC 6308 (6308CphA) is cloned in pMMB206 [71]. All expression vectors were transformed into *E. coli* BL21(DE3), and these strains were used for all expression and cyanophycin production experiments.

Cyanophycin production from the different CphAs was carried out in 25 mL of Terrific Broth (TB) inside of 125 mL baffled shake flasks. TB was supplemented with 50 mg/mL kanamycin for AbCphA, SuCphA, and TmCphA, and with 35 mg/mL chloramphenicol for 6308CphA. Cultures were grown at 37°C to an OD600 of approximately 0.6, followed by induction of CphA expression by the addition of 1mM IPTG. Following induction, the cultures were shifted to 30°C for 48 hours for cyanophycin production. Cells were then harvested and cell pellets stored at -20°C for further processing.

Cyanophycin purification

Cell pellets from cyanophycin production were resuspended in 1M HCl at a ratio of 10mL of HCl per gram of wet cell weight. Pellets were briefly vortexed and then incubated in a shaker for six hours at room temperature. Samples were then centrifuged at 10,000g for 15 minutes to remove cell debris, and the supernatant was collected. To help buffer the resulting solution, Tris pH 8.0 was added to each sample to a final concentration of 0.5M Tris-HCl pH 8.0. This was followed by neutralizing the sample by adding 10M NaOH. At this stage, the insoluble fraction precipitated and was collected by centrifugation at 10,000g for 15 minutes. The supernatant, which contains the soluble cyanophycin, was added to twice the volume of ethanol, which causes the soluble cyanophycin fraction to precipitate. Soluble cyanophycin was harvested by centrifugation at 10,000g for 15 minutes. The insoluble and soluble cyanophycin fractions were lyophilized, weighed, and stored for further analysis.

Expression and purification of TmCphA

TmCphA was overexpressed using the same strain as for cyanophycin production described above. Cells were grown in LB media supplemented with 50 ug/mL kanamycin at 37ºC. Expression of TmCphA was induced at OD600 of 0.6 by the addition of 1mM IPTG, and protein expression was carried out at 30ºC for 20 hours. Cells were harvested, resuspended in buffer A (150mM NaCl, 0.5mM TCEP, 50mM Tris-HCl, pH 8.0) and lysed by sonication. The lysed sample was centrifuged at 10,000g for 15 minutes and the supernatant was flown through a 5mL Ni-NTA column along with 2% buffer B (150mM NaCl, 0.5mM TCEP, 500mM imidazole, 50mM Tris-HCl, pH 8.0). The Ni-NTA column was then washed with 80% buffer A, 20% buffer B, followed by elution with 100% buffer B. The peak fraction was collected and further purified by SEC on a Superose 6 column equilibrated in buffer A. The SEC peak fraction corresponding to approximately 400 kDa was collected for further analysis.

TmCphA Activity Assay

The assay was performed as described previously[9], with minor modification. A reaction mixture containing 12.5 mM ATP, 2 mM $MgCl_2$, 12.5 mM L-aspartic acid, 12.5 mM L-arginine, 12.5 mM KCl, 100 mM Tris-HCl pH 8.0, 1.5 mg/mL purified cyanophycin was added to purified TmCphA (100 μg/mL final). This reaction mixture with TmCphA was incubated for 2 hours, followed by analysis by SDS-PAGE.

Bioreactor production

A 1L Sartorious Biostat A Plus (Sartorious, G*ö*ttingen, Germany) bioreactor, was used to test TmCphA scale-up. 20 mL seed cultures were grown overnight at 37ºC in LB supplemented with 50µg/L kanamycin. Seed cultures were added to the bioreactor with 1L of TB, supplemented with kanamycin, and grown at 37ºC with 1 vessel volume per minute (vvm) air sparging, 500 rpm agitation, and held at pH 7.0 until the OD reached 0.5. The pH was controlled through the addition of 6M NaOH or 6M HCl. After reaching an OD of 0.5, cultures were induced with 1mM IPTG, the temperature was reduced to 30ºC, and cultures were left to grow for 48 hours with all other conditions remaining constant. After 48 hours, cells were harvested and cyanophycin was processed as described above.

Mass Spectrometry

25mg/mL insoluble and soluble cyanophycin were each resuspended in 50mM NH_4CO_3 and 15µM cyanophycinase and samples were digested overnight at room temperature with mild shaking. Peak areas (including Na+ adducts) were used to calculate the relative abundance of β-Asp-Arg and β-Asp-Lys subunits.

CHAPTER 3: GAIN OF FUNCTION BY RATIONAL DESIGN IN CYANOPHYCIN SYNTHETASE FROM *ACENITOBACTER BAYLYI*

Author list: Kyle Swain, Itai Sharon, Wyatt Blackson, Alison Haymaker, T. Martin Schmeing, David R. Nielsen, Brent L. Nannenga

Abstract

One of the key priorities in combating the pollution crisis is finding sustainable, biofriendly alternatives to current plastic production approaches. Cyanophycin, a biopolymer composed of a poly-L-aspartic acid backbone with L-arginine side chains, is a promising replacement for synthesized polyacrylate. It is produced by a nonribosomal peptide synthatase, Cyanophycin synthetase (CphA). In this work we modify CphA from *Acenitobacter baylyi* to create two novel mutants, G163K and G217K. G163K produces CGP with high monodispersity while G217K increases CGP production by up to 34% during *in vivo* trials.

Highlights

- Cyanophycin synthetase (CphA) is an enzyme that produces a biopolymer called cyanophycin
- Cyanophycin can be used as a bio-based substitute for polyacrylate
- G163K mutation in CphA from *A. baylyi* can produce a more monodisperse cyanophycin
- G217K mutation in CphA from *A. baylyi* can produce 34% more cyanophycin per *in vivo* trial

3.1 Introduction

Finding renewable alternatives for current petroleum-based plastics will be one step towards creating a more sustainable future [7]. Bio-based sources, such as nonribosomal peptide synthases (NRPSs), can provide complex product molecules like plastics without requiring the cellular guidance and machinery of many other enzymes [3,72]. Cyanophycin synthetase (CphA), an NRPS derived from various molds, fungi, and cyanobacteria, can produce cyanophycin, a biopolymer with a poly-L-aspartic acid backbone decorated with poly-L-arginine side chains [10]. In biological sources, cyanophycin is used as a nitrogen and carbon storage system for times when these nutrients are not readily available [25,73]. CphAs are further subdivided into CphA1s, or enzymes which can synthesize cyanophycin *de novo* from free aspartic acid and arginine, and CphA2s, enzymes which can ligate prebuilt aspartic acid-arginine dipeptides [22]. For their ability to operate with simpler building blocks, this study focused on CphA1s and in particular, CphA from *Acetinobacter baylyi* (abCphA) due to its ability to be recombinantly expressed in *E coli* [9].

AbCphA is a 400kDa tetrameric NRPS that contains three functional domains, characterized in 2021 by Sharon et. al, with a structure solved to 4.4Å by cryo-EM [9]. The N domain serves as a binding point that holds a growing cyanophycin chain as it alternates in a "flip-flop" motion between the other two domains [74]. It also has recently been demonstrated to generate starting primers during *in vivo* experiments, as CphA in vitro is not capable of starting the polymerization reaction without available primers [65]. The M domain shows similarity to Mur ligases, and uses ATP to add the decorative

arginine side chains to the aspartic acid backbone [63,64]. The G domain also uses ATP and constructs cyanophycin's aspartic acid backbone [9].

Cyanophycin natively has a size that ranges from 25-100 kDa, although *E. coli* based expression usually results in shorter chains in the 25-40 kDa window [14,63]. There is evidence of some side chain variation, resulting in cyanophycin chains that have higher lysine or other amino acid content in the place of arginine [20]. This results in the production of soluble cyanophycin, where the lysine content is higher than 25%, rather than the more typical insoluble material where the arginine content dominates [25]. The insoluble fraction is particularly attractive as a biopolymer due to less resource-intensive recovery methods and more uniform polymer composition [8,60]. At the same time, production and start-up costs are prohibitive to making this a widespread, readily available bio-based plastic [26]. One way to potentially increase yields and make cyanophycin a more attractive prospect is to improve CphA efficiency or stability through rational design or directed evolution.

In the case of CphA, there is a history of mutations causing beneficial effects[5,12,68]. Hai et. al were able to delete 31 c-terminal residues from a CphA derived from *Nostoc ellipsosporum*, generating a 2.2-fold-higher enzyme activity than the wild type [68]. In the same work, though, a 59 residue c-terminal truncation completely eliminated activity [68]. Another work in 2017 noted that a deletion of just one c-terminal residue, 874, in CphA from *Synechocystis spp. 6308* was able to increase CGP production by nearly 15% versus the wild type [71]. The most recent and exciting example of protein engineering on a CphA variant was done in 2022 by Sharon et. al when they used a single point mutation to change a CphA2 from primer-dependent to primer-independent function [22]. With this in

mind, we selected abCphA as a potential target enzyme for mutagenesis due to its soluble expression and the well documented substrate specificity, which allows little to no lysine or other amino acid contaminants to replace arginine in CGP [75].

3.2 Materials and Methods

The gene for abCphA (7LGM, PDB) was synthesized by Genscript (NJ, USA), cloned into a pET-28 expression vector with a c-terminal 6x-HIS tag using Gibson assembly, and transformed by heat shock into both competent NEB-5α and BL21-DE3 *E. coli* strains (NEB, MA, USA) for plasmid and protein expression, respectively. To ensure correct gene orientation, the plasmid was sequenced by Genewiz (NJ, USA) and confirmed to be accurate.

Mutagenesis was done using Q5-SDM kits from New England Biolabs, following their recommended protocols (NEB, MA, USA). Primers were designed using the NEB NEBasechanger tool, visualized in Snapgene, and synthesized by Integrated DNA Technologies (IA, USA). Mutants were sequenced by Genewiz for confirmation of the appropriate mutation.

Protein Expression and Purification

AbCphA was overexpressed in BL21-DE3 *E. coli*. Cells were grown in LB media supplemented to 50 µg/mL kanamycin at 37ºC. Expression was induced at OD600 of 0.6 by adding 1mM IPTG, and the temperature was reduced to 30ºC for 20 hours. Cells were harvested by centrifugation (5000g, 15min), resuspended in Buffer A (150mM NaCl, 0.5mM TCEP, 50mM Tris-HCl, pH 8.0) and lysed by sonication. The resulting sample was centrifuged at 10,000g for 15 minutes and the supernatant was collected. It was flowed through a 5mL Ni-NTA column supplemented with 2% buffer B (150mM NaCl,

0.5mM TCEP, 500mM imidazole, 50mM Tris-HCl, pH 8.0). The column was then washed with 80% buffer A, 20% buffer B, and then abCphA was eluted with 100% buffer B. Peaks were collected and flowed through a size-exclusion column, Superose 6, in buffer A. Any peak fraction corresponding to approximately 400 kDa was collected.

Cyanophycin Production

Cyanophycin was produced in 25mL of Terrific Broth (TB), supplemented with 50 mg/mL kanamycin, in 125mL baffled shake flasks. Cells were grown at 37ºC until an OD600 of 0.6, then induced using 1mM IPTG. The temperature was lowered to 30ºC for 48 hours. Cells were harvested and stored at -20ºC.

Cell pellets were resuspended in 1M HCl at a ratio of 10mL per gram of wet cell weight. After brief vortexing, pellets were shaken at room temperature for six hours. They were centrifuged at 10,000g for 15 minutes and the supernatant was collected. Tris pH 8.0 was added to a final concentration of 0.5M. Each sample was neutralized by adding 10M NaOH until the sample pH was 8.0. Now, the neutralized sample was centrifuged for 10,000g for 15 minutes. The precipitate is insoluble cyanophycin. The supernatant was removed and added to 2 volumes of ethanol, which causes the soluble cyanophycin to precipitate. The solution was centrifuged for 15 minutes at 10,000g with the precipitate collected as soluble cyanophycin. All cyanophycin fractions were lyophilized and then weighed.

Activity Assay

This assay was performed as previously described, with modifications [9]. A reaction mixture containing 12.5 mM ATP, 2 mM $MgCl_2$, 12.5 mM L-aspartic acid, 12.5 mM L-arginine, 12.5 mM KCl, 100 mM Tris-HCl pH 8.0, 1.5 mg/mL purified cyanophycin was

added to purified abCphA (100 μg/mL final). This incubated at room temperature for 2 hours.

Mass Spectrometry

25 mg/mL soluble and insoluble cyanophycin were resuspended in 50mM NH_4CO_3 and 15μM cyanophycinase, then left to react overnight at room temperature with mild agitation. To calculate relative β-Asp-Arg and β-Asp-Lys subunit abundance, peak areas including Na+ adducts were used.

3.3 Results

Nine mutants were created, G163KQ, A202NS, G217KQRE, and a combination mutant of G163K and G217K. Only four proved to be stable and soluble after purification, G163K, G217E, G217K, and the combination mutant. The final, purified enzymes are visualized in Figure 3.1.

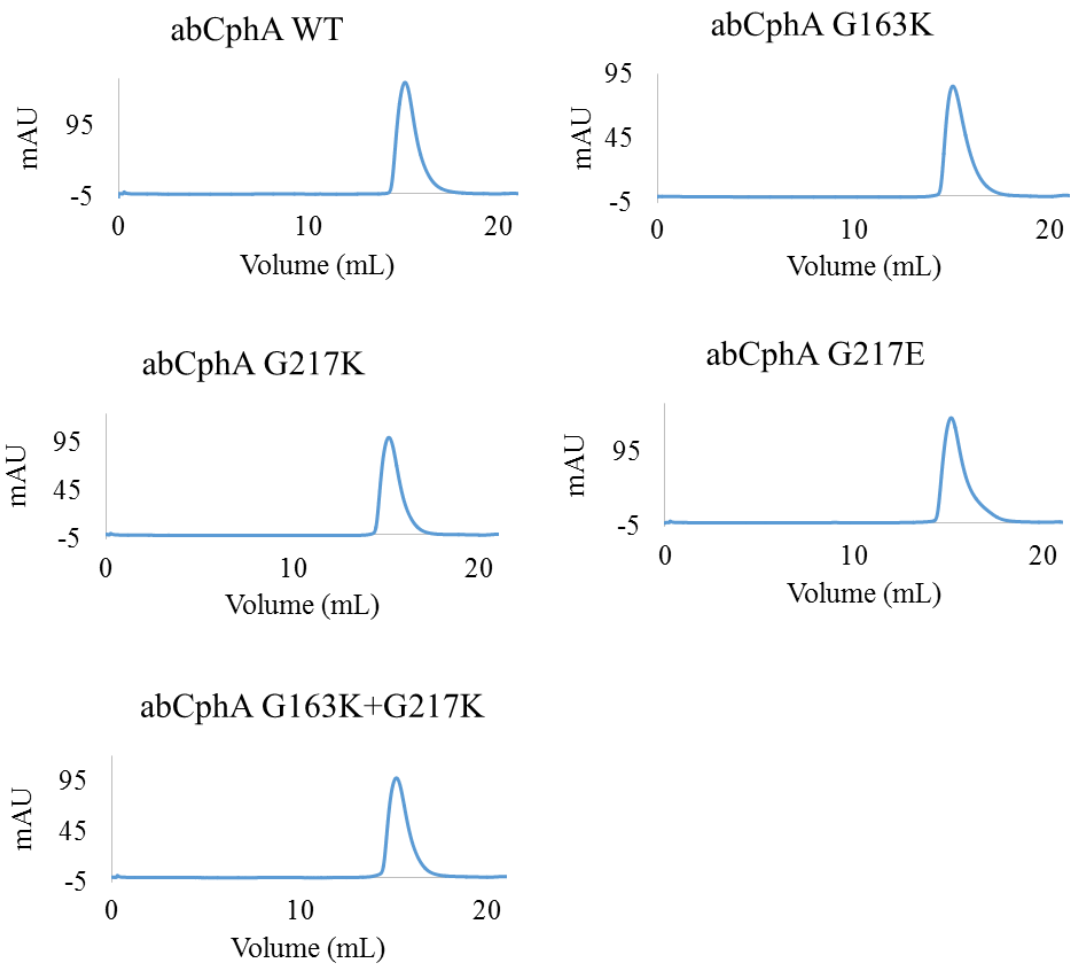


Figure 3.1: SEC curves of all mutant purifications, showing clean, stable elutions following the standardized purification procedure.

G163K and G217K were placed in comparative activity assays and run against wild type abCphA (WT) with samples in duplicate. G163K generated a final OD600 reading of 0.61 while G217K was 0.58, both significantly higher than the WT at 0.48. This implies either that the mutants have higher activities or better stability, resulting in higher CGP productivity. The results of an activity test of all five mutants are visualized in Figure 3.2.

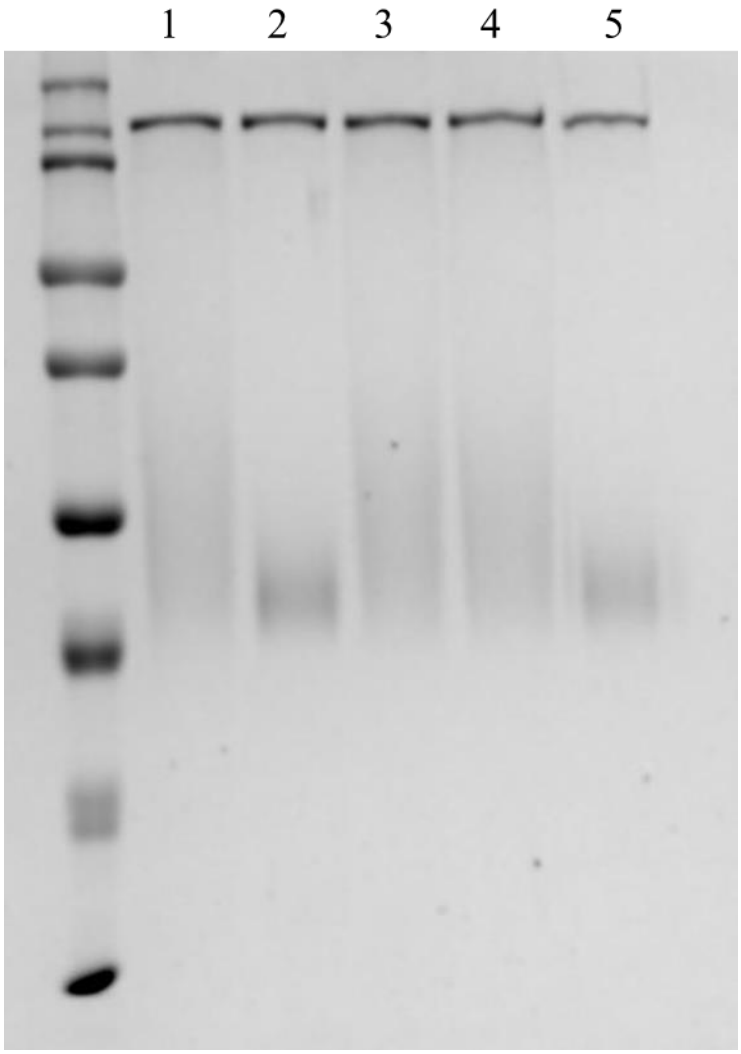


Figure 3.2: CphA activity assay. The monomeric 100kDa CphA, dissociated from the native tetramer by the SDS, can be seen with a cyanophycin smear of varying chain lengths. Lane 1 is WT abCphA, lane 2 is G163K, lane 3 is G217K, lane 4 is G217E, and lane 5 is the combination mutant.

One interesting result was that G163K and the combination mutant, when run in vitro, generate cyanophycin of significantly shorter chain lengths than the standard size of WT, G217K, and G217E cyanophycin.

In order to further examine these phenomena, G163K and G217K were produced *in vivo*, in 25mL culture volumes of TB, using the cyanophycin production procedure. Average

productivity values for insoluble cyanophycin and soluble cyanophycin were established by running 9 replicates of the WT and 6 replicates each for the G217K and G163K mutants.

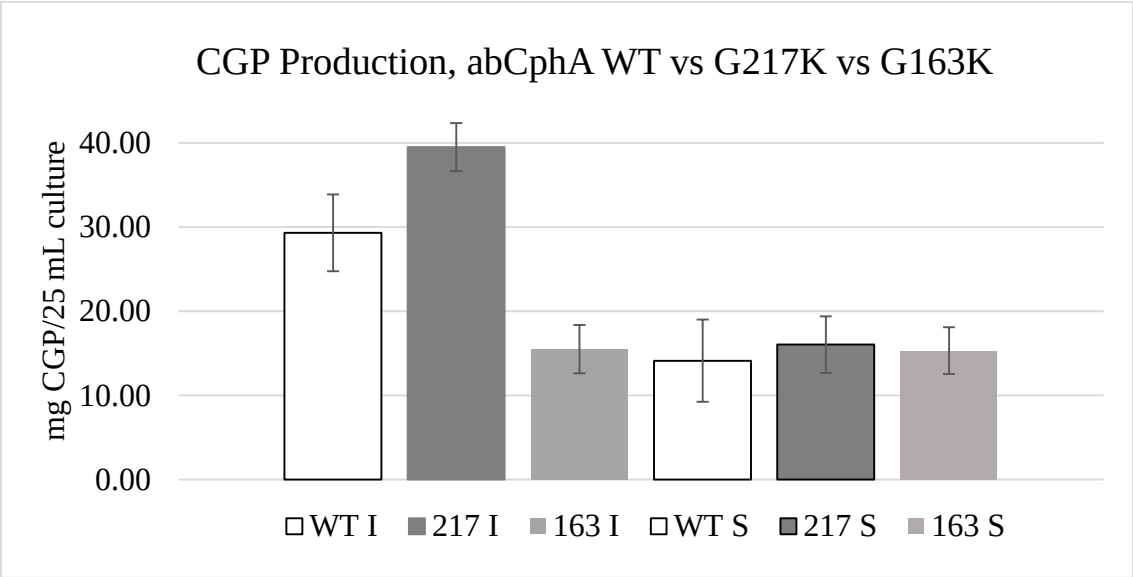


Figure 3.3: Cyanophycin production in WT, the G217K mutant, and G163K mutant. It demonstrates the higher productivity of the G217K mutant, particularly for insoluble cyanophycin, and also the significantly reduced insoluble cyanophycin production by the G163K mutant.

G217K outcompetes WT by 34% in insoluble cyanophycin production, while the G163K mutation causes a reduction of 47% production of insoluble cyanophycin. The soluble cyanophycin fractions remain unchanged. As it has been previously demonstrated that single individual deletions or mutations can cause significant effects on cyanophycin production, this is in line with previous findings.

At the same time, the G217E mutant showed a 20% decrease for *in vivo* insoluble cyanophycin production and the G163K/G217K combination mutant had similar productivity and polymer monodispersity to the G163K mutant.

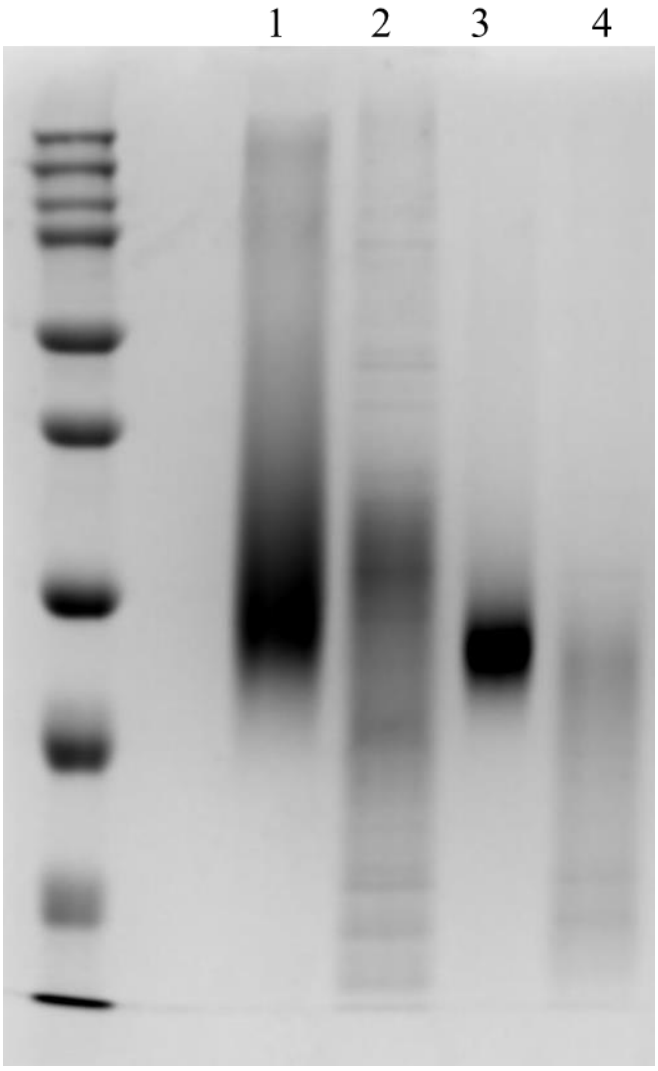


Figure 3.4: SDS-PAGE gel of *in vivo* cyanophycin samples. Lane 1 is the insoluble cyanophycin for abCphA WT, lane 2 is the soluble cyanophycin for abCphA WT, lane 3 is the insoluble cyanophycin for G163K, lane 4 is the soluble cyanophycin for G163K.

All samples were run on mass spectrometry by the Schmeing lab at McGill University. It was found that all cyanophycin produced by WT abCphA and any mutants shared similar compositions in both soluble and insoluble fractions.

In order to see if increased structural resolution provided more insight as to the effects of specific mutations, abCphA was processed using cryo-EM to produce a more complete structure. The current standard, 4.4Å, does not provide adequate resolution to see the

orientation of active site amino acids or the bound waters which could help answer several of our questions. Two datasets were collected, with one providing far superior particle quality so that one was selected for cryosparc processing. As expected, abCphA dissociated into dimers despite its tetrameric architecture demonstrated on the sizing traces shown earlier.

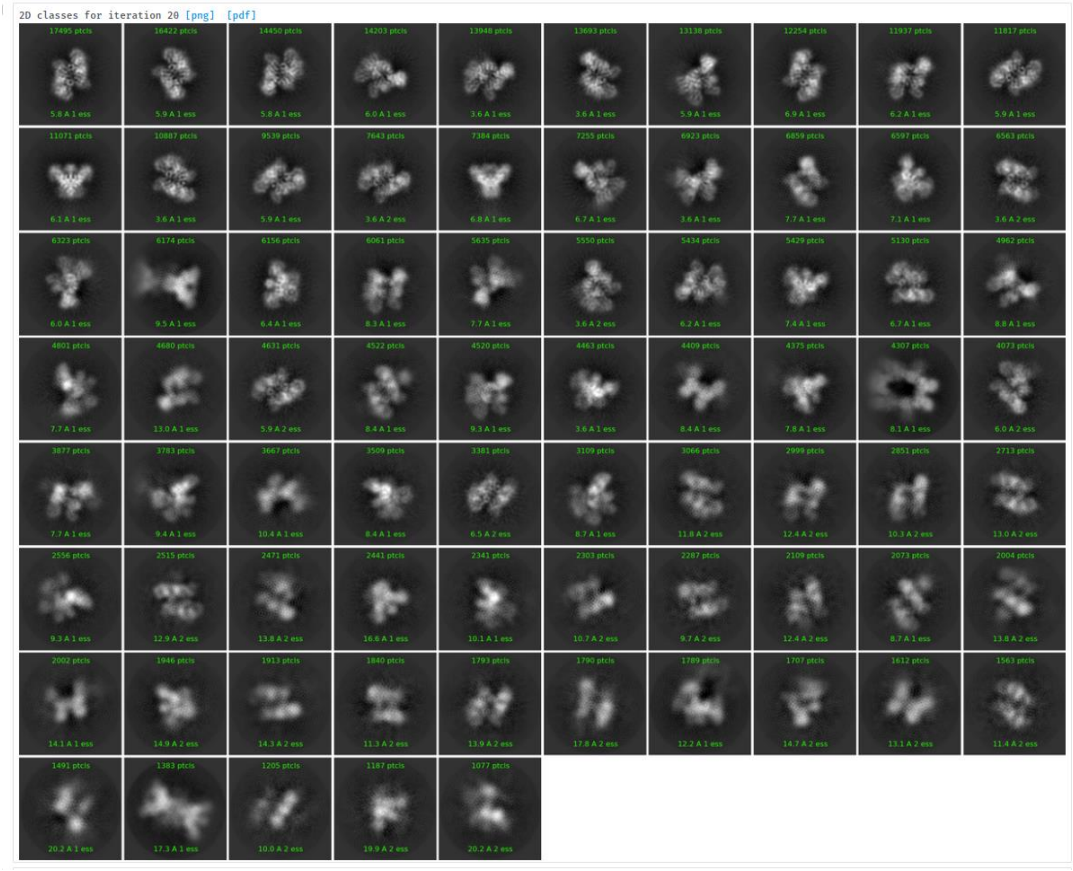


Figure 3.5: 2-dimensional class averages of abCphA in CryoSPARC showing obvious dimers rather than the expected tetramers

The structure was refined using the current 4.4Å structure as an ab-initio and model for computational particle picking. The resulting particles, when refined, produced a 3.5Å structure. While this is an improvement and worth noting, it does not provide significant

structural insights on the desired parts of the active site so further refinement or more particle image collection would be needed to proceed with this arm of the project.

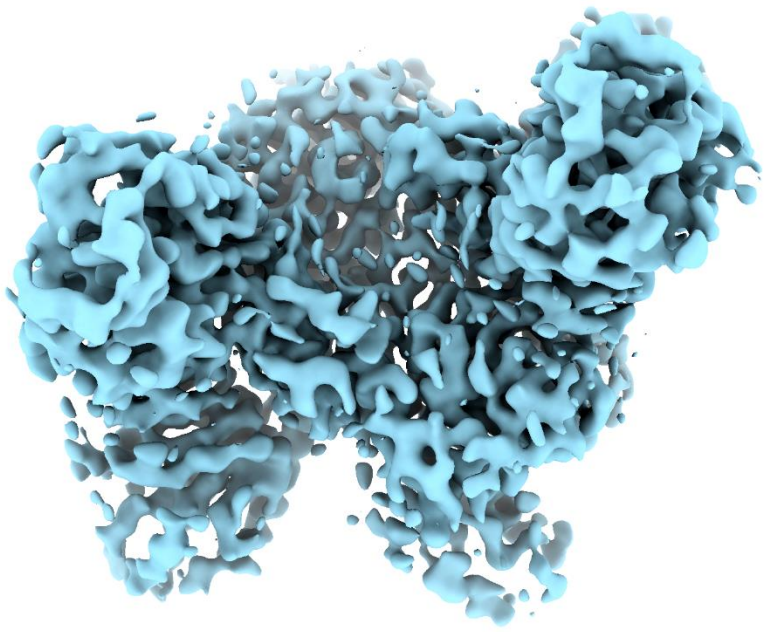


Figure 3.6: abCphA electron density map in Chimera at 3.5Å

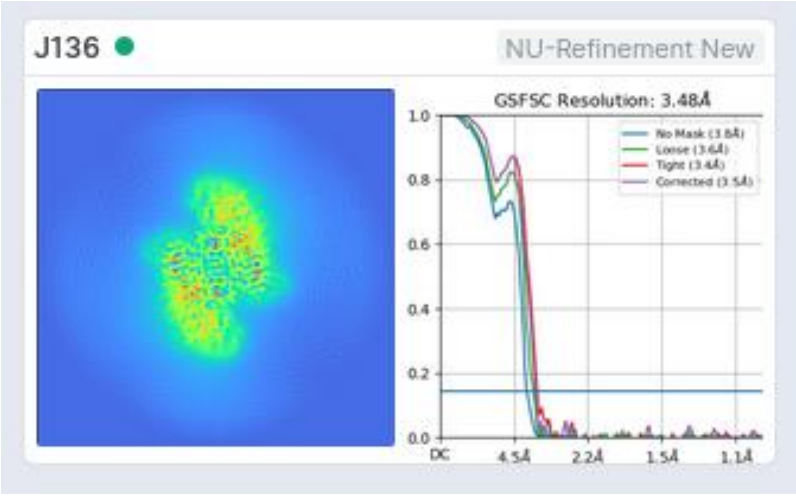


Figure 3.7: CryoSPARC abCphA final resolution and quality figure

3.4 Discussion

Based on the previous results, it is obvious that individual residue changes can affect the overall function of CphA. In particular, of note, are the G163K and G217K mutations that cause either a more monodisperse cyanophycin, or increase production by up to 34%, respectively. We selected these mutation sites in order to affect the cyanophycin production pathway as determined by Sharon et al [9]. While most of the mutations destabilized the enzyme entirely and resulted in poor soluble expression, the successful ones had surprising effects.

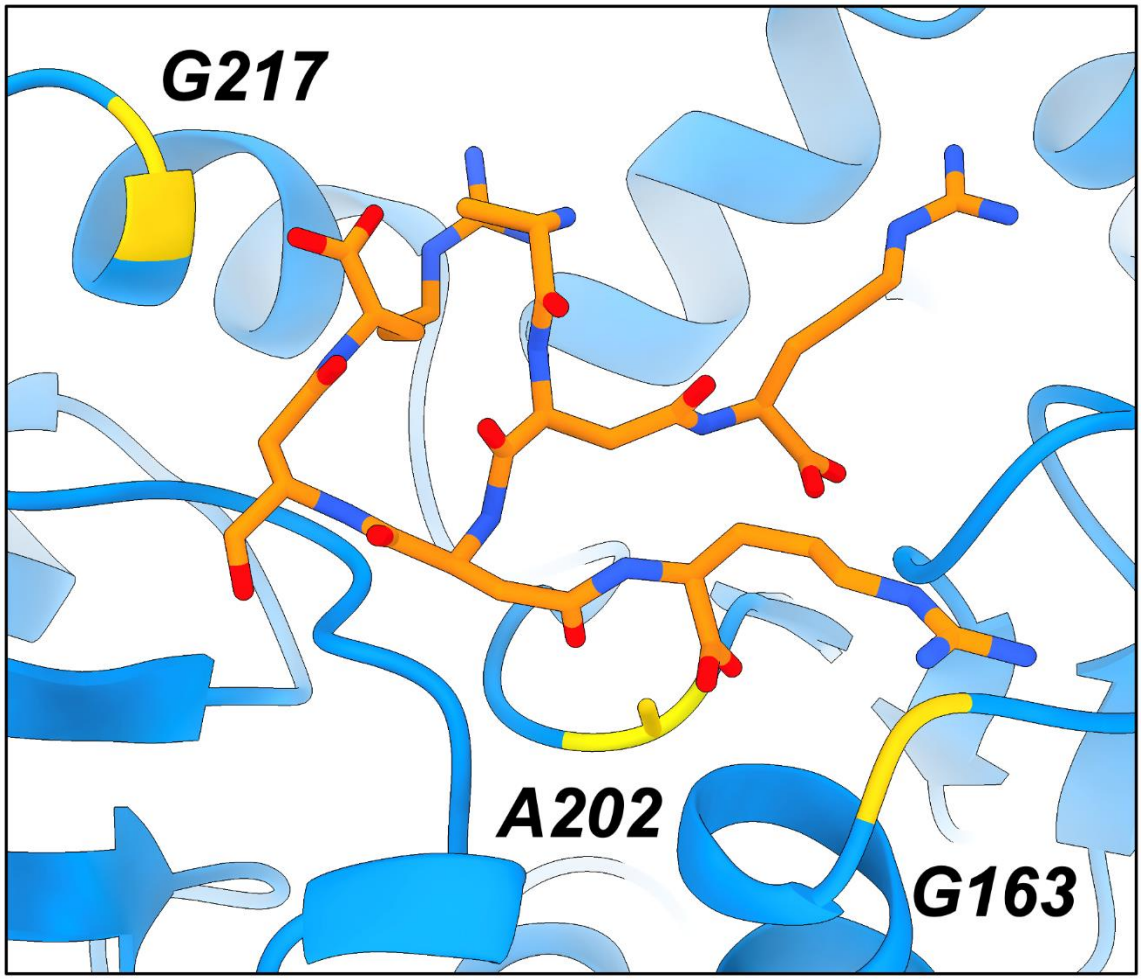


Figure 3.8: Chimera model of the mutation sites G163, A202, G217 interacting with modeled substrate.

As seen in Figure 3.8, all residues are on the cyanophycin production pathway but in different domains. G163K is in the N domain and we hypothesize that this downregulates and standardizes chain length by assisting with cyanophycin dissociation from CphA earlier in the extension cycle. The placement high up along the N domain would help the polymer chain detach earlier than the WT. At the same time, G217K showed a marked increase in cyanophycin production over the WT. It is possible that the positively charged lysine helps increase the "flip-flop" rate of cyanophycin production between the G and M domains, allowing for higher productivity per molecule over the lifespan of a single CphA. This is supported by the G217E mutant's reduction in productivity, which implies that the charge on the 217 residue may play a role in regulating cyanophycin production rate. However, this could require a minimum chain length in order to take effect, as the combination mutant showed no productivity increase over the G163K individual mutant. Further research into these mutants' mechanisms is necessary to understand more about how to increase CphA efficiency and how to up or downregulate cyanophycin chain length. There is potential for creating a tunable CphA that produces a monodisperse, desired chain length.

The mass spectrometry results demonstrate that all the produced cyanophycins have similar compositional ratios of dipeptide subunits. The mutations do not effect chain composition so they do not create strong enough steric or ionic blockages to stop the "flip-flop" mechanism that produces cyanophycin.

3.5 Conclusion

Increasing the availability and cost-effectiveness of bioplastics is a consistent goal in pursuit of more renewable alternatives to current materials. One of these bio-sourced polymers, cyanophycin, requires a prohibitive amount of production cost in the current market. The goal, then, is to find ways to either increase the product per unit cost or reduce the feedstock necessary to produce a similar amount of material.

In this paper, we have demonstrated a way to increase the efficiency of an enzyme that produces cyanophycin, abCphA. This mutant, termed G217K, shows up to 34% improvement over the wild type abCphA, which is an exciting development in terms of increased productivity. As well, we have discovered and characterized a second mutant, G163K, which demonstrates a novel, shorter chain length, monodisperse cyanophycin although at the cost of almost 50% less CGP produced *in vivo*. Further experiments have demonstrated that the mutant G217E shows a 20% reduction in productivity, and a combination mutant of G163K and G217K reflect a similar yield of monodisperse polymer as G163K. These exciting results demonstrate that abCphA is a potentially tunable enzyme with opportunities for improved efficiency and capable of producing polymers with different chain lengths than the wild type.

CHAPTER 4:

CHAIN-LENGTH DEPENDENT PROGRAMMED BIOSYNTHESIS OF β-γ POLYENES BY AN ITERATIVE TYPE I POLYKETIDE SYNTHASE

Author contribution: This work was highly collaborative with significant contributions from the Lin, Nannenga, and Wang labs. The author of this thesis contributed to the crystal structure of AlnB_DH, the cryo-EM structure of AlnB holoenzyme, and many AlnB_DH mutants with associated activity assessments in the mutant screening program to help tease out the functional mechanism of AlnB_DH. The paper has been reproduced in its entirety due to being in the submission/review process. There is an update at the conclusion regarding recent findings pertaining to this submission and possible edits to the paper in the near future.

Authors: Bin Wang[#,*], Kyle Swain[#], Fang Guo, Wyatt Blackson, Chunshuai Huang, Bing Chen, Teresa A. Martin, Brent Nannenga[*], and Huimin Zhao[*]

4.0 Summary

A large number of polyketides are produced by the iteratively acting type I polyketide synthases (iT1PKSs) through highly programmed but poorly understood biosynthetic machinery. Here we report that the bifunctional α,β-dehydratase/isomerase domain from an iT1PKS, AlnB_DH, can switch its enzymatic function in an unprecedented chain-length dependent manner and act as the central link in the biosynthesis of several unusual β,γ-polyene products. The molecular basis for the functional switch and multidomain collaboration in driving the thermodynamically unfavorable β,γ-double bond formation was elucidated via protein crystallography, site-directed mutagenesis, and *in vitro* enzymatic assays. This work decodes the intrinsic programming of AlnB_DH as well as

the meta-programming of the holo AlnB, and has laid the foundation for the re-programming of iT1PKSs for synthesis of novel polyketides.

4.1 Introduction

Microbial natural products are a major source of the modern medicines and pharmaceutical agents [76,77]. They are marvelously diverse in structure but synthesized by only a few classes of biosynthetic machineries. Among the most successful of these are the polyketides that are synthesized by multifunctional type I polyketide synthases (T1PKSs), which either operate in an iterative mode (iT1PKSs) or are organized to assembly-lines to build carbon chains step-by-step [30,78]. Structurally, a typical iT1PKS contains only one module composed of ketosynthase (KS), acyltransferase (AT), acyl carrier protein (ACP), and accessory domains such as dehydratase (DH), methyltransferase (MT), enoylreductase (ER), and ketoreductase (KR); in mechanism, the single module would run for several extension cycles. In comparison to the mammalian fatty acid synthase (mFAS) which forces all of its domains to work in all extension cycles and produces the fully reduced palmitic acid, an iT1PKS is usually programmed in the control of chain-length and oxidoreductive state depending on the intrinsic programming of individual domains as well as the meta-programming from all of active domains together. However, for the majority of iT1PKSs, the programming is poorly understood, and elucidation of the molecular basis remains very challenging [34].

Naturally occurring β-γ double bond bearing polyketides are quite rare [79-81]. Recently, we characterized the β-γ pentaenic allenomycins and the associated iT1PKS biosynthetic gene cluster, *aln* [32]. The sole AlnB_DH was predicted to catalyze the formation of both α-β and β-γ double bonds in the polyene structures. In this work, we focus on understanding the

intrinsic programming of AlnB_DH, including: 1) how it determines when to perform α-β or β-γ dehydrations and what the molecular basis is for the switch of the catalytic mechanisms? 2) as the formation of β-γ-double bond is both thermodynamically and kinetically unfavorable, how it completes the reaction? Answering the two questions has led us to decode the meta-programming of the β-γ-polyene biosynthesis.

4.2 Materials and Methods

Bacterial strains and culture conditions

Escherichia coli strains were cultured in Luria-Bertani broth at 37°C supplemented with 50 μg/ml of kanamycin. *E. coli* NEB5α was used for cloning, *E. coli* BAP1[82] for protein expression of alnB-his and his-alnB_ACPv2, and BL21(DE3) for expression of both N-his and C-his tagged alnB_DH, alnB_ER, alnB_KS, alnB_KSAT, alnB_KSATDH^0, and the related mutants.

Plasmid construction and protein expression

All of the wild type expression plasmids were constructed by Gibson Assembly [New England Biolabs (NEB), Ipswich, MA], and the mutant plasmids were constructed by KDL (NEB) except pET28a-his-alnB_KSATDH^0 which was built also by Gibson Assembly (NEB). The fragments were PCR amplified using Q5 hot-start high-fidelity DNA polymerase (NEB) and primers listed in the table below. pET28a was cut by *Nco*I+*Hin*dIII to prepare C-his tag backbone while cut by *Bam*HI+*Hin*dIII (NEB) to generate N-his tag backbone. Assembled plasmids were transformed into NEB5α high-efficiency competent cells (NEB), and correct constructs were obtained by plasmid mini-prep, restriction digestion, and DNA sequencing using universal primers for pET28a. Then they were transformed into *E. coli* BAP1 and BL21 (DE3), respectively. Protein

expression was induced with 0.1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD_{600} = 0.6 and was incubated at 20°C, 200 rpm for overnight. Protein purification was based on Ni-NTA his-tag affinity chromatography and size-exclusion chromatography. Protein purity was assessed by SDS-PAGE.

In Vitro Reaction Assay for AlnB_DH and Mutants

Table 4.1

Reagent	Stock conc.	Add volume, µl	Final conc.
H_2O		32	
Tris-HCl, pH 8.0	1 M	2.5	50 mM
TCEP, pH 7.2	0.5 mM	0.5	10 □M
PANT substrates	15 mM	1	600 □M
His-AlnB_DH or mutants in 40% glycerol	500 □M	5	100 □M
Total		25	

Reaction was performed at 25°C for 1 h., and then an equal volume of MeOH was added to stop the reaction, precipitate protein, and dissolve the products. After centrifugation at 13000×*g* for 5 min through an Amicon Ultra 30kDa filter (GE), 20 µl of flow-through was injected onto the HPLC.

Crystallization and structure determination

Cells were initially grown at 37ºC until the optical density (OD_{600}) reached 1.0. AlnB_DH expression was induced by the addition of 0.1 mM IPTG and the temperature was reduced to 25ºC for 18-24 hours. Cells were harvested by centrifugation (4000×g, 15 minutes), and frozen at -20ºC. Cell pellets were resuspended in lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl), and lysed by sonication on ice at 40% amplitude, 2s-on 2s-off pulses for 5 minutes. The lysate was clarified by centrifugation (20000×g, 15 minutes), filtered at 0.45 µm, and loaded onto a Histrap HP column (Cytiva). It was washed

extensively with lysis buffer mixed with 2% buffer B (lysis buffer with 500 mM imidazole), and then eluted with 100% buffer B. Fractions showing high OD_{280} were combined, diluted 1:10 in lysis buffer to reduce imidazole concentration, and then concentrated using an Amicon Ultra 30 kDa at 8000×g for 30 minutes to approximately 60 mg/mL (OD_{280}).

For crystallization, a truncated construct of AlnB-DH was constructed consisting of residues 35 to 328. This construct was expressed and purified as described previously with an additional SEC purification step. Following elution from the Histrap column, the sample was applied to a Superdex200 16/60 column (GE Healthcare) equilibrated in lysis buffer, and the resulting fractions were concentrated to 12 mg/mL. This AlnB-DH construct was crystallized using the hanging drop vapor diffusion method. Enzyme at 12 mg/mL was added at a 2:1 µL ratio with well solution (0.2 M Ammonium Sulfate, 0.1 M Tris pH 6.0, 29% PEG 3350) and equilibrated against a well solution reservoir of 500 µL for 5-7 days at 25ºC. Crystals formed as either thin rods or stacked plates. The thin rods were harvested and cryoprotected using a solution of 0.2 M Ammonium Sulfate, 0.1 M Tris pH 6.0, 25% PEG 3350, 40% glycerol, and crystals were looped and flash frozen in liquid nitrogen. Diffraction data were collected on microfocus beamline 23-ID-B at the Advanced Photon Source. Diffraction data were processed using XDS [83] and the data set was merged and scaled using AIMLESS [84]. The data were phased using the BALBES molecular replacement pipeline [85]. Structure refinement was performed using REFMAC [86] and manual model building was performed using Coot [43].

Table 4.2 Crystallographic data collection, processing, and structure refinement

Diffraction source	APS 23-ID-B (GM/CA)

Wavelength	1.0332
Temperature	80 K
Detector	Eiger 9M
Rotation range per image	0.20°
Exposure time per image	0.2 s
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	63.94, 73.97, 129.90
α, β, ɣ (°)	90, 90, 90
Resolution (Å)	48.1 – 2.35 (2.44 – 2.35)
Molecules in the asymmetric unit	2
R_{merge}	0.125 (1.870)
$CC_{1/2}$	0.996 (0.346)
I/σ_I	5.8 (0.4)
Completeness (%)	94.3 (67.4)
Redundancy	3.3 (2.2)
Total no. of reflections	81125
No of unique reflections	24748
Structure Refinement	
Resolution	48.37 – 2.55
R_{work}/R_{free}	0.2549/0.3096
RMS deviations	
Bond length (Å)	0.016
Bond Angle (°)	2.039
Ramachandran statistics	
Outliers (%)	0.38
Favored (%)	94.74

Values in parentheses refer to the highest resolution shell

Further materials and methods are included in Appendix A, as provided by Bin Wang, Huimin Zhao, and the research team at University of Illinois Urbana-Champagne.

As well, Appendix A contains a 3.8Å cryo-EM structure of full length AlnB wild type, plus extensive purification results from AlnB and AlnB DH/DH Crys 3 from the Nannenga lab.

4.3 Results and Discussion

An *alnB* expression system in the *E. coli* BAP1 [82] strain was previously established [32]. In this study, extensive *in vivo* and *in vitro* experiments on *alnB* discovered a series of

heptaketide (**1**), octaketide (**2**, major), and nonaketide (**3**) β-γ polyene compounds, as well as two unmethylated congeners (**2'** and **3'**), based on UV-vis and HRMS analyses (**Fig. 4.1**, **fig. S1**, and **table S1**). We managed to purify and characterize **2** by ^{1}H NMR and COSY experiments to confirm its unusual β-γ polyene structure (**Fig 4.1d** and **fig. S2**). Collectively, all five structures contain a C_4 alkane tail (a result of α-β double bonds being reduced by AlnB_ER) and a β-γ polyene body, which indicates that the sole AlnB_DH is involved in the two forms of double bonds and probably it carries out α-β double bond formation in the 1st (C_4) and 2nd (C_6) iterations and then switches to β-γ (C_8) for the remainder of the iterations. Worthy of note, the methyl group that occurs ubiquitously on the β-γ double bonds presumably to localize the latter [79-81], is dispensable for AlnB reaction as indicated by the comparable productivity of **2** and **2'** (**Fig. 1b**); and in accordance, nonmethylated enoyl substrates were synthesized for the following study.

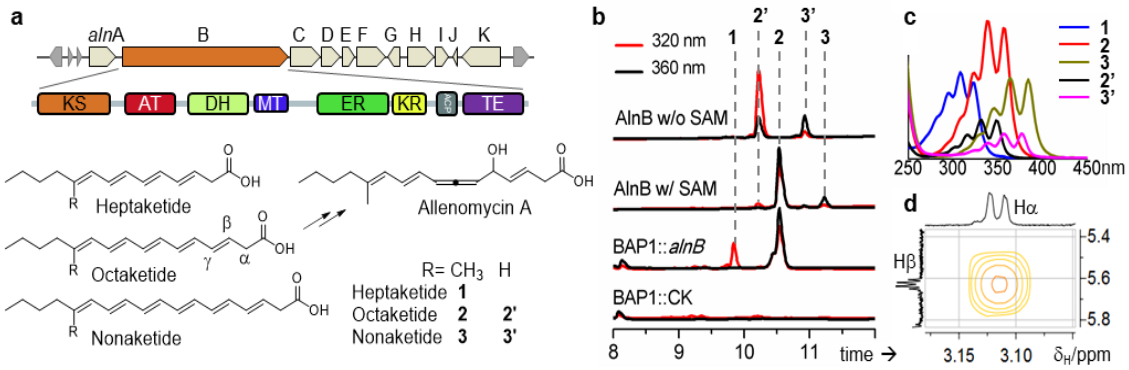


Figure 4.1 The iterative type I polyketide synthase (iT1PKS) AlnB and its associated β-γ polyene products**. a**) The *aln* gene cluster employs AlnB which contains a full set of domains to synthesize the β-γ polyene intermediates and together with tailoring enzymes to produce allenomycin A. **b**) HPLC analysis of peaks produced by *alnB* in *E. coli* and *in vitro*. **c**) UV-Vis spectra of the β-γ polyene compounds showing signature fine structures typical of unconjugated polyenes. **d**) COSY spectrum of compound **2** providing evidence for the HOOC-Cα**H**$_2$-Cβ**H**=CγH structure.

To investigate AlnB_DH activities and verify the catalytic switch, both α,β- and β,γ- double bond substrates linked to pantetheine (PANT) were chemically synthesized, as mimics of ACP-tethered substrates (**4**, **5**, **6**, and **7** in **Fig. 4.2a**, **fig. S3-S11** and **table S1**). AlnB_DH was then cloned and expressed in *E. coli* BL21(DE3), and purified for the *in vitro* reaction. The HPLC analysis showed considerable turnover from C_8 set, **6** and **7**, to the βOH- **9** (**Fig. 4.2b** and **table S1**), whereas in C_6 set only **4** was turned over to the βOH- **8** (**Fig. 4.2b** and **table S1**). This result agrees well with the β-γ-polyene structures that chain-lengths from C_6 to C_8 can effectively trigger the switch from α-β to β-γ-double bond formation. It also hints on two points that, 1) the mechanism switch is chain-length dependent that might involve binding pocket geometry rather than the active site chemistry as reported in FabA [87], and 2) the β-γ double bond hydration might proceed via a two-step route, i.e., β-γ toα-β-double bond isomerization plus α-β-hydration which currently is more recognized than the direct β-γ –hydration [88,89].

However, the mechanism of β-γ-double bond dehydration/hydration is not conclusive so far due to the universally low catalytic activity to the β-γ substrates which complicated the elucidation of the HRMS and NMR data [79,90]. Given that the α-β-double bond hydration is much faster than the β-γ- to α-β-double bond isomerization, we reason that the reaction tendency of the hydration reaction could reflect the overall reaction direction. We measured the *K* values of [βOH-] to [α,β-] and that at the equilibrium (K_{eq} = 1.6) in the time-course experiment (**Fig. 4.2c**). The result showed that *K* grew from 0.8 to 1.6 reaching the equilibrium, indicating a tendency from α,β- (**6**) to βOH- (**9**) and thus a previous β,γ- (**7**) to α,γ- (**6**) isomerization, i.e., AlnB_DH is a bifunctional α,β-DH/EI.

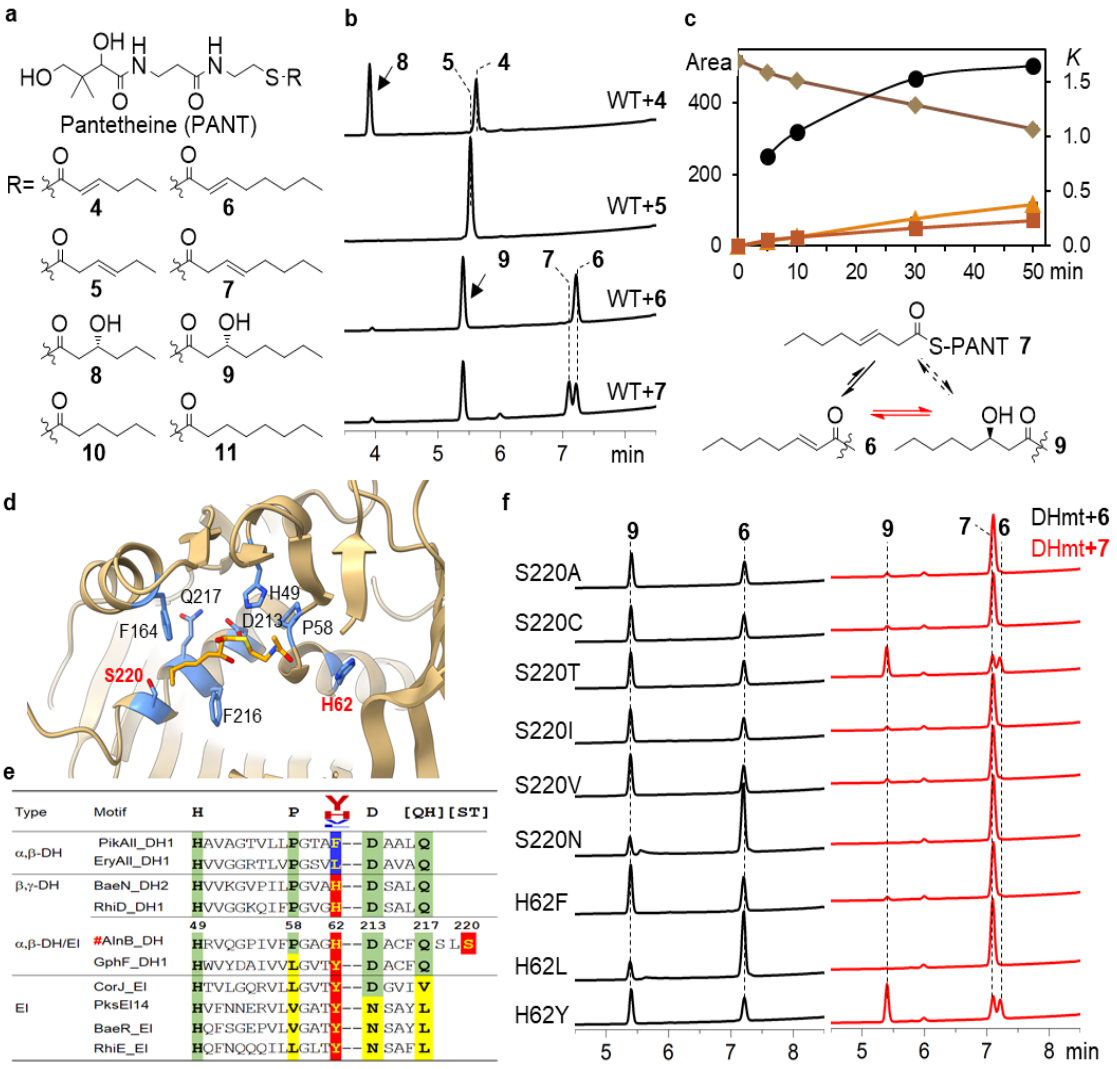


Figure 4.2. Structural and enzymatic analyses of AlnB_DH. **a**) Pantetheine-linked substrate surrogates used in this study. **b**) HPLC analysis of *in vitro* assays by the wild type AlnB_DH. **c**) The time-course plot of AlnB_DH toward **7**, in which the left axis is the area calculated at 235 nm absorption, bottom axis is reaction time points (replicates were performed), and right axis is the calculated K = area[βOH-]/area[α,β-]. **d**) Crystal structure of AlnB_DH with docked **9** (gold) and surrounding residues as sticks (light blue). **e**) Representative DH sequences of α,β-dehydratases (α,β-DHs), β,γ-dehydratases (β,γ-DHs), enoyl isomerases (EIs), and α,β-DH/EI bifunctional enzymes. A sequence logo[91] was extracted from multiple sequence alignment at position 62 as [YHVL]. **f**) HPLC analyses of AlnB_DH mutants toward **6** and **7**.

The molecular basis underlying enoyl isomerase (EI) catalysis remains elusive. Previously, comparative studies of α,β-DHs, α,β-DH/EIs, and EIs employed a series of domain

swapping experiments and site-specific mutagenesis but failed to annihilate the EI activity. Therefore, it was proposed that the geometry of the binding tunnel determines the EI activity rather than particular single residues around the tunnel or active sites [79,92,93]. In this work, AlnB_DH was successfully crystallized and shown as a homodimer (**fig. S12** and **table S2**), with each monomer composed of a typical double-hotdog fold and substrate binding tunnel opening outward.

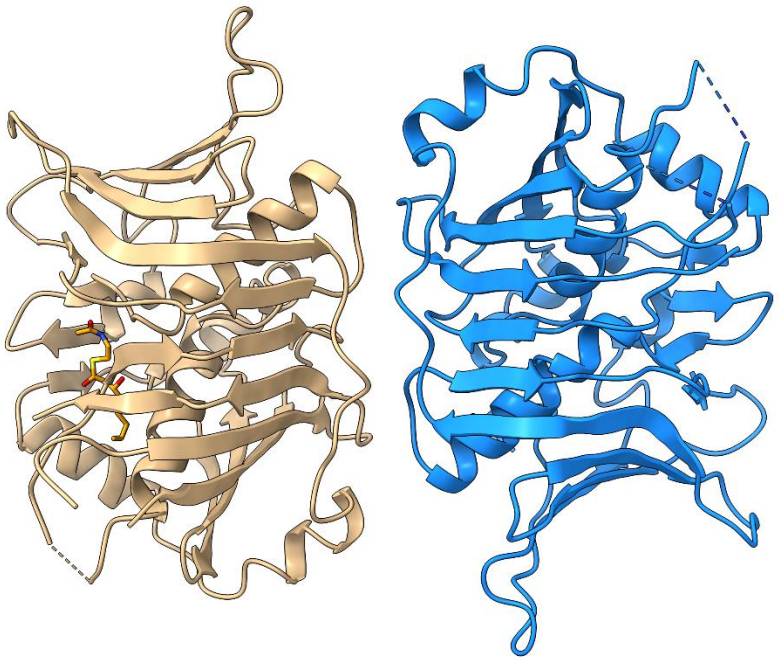


Figure 4.3. Crystal structure of AlnB_DH with substrate modeled into left monomer binding pocket

The catalytic H49 stacks with P58 (HX_8P motif) and the catalytic D213 is hydrogen bonded to Q217 (DX_3[QH] motif) (**Fig. 4.2d**). Since the EI activity is triggered by chain-length, **9** was then docked into the binding tunnel by referring to FabA (PDB 4b0i)[87]. Structurally, AlnB_DH possesses a long narrow hydrophobic tunnel as FabA except, a polar S220 which sits at the same side of D213 and Q217 (DX_3[QH]X_2**S**). The hydroxyl group of S220

protrudes into the tunnel right against the acyl distal of **9** in the docking simulation and is fixed by hydrogen-bonding to F216 backbone carbonyl, which suggests a role of gatekeeper, like I108 in FabF and M154 in mtKAS functioning by increasing the rotational energy barrier [94]. To verify the correlation of S220 to the EI activity, mutants S220[ACTIVN] with sidechains differed by size and polar were constructed and tested (**Fig. 4.2f**). The HPLC analysis showed that, all nonpolar mutants (S220[ACIV]) decreased their activities toward **7** down to the minimum regardless of sidechain size. In contrast, while both polar S220[TN] mutants maintain the hydrogen-bond the performance of S220T was identical to the wild type whereas the larger sidechain of S220N was deleterious even to the activity toward **6**. This result confirms that residues at position 220 are directly involved. On one hand, [ACIV]220 suggest a low rotational energy barrier which permits **6** readily for α,β-hydration; on the other hand, fixed rotamers of [ST]220 significantly increase the energy barrier, which leads to destabilized alignment of substrate to the α,β-positions but an increased probability of its re-alignment to the β,γ-positions to enhance the protonation at Cγ and complete the isomerization.

It is noticeable that [ACIV]220 remain active to **7** albeit minimally. We examined the active site machinery but found mutants, H49[AF] or D213[AN], completely lost activity to all four substrates, which is consistent with previous reports that both of them are involved in the EI activity [92,93]. Based on multiple sequence alignment on a curated collection of DHs and EIs, we noticed a pattern at H62 that is mainly composed of [HYL] among which nonpolar residues dominate in α,β-DHs while polar aromatic residues appear in other DHs and EIs (**Fig. 4.2e**). H62 is located on the α-helix that carries P58 at the tip which then stacks with and determines the orientation of the catalytic residue H49

(HX_8PX_3H). Subsequently, H62[LY] were constructed and tested, and it was shown that H62L was deleterious to the structure while H62Y maintained both activities as the wild type (**Fig. 4.2f**). An additional H62F was also tested and surprisingly it retained a minimal EI activity (**Fig. 4.2f**). Therefore, the residues at 62 from aliphatic nonaromatic to polar aromatic residues suggest a gain-of-function.

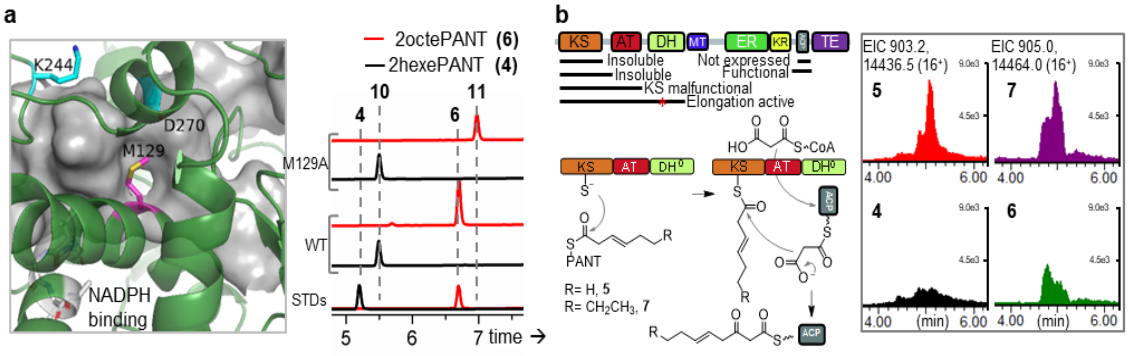


Figure 4.4. Analysis of the programming mechanisms of AlnB_ER and AlnB_KS**. a**) Docking **6** to a homology-based model of AlnB_ER, in which M129 truncates the binding tunnel to inner and outer spaces; right is the *in vitro* assays for the wild type and M129A. **b**) Scheme of AlnB module composition, cloned versions of domains, and elongation mechanism (left); right is the extracted ion chromatography (EIC) analysis of the elongation reactions using four compounds as substrates, the peaks were extracted as 16^+.

Although the isomerization from **7** to **6** is thermodynamically favorable [95] and [HY]62 plus [ST]220 can effectively carry it out, the reactivity is extremely low, not to mention the reverse isomerization from **6** to **7**. Given that enzymes are good at coupling an endergonic reaction with exergonic reactions, AlnB_ER and AlnB_KS are probably the driving forces to facilitate the isomerization. As previously reported, ERs were broadly toleran t with diverse structures but the tunnel sizes are generally related to chain lengths [96]. AlnB_ER was then modeled and docked with **6**. A large residue M129 was revealed in the halfway of a deep tunnel, resembling gatekeepers in FabF and mtKAS [94] (**Fig. 4.4a**). The wild type

ER and M129A were constructed and tested, and the HPLC analysis showed that the wild type can reduce only **4** whereas M129A gained the activity to reduce both **4** and **6** (**Fig. 4.4a** and **table S1**), i.e., AlnB_ER is "on" at the C_6 stage but switches to "off" at the C_8 stage, subjecting **6** to be recycled by AlnB_DH for the isomerization to **7**. Next, AlnB_KS was examined. KS was reported to catalyze a two-step reaction, transthiolation and elongation, but only the latter confers substrate selectivity [97]. In order to get functional domains, quite a few versions were tried, and in the end AlnB_KS-AT-DH^0 (w/ DHmtH49A) and AlnB_ACPv2 were constructed (**Fig. 4.4b**). The elongation experiment was monitored by LC-HRMS detection of ACP derivatives and the result showed that AlnB_KS prefers the β,γ-double bond bearing substrates **5** and **7** (**Fig. 4.4b** and **fig. S13**). Therefore, the exergonic chain elongation reaction selectively drives the isomerization from α,β to β,γ-double bond. Intriguingly, these isomers are physically similar, so to distinguish the subtle differences, AlnB_KS must have evolved a special binding tunnel as speculated in FabA [92], unless it resorts to a chemical differentiation strategy.

The complex programming of AlnB in the biosynthesis of β,γ-polyene natural products represents only one of the strategies evolved by fungal and bacterial iT1PKSs. By phylogenetic analysis, we also revealed that the pair of upgraded motifs, HX_8PX_3[HY] and DX_3[QH]X_2[ST], in AlnB_DH, and the gatekeeper residue M129 in AlnB_ER, are conserved and widely distributed among the genera of *Streptomyces* and *Nocardia*, and so is the AlnB-like module organization, which clearly indicates a high degree of co-evolution (**fig. S14**).

4.4 Conclusion*

To conclude, our result depicts a scenario that, before the stage of C_8, AlnB_DH produces α,β-enoyl which is then reduced by AnB_ER and goes to the next iteration; however, at the stage of C_8 AlnB_DH switches (chain-length dependently) to produce both α,β- (major) and β,γ-enoyl (minor) isomers by employing a pair of upgraded motifs, HX_8PX_3[HY] and DX_3[QH]X_2[ST]. On one hand, AlnB_ER switches off (chain-length dependently) toward the α,β-enoyl and, on the other hand, AlnB_KS preferentially depletes the β,γ-enoyl, which eventually drive the α,β- to β,γ-enoyl isomerization. This model resembles a "push-and-pull" mechanism at the multidomain level in order to drive a thermodynamically unfavorable reaction and the central link is AlnB_DH in the programmed biosynthesis of β,γ-polyene polyketides.

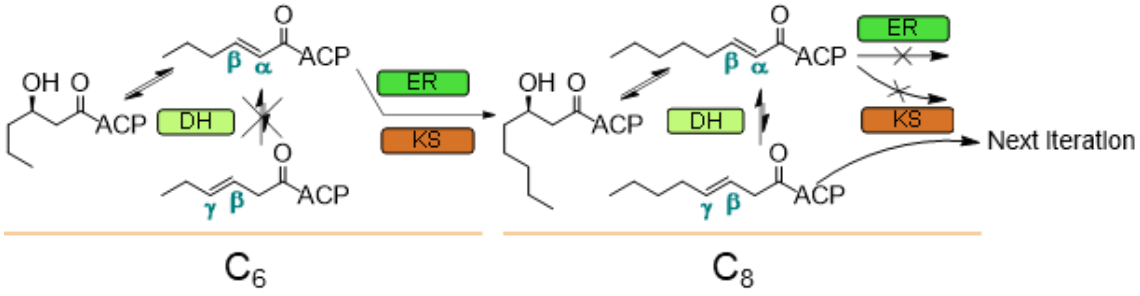


Figure 4.5: Reaction scheme of AlnB_DH, AlnB_ER, AlnB_KS

*Recent developments have provided an alternative theory that chain length regulates the handoff of the polyketide chain from one modular subunit to the next (i.e. from ER to DH or DH to KS) rather than DH working as a bifunctional enzyme with a gradual gain of function mediated by chain length. These developments also indicate that the previous β,γ double bond does not necessarily code for the formation of a second bond. More research is needed in regards to the interactions between the ER, DH, and KS domains.

REFERENCES

1 Cooper, G. *The Central Role of Enzymes as Biological Catalysits*, <https://www.ncbi.nlm.nih.gov/books/NBK9921/> (2000).

2 Finking, R. & Marahiel, M. A. Biosynthesis of Nonribosomal Peptides. *Annual Review of Microbiology* **58**, 453-488, doi:10.1146/annurev.micro.58.030603.123615 (2004).

3 Koglin, A. & Walsh, C. T. Structural insights into nonribosomal peptide enzymatic assembly lines. *Natural Product Reports* **26**, 987-1000, doi:10.1039/B904543K (2009).

4 Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A. & Essen, L.-O. Structural and Functional Insights into a Peptide Bond-Forming Bidomain from a Nonribosomal Peptide Synthetase. *Structure* **15**, 781-792, doi:10.1016/j.str.2007.05.008 (2007).

5 Reimer, J. M. *et al.* Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366**, eaaw4388, doi:doi:10.1126/science.aaw4388 (2019).

6 Bozhüyük, K. A. J. *et al.* Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nature Chemistry* **11**, 653-661, doi:10.1038/s41557-019-0276-z (2019).

7 Raddadi, N. & Fava, F. Biodegradation of oil-based plastics in the environment: Existing knowledge and needs of research and innovation. *Science of The Total Environment* **679**, 148-158, doi:https://doi.org/10.1016/j.scitotenv.2019.04.419 (2019).

8 Kwiatos, N. & Steinbüchel, A. Cyanophycin Modifications—Widening the Application Potential. *Frontiers in Bioengineering and Biotechnology* **9**, doi:10.3389/fbioe.2021.763804 (2021).

9 Sharon, I. *et al.* Structures and function of the amino acid polymerase cyanophycin synthetase. *Nature Chemical Biology* **17**, 1101-1110, doi:10.1038/s41589-021-00854-y (2021).

10 Du, J., Li, L. & Zhou, S. Microbial production of cyanophycin: From enzymes to biopolymers. *Biotechnology Advances* **37**, 107400, doi:https://doi.org/10.1016/j.biotechadv.2019.05.006 (2019).

11 Simon, R. D. Cyanophycin Granules from the Blue-Green Alga Anabaena cylindrica: A Reserve Material Consisting of Copolymers of Aspartic Acid and

Arginine. *Proc Natl Acad Sci U S A* **68**, 265-267, doi:10.1073/pnas.68.2.265 (1971).

12 Watzer, B. & Forchhammer, K. Cyanophycin Synthesis Optimizes Nitrogen Utilization in the Unicellular Cyanobacterium Synechocystis sp. Strain PCC 6803. *Appl Environ Microbiol* **84**, doi:10.1128/aem.01298-18 (2018).

13 Mueller, T. J., Welsh, E. A., Pakrasi, H. B. & Maranas, C. D. Identifying Regulatory Changes to Facilitate Nitrogen Fixation in the Nondiazotroph Synechocystis sp. PCC 6803. *ACS Synthetic Biology* **5**, 250-258, doi:10.1021/acssynbio.5b00202 (2016).

14 Berg, H. *et al.* Biosynthesis of the cyanobacterial reserve polymer multi-L-arginyl-poly-L-aspartic acid (cyanophycin): mechanism of the cyanophycin synthetase reaction studied with synthetic primers. *Eur J Biochem* **267**, 5561-5570, doi:10.1046/j.1432-1327.2000.01622.x (2000).

15 Simon, R. D. The biosynthesis of multi-L-arginyl-poly(L-aspartic acid) in the filamentous cyanobacterium Anabaena cylindrica. *Biochim Biophys Acta* **422**, 407-418, doi:10.1016/0005-2744(76)90151-0 (1976).

16 Sallam, A. & Steinbüchel, A. Dipeptides in nutrition and therapy: cyanophycin-derived dipeptides as natural alternatives and their biotechnological production. *Appl Microbiol Biotechnol* **87**, 815-828, doi:10.1007/s00253-010-2641-0 (2010).

17 Hai, T., Oppermann-Sanio, F. B. & Steinbüchel, A. Molecular characterization of a thermostable cyanophycin synthetase from the thermophilic cyanobacterium Synechococcus sp. strain MA19 and in vitro synthesis of cyanophycin and related polyamides. *Appl Environ Microbiol* **68**, 93-101, doi:10.1128/aem.68.1.93-101.2002 (2002).

18 Merritt, M. V., Sid, S. S., Mesh, L. & Allen, M. M. Variations in the amino acid composition of cyanophycin in the cyanobacterium Synechocystis sp. PCC 6308 as a function of growth conditions. *Arch Microbiol* **162**, 158-166, doi:10.1007/bf00314469 (1994).

19 Steinle, A., Bergander, K. & Steinbüchel, A. Metabolic engineering of Saccharomyces cerevisiae for production of novel cyanophycins with an extended range of constituent amino acids. *Appl Environ Microbiol* **75**, 3437-3446, doi:10.1128/aem.00383-09 (2009).

20 Wiefel, L. & Steinbüchel, A. Solubility behavior of cyanophycin depending on lysine content. *Applied and environmental microbiology* **80**, 1091-1096, doi:10.1128/AEM.03159-13 (2014).

21 Klemke, F. *et al.* CphA2 is a novel type of cyanophycin synthetase in N2-fixing cyanobacteria. *Microbiology* **162**, 526-536, doi:https://doi.org/10.1099/mic.0.000241 (2016).

22 Sharon, I., Grogg, M., Hilvert, D. & Schmeing, T. M. Structure and Function of the β-Asp-Arg Polymerase Cyanophycin Synthetase 2. *ACS Chemical Biology* **17**, 670-679, doi:10.1021/acschembio.1c01007 (2022).

23 Aboulmagd, E., Oppermann-Sanio, F. B. & Steinbüchel, A. Purification of Synechocystis sp. strain PCC6308 cyanophycin synthetase and its characterization with respect to substrate and primer specificity. *Appl Environ Microbiol* **67**, 2176-2182, doi:10.1128/aem.67.5.2176-2182.2001 (2001).

24 Arai, T. & Kino, K. A cyanophycin synthetase from Thermosynechococcus elongatus BP-1 catalyzes primer-independent cyanophycin synthesis. *Appl Microbiol Biotechnol* **81**, 69-78, doi:10.1007/s00253-008-1623-y (2008).

25 Frommeyer, M., Wiefel, L. & Steinbüchel, A. Features of the biotechnologically relevant polyamide family "cyanophycins" and their biosynthesis in prokaryotes and eukaryotes. *Crit Rev Biotechnol* **36**, 153-164, doi:10.3109/07388551.2014.946467 (2016).

26 Mooibroek, H. *et al.* Assessment of technological options and economical feasibility for cyanophycin biopolymer and high-value amino acid production. *Applied Microbiology and Biotechnology* **77**, 257-267, doi:10.1007/s00253-007-1178-3 (2007).

27 Schwamborn, M. Chemical synthesis of polyaspartates: a biodegradable alternative to currently used polycar□ylate homo- and copolymers. *Polymer Degradation and Stability* **59**, 39-45, doi:https://doi.org/10.1016/S0141-3910(97)00184-5 (1998).

28 Du, J., Li, L. & Zhou, S. Enhanced cyanophycin production by Escherichia coli overexpressing the heterologous cphA gene from a deep sea metagenomic library. *J Biosci Bioeng* **123**, 239-244, doi:10.1016/j.jbiosc.2016.08.008 (2017).

29 Ventola, C. L. The antibiotic resistance crisis: part 1: causes and threats. *P T* **40**, 277-283 (2015).

30 Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat Prod Rep* **18**, 380-416 (2001).

31 Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J Ind Microbiol Biotechnol* **43**, 155-176, doi:10.1007/s10295-015-1723-5 (2016).

32 Wang, B., Guo, F., Huang, C. & Zhao, H. Unraveling the iterative type I polyketide synthases hidden in *Streptomyces*. *Proceedings of the National Academy of Sciences* **117**, 8449-8454, doi:10.1073/pnas.1917664117 (2020).

33 Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Natural Product Reports* **29**, 1050, doi:10.1039/c2np20019h (2012).

34 Yang, X. L. *et al.* Molecular basis of methylation and chain-length programming in a fungal iterative highly reducing polyketide synthase. *Chem Sci* **10**, 8478-8489, doi:10.1039/c9sc03173a (2019).

35 McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta Crystallographica Section F Structural Biology Communications* **70**, 2-20, doi:10.1107/s2053230x13033141 (2014).

36 Gawas, U. B., Mandrekar, V. K. & Majik, M. S. in *Advances in Biological Science Research* 69-84 (2019).

37 Pichlo, C., Montada, A. A., Schacherl, M. & Baumann, U. Production, Crystallization and Structure Determination of C. difficile PPEP-1 via Microseeding and Zinc-SAD. *Journal of Visualized Experiments*, doi:10.3791/55022 (2016).

38 Skarina, T., Xu, X., Evdokimova, E. & Savchenko, A. in *Structural Genomics and Drug Discovery: Methods and Protocols* (ed Wayne F. Anderson) 159-168 (Springer New York, 2014).

39 Bijelic, A. & Rompel, A. Polyoxometalates: more than a phasing tool in protein crystallography. *ChemTexts* **4**, 10, doi:10.1007/s40828-018-0064-1 (2018).

40 Pflugrath, J. W. Practical macromolecular cryocrystallography. *Acta Crystallogr F Struct Biol Commun* **71**, 622-642, doi:10.1107/S2053230X15008304 (2015).

41 Nave, C. Protein crystallography on a synchrotron. *Synchrotron Radiation News* **2**, 24-28, doi:10.1080/08940888908261212 (1989).

42 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* **60**, 2126-2132, doi:doi:10.1107/S0907444904019158 (2004).

43 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

44 Rossmann, M. G., Battisti, A. J. & Plevka, P. in *Recent Advances in Electron Cryomicroscopy, Part B Advances in Protein Chemistry and Structural Biology* 101-121 (2011).

45 Milne, J. L. S. *et al.* Cryo-electron microscopy--a primer for the non-microscopist. *FEBS J* **280**, 28-45, doi:10.1111/febs.12078 (2013).

46 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

47 Doerr, A. Single-particle cryo-electron microscopy. *Nature Methods* **13**, 23-23, doi:10.1038/nmeth.3700 (2016).

48 Egelman, E. H. The Current Revolution in Cryo-EM. *Biophys J* **110**, 1008-1012, doi:10.1016/j.bpj.2016.02.001 (2016).

49 Sørensen, H. P. & Mortensen, K. K. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial Cell Factories* **4**, 1, doi:10.1186/1475-2859-4-1 (2005).

50 Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **5**, 172-172, doi:10.3389/fmicb.2014.00172 (2014).

51 Jia, B. & Jeon, C. O. High-throughput recombinant protein expression in Escherichia coli : current status and future perspectives. *Open Biology* **6**, 160196, doi:10.1098/rsob.160196 (2016).

52 Briand, L. *et al.* A self-inducible heterologous protein expression system in Escherichia coli. *Scientific Reports* **6**, 33037, doi:10.1038/srep33037 (2016).

53 Frey, K. M., Oppermann-Sanio, F. B., Schmidt, H. & Steinbüchel, A. Technical-Scale Production of Cyanophycin with Recombinant Strains of Escherichia coli. *Applied and Environmental Microbiology* **68**, 3377-3384, doi:10.1128/aem.68.7.3377-3384.2002 (2002).

54 Wingfield, P. T. Overview of the purification of recombinant proteins. *Curr Protoc Protein Sci* **80**, 6.1.1-6.1.35, doi:10.1002/0471140864.ps0601s80 (2015).

55 Trabbic-Carlson, K., Liu, L., Kim, B. & Chilkoti, A. Expression and purification of recombinant proteins from Escherichia coli: Comparison of an elastin-like polypeptide fusion with an oligohistidine fusion. *Protein Sci* **13**, 3274-3284, doi:10.1110/ps.04931604 (2004).

56 Proteomics, C. *Size Exclusion Chromatography (SEC) Service*, <https://www.creative-proteomics.com/pronalyse/size-exclusion-chromatography-sec-service.html> (2013).

57 Weber, K. New structural model of E. coli aspartate transcarbamylase and the amino-acid sequence of the regulatory polypeptide chain. *Nature* **218**, 1116-1119, doi:10.1038/2181116a0 (1968).

58 Immunotherapies, R. *SDS-PAGE Technique*, <https://www.rockland.com/resources/sds-page-technique/> (

59 Recht, M. I. *et al.* Measurement of enzyme kinetics and inhibitor constants using enthalpy arrays. *Anal Biochem* **388**, 204-212, doi:10.1016/j.ab.2009.02.028 (2009).

60 Aravind, J., Saranya, T., Sudha, G. & Palanisamy, K. 49-59 (2016).

61 Brady, C. L. *et al.* Transfer of Pantoea citrea, Pantoea punctata and Pantoea terrea to the genus Tatumella emend. as Tatumella citrea comb. nov., Tatumella punctata comb. nov. and Tatumella terrea comb. nov. and description of Tatumella morbirosei sp. nov. *Int J Syst Evol Microbiol* **60**, 484-494, doi:10.1099/ijs.0.012070-0 (2010).

62 van Heijenoort, J. Recent advances in the formation of the bacterial peptidoglycan monomer unit. *Nat Prod Rep* **18**, 503-519, doi:10.1039/a804532a (2001).

63 Ziegler, K. *et al.* Molecular characterization of cyanophycin synthetase, the enzyme catalyzing the biosynthesis of the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartate (cyanophycin). *Eur J Biochem* **254**, 154-159, doi:10.1046/j.1432-1327.1998.2540154.x (1998).

64 Smith, C. A. Structure, function and dynamics in the mur family of bacterial cell wall ligases. *J Mol Biol* **362**, 640-655, doi:10.1016/j.jmb.2006.07.066 (2006).

65 Sharon, I. *et al.* A cryptic third active site in cyanophycin synthetase creates primers for polymerization. *Nature Communications* **13**, 3923, doi:10.1038/s41467-022-31542-7 (2022).

66 Kroll, J., Klinter, S. & Steinbüchel, A. A novel plasmid addiction system for large-scale production of cyanophycin in Escherichia coli using mineral salts medium. *Applied Microbiology and Biotechnology* **89**, 593-604, doi:10.1007/s00253-010-2899-2 (2011).

67 Aboulmagd, E., Oppermann-Sanio, F. B. & Steinbüchel, A. Molecular characterization of the cyanophycin synthetase from Synechocystis sp. strain PCC6308. *Arch Microbiol* **174**, 297-306, doi:10.1007/s002030000206 (2000).

68 Hai, T., Frey, K. M. & Steinbüchel, A. Engineered cyanophycin synthetase (CphA) from Nostoc ellipsosporum confers enhanced CphA activity and cyanophycin accumulation to Escherichia coli. *Appl Environ Microbiol* **72**, 7652-7660, doi:10.1128/aem.01132-06 (2006).

69 Krehenbrink, M. & Steinbüchel, A. Partial purification and characterization of a non-cyanobacterial cyanophycin synthetase from Acinetobacter calcoaceticus strain ADP1 with regard to substrate specificity, substrate affinity and binding to cyanophycin. *Microbiology* **150**, 2599-2608, doi:https://doi.org/10.1099/mic.0.27140-0 (2004).

70 Frey, K. M., Oppermann-Sanio, F. B., Schmidt, H. & Steinbüchel, A. Technical-scale production of cyanophycin with recombinant strains of Escherichia coli. *Appl Environ Microbiol* **68**, 3377-3384, doi:10.1128/aem.68.7.3377-3384.2002 (2002).

71 Khlystov, N. A. *et al.* Material properties of the cyanobacterial reserve polymer multi-l-arginyl-poly-l-aspartate (cyanophycin). *Polymer* **109**, 238-245, doi:https://doi.org/10.1016/j.polymer.2016.11.058 (2017).

72 Marahiel, M. A. & Essen, L. O. in *Methods in Enzymology* Vol. 458 337-351 (Academic Press, 2009).

73 Sharon, I., Grogg, M., Hilvert, D. & Schmeing, T. The structure of cyanophycinase in complex with a cyanophycin degradation intermediate. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1866**, 130217, doi:10.1016/j.bbagen.2022.130217 (2022).

74 Miyakawa, T. *et al.* Structural bases for aspartate recognition and polymerization efficiency of cyanobacterial cyanophycin synthetase. *Nature Communications* **13**, doi:10.1038/s41467-022-32834-8 (2022).

75 Elbahloul, Y. & Steinbüchel, A. Engineering the genotype of Acinetobacter sp. strain ADP1 to enhance biosynthesis of cyanophycin. *Applied and environmental microbiology* **72**, 1410-1419, doi:10.1128/AEM.72.2.1410-1419.2006 (2006).

76 Scott, T. A. & Piel, J. The hidden enzymology of bacterial natural product biosynthesis. *Nat. Rev. Chem.* **3**, 404-425, doi:10.1038/s41570-019-0107-1 (2019).

77 Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43**, 155-176, doi:10.1007/s10295-015-1723-5 (2016).

78 Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew Chem. Int. Ed. Engl.* **48**, 4688-4716, doi:10.1002/anie.200806121 (2009).

79 Dodge, G. J., Ronnow, D., Taylor, R. E. & Smith, J. L. Molecular basis for olefin rearrangement in the gephyronic acid polyketide synthase. *ACS Chem. Biol.* **13**, 2699-2707, doi:10.1021/acschembio.8b00645 (2018).

80 Moldenhauer, J. *et al.* The final steps of bacillaene biosynthesis in *Bacillus amyloliquefaciens* FZB42: direct evidence for β,γ dehydration by a *trans*-

acyltransferase polyketide synthase. *Angew Chem. Int. Ed. Engl.* **49**, 1465-1467, doi:10.1002/anie.200905468 (2010).

81 Lohr, F. *et al.* α,β→β,γ-double bond migration in corallopyronin A biosynthesis. *Chem. Sci.* **4**, 4175-4180 (2013).

82 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* **291**, 1790-1792, doi:10.1126/science.1058092 (2001).

83 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).

84 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallographica Section D Biological Crystallography* **69**, 1204-1214, doi:10.1107/s0907444913000061 (2013).

85 Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. BALBES: a molecular-replacement pipeline. *Acta Crystallographica Section D* **64**, 125-132, doi:doi:10.1107/S0907444907050172 (2008).

86 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D* **67**, 355-367, doi:doi:10.1107/S0907444911001314 (2011).

87 Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M. & Smith, J. L. Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure* **4**, 253-264, doi:10.1016/s0969-2126(96)00030-5 (1996).

88 Brock, D. J., Kass, L. R. & Bloch, K. β-hydroxydecanoyl thioester dehydrase. II. Mode of action. *J. Biol. Chem.* **242**, 4432-4440 (1967).

89 Labonte, J. W. & Townsend, C. A. Active site comparisons and catalytic mechanisms of the hot dog superfamily. *Chem. Rev.* **113**, 2182-2204, doi:10.1021/cr300169a (2013).

90 Gay, D. C., Spear, P. J. & Keatinge-Clay, A. T. A double-hotdog with a new trick: structure and mechanism of the *trans*-acyltransferase polyketide synthase enoyl-isomerase. *ACS Chem. Biol.* **9**, 2374-2381, doi:10.1021/cb500459b (2014).

91 Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188-1190, doi:10.1101/gr.849004 (2004).

92 Lu, Y.-J., White, S. W. & Rock, C. O. Domain swapping between *Enterococcus faecalis* FabN and FabZ proteins localizes the structural determinants for

isomerase activity. *J. Biol. Chem.* **280**, 30342-30348, doi:10.1074/jbc.M504637200 (2005).

93 Kimber, M. S. *et al.* The structure of (3*R*)-hydroxyacyl-acyl carrier protein dehydratase (FabZ) from *Pseudomonas aeruginosa*. *J. Biol. Chem.* **279**, 52593-52602, doi:10.1074/jbc.M408105200 (2004).

94 Heil, C. S., Wehrheim, S. S., Paithankar, K. S. & Grininger, M. Fatty acid biosynthesis: chain-length regulation and control. *ChemBioChem* **20**, 2298-2321 (2019).

95 Rando, R. R. & Bloch, K. Mechanism of Action of β-Hydroxydecanoyl Thioester Dehydrase. *J. Biol. Chem.* **243**, 5627-5634, doi:https://doi.org/10.1016/S0021-9258(18)91913-7 (1968).

96 Roberts, D. M. *et al.* Substrate selectivity of an isolated enoyl reductase catalytic domain from an iterative highly reducing fungal polyketide synthase reveals key components of programming. *Chem. Sci.* **8**, 1116-1126, doi:10.1039/c6sc03496a (2017).

97 Jenner, M. *et al.* Acyl-chain elongation drives ketosynthase substrate selectivity in *trans*-acyltransferase polyketide synthases. *Angew Chem. Int. Ed. Engl.* **54**, 1817-1821, doi:https://doi.org/10.1002/anie.201410219 (2015).

98 Roberts, D. M. *et al.* Substrate selectivity of an isolated enoyl reductase catalytic domain from an iterative highly reducing fungal polyketide synthase reveals key components of programming. *Chem. Sci.* **8**, 1116-1126 (2017).

99 Gay, D. C., Spear, P. J. & Keatinge-Clay, A. T. A double-hotdog with a new trick: Structure and mechanism of the *trans*-acyltransferase polyketide synthase enoyl-isomerase. *ACS Chem. Biol.* **9**, 2374-2381, doi:10.1021/cb500459b (2014).

APPENDIX A

DATA COLLECTED AUGUST 2021-SEPTEMBER 2022

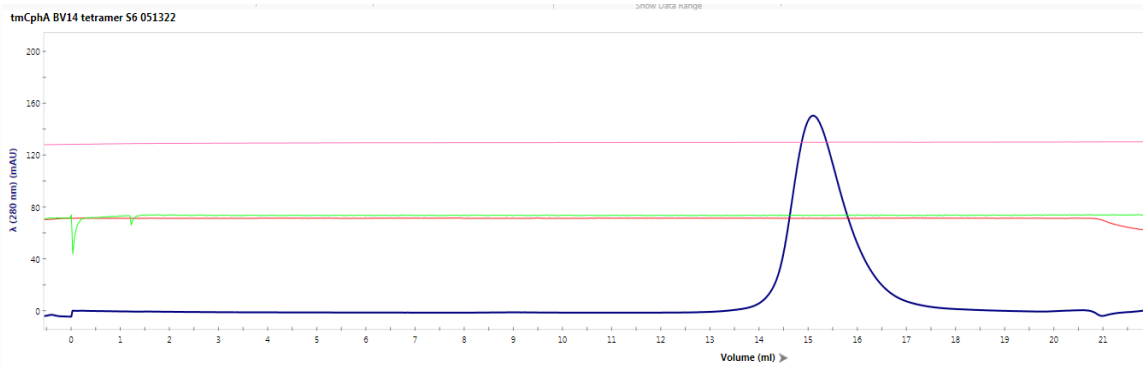


Figure S25: Sizing trace of tmCphA demonstrating purity of the tetramer

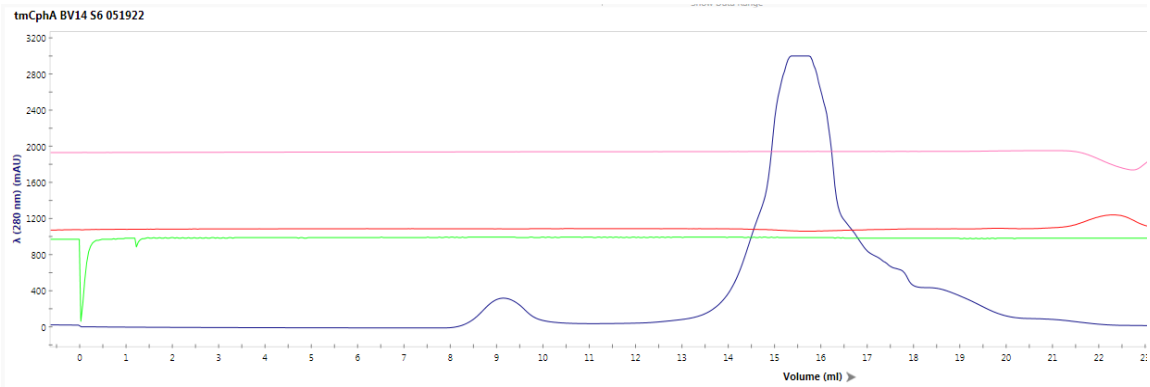


Figure S26: Sizing trace of tmCphA demonstrating the overall productivity of the strain. The first small peak is the aggregates and the void volume, while the second peak is the tmCphA.

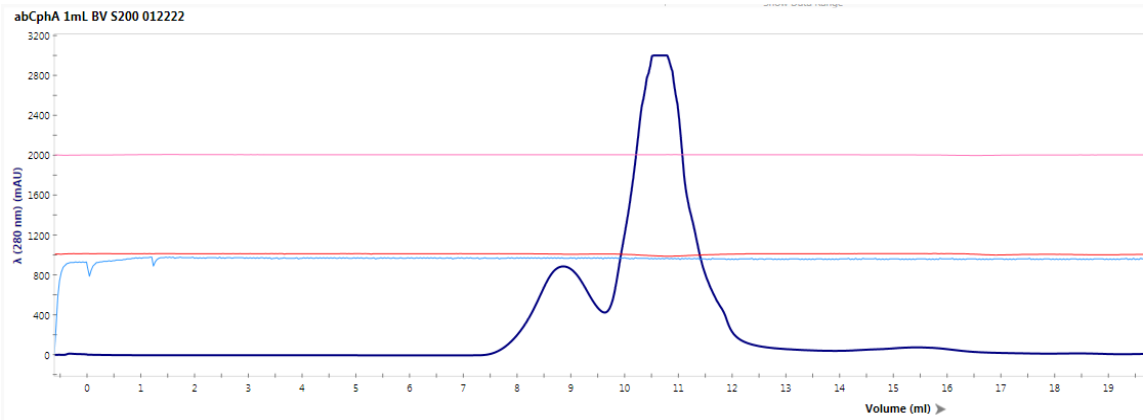


Figure S27: abCphA SEC run. The first peak is the void volume while the second peak is the purified abCphA showing tetrameric architecture based on the size.

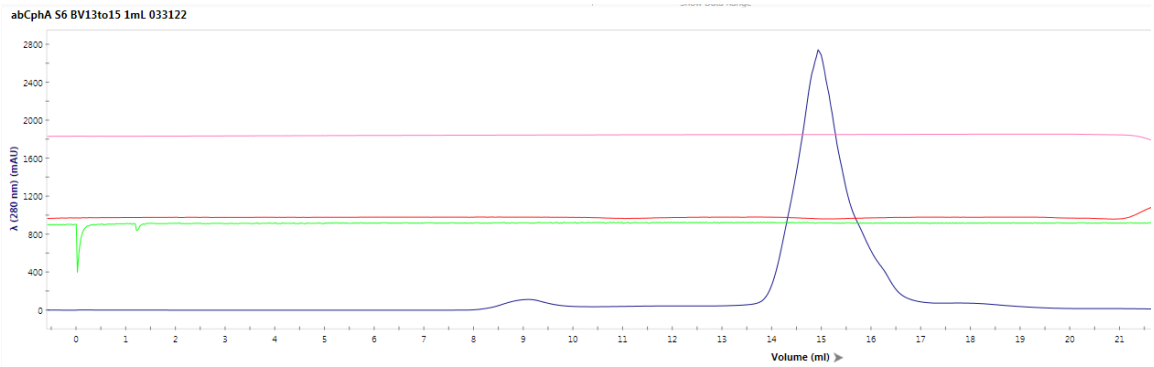


Figure S28: abCphA G217K SEC run demonstrating sample purity.

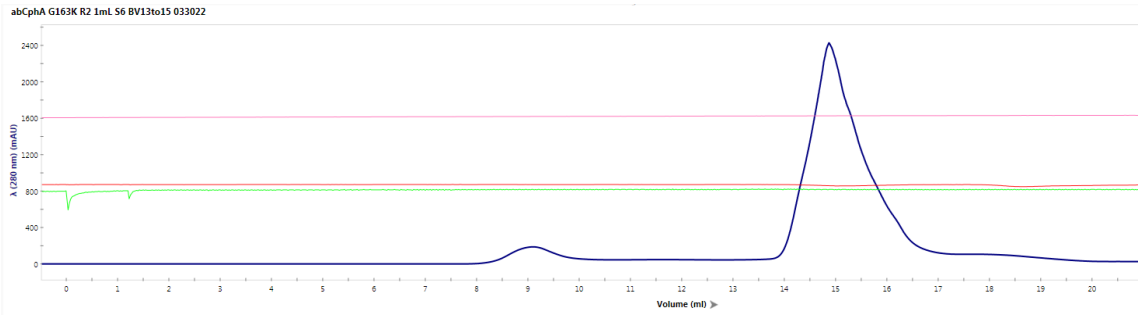


Figure S29: abCphA G163K SEC run demonstrating sample purity 72

APPENDIX B

DATA COLLECTED JANUARY 2021-FEBRUARY 2022

EXTENDED MATERIALS AND METHODS FOR CHAPTER 4

Materials and Methods

plasmid	primer name	primer sequence
pET28a-alnB-his	alnB-F11	gtttaactttaagaaggagatataccatggTGAATGCTGCCCA CGGCT
	alnB-R11	cgacggcttcggcGACCATGCGATTGCACCG
	alnB-F21	caatcgcatggtcGCCGAAGCCGTCGCCGAA
	alnB-R21	cagtggtggtggtggtggtgctcgagtgcggccgcaagcttGTGGC CGTCCACAAGGCTC
pET28a-his-alnB_DH	alnB_DH-F	agcaaatgggtcgcg*gatcc*CTGGAAACGCCGTTCTCG
	alnB_DH-R	tcgagtgcggccgca*agctt*TCACGTCCGCTCGCGTAC
pET28a-his-alnB_DHmtH49A	alnB_DHmtH49A-F2	GCTCGAGGACgcgCGCGTGCAGGGgccg
	alnB_DHmtH49A-R2	cacgggaacgtatcggtgtccatcgtcagt
pET28a-his-alnB_DHmtD213A	alnB_DHmtD213A-F	GGCGCTGCTCgcgGCGTGCTTCC
	alnB_DHmtD213A-R	GGGTGGATCAGGTGCCGC
pET28a-his-alnB_DHmtH49F	alnB_DHmtH49F-F	GCTCGAGGACttcCGCGTGCAGG
	alnB_DHmtH49F-R	CACGGGAACGTATCGGTG
pET28a-his-alnB_DHmtD213N	alnB_DHmtD213N-F	GGCGCTGCTCaatGCGTGCTTCC
	alnB_DHmtD213N-R	GGGTGGATCAGGTGCCGC
	alnB_DHmtP58-R	tcggcccctgcacgcgGTGgtcctcga
pET28a-his-AlnB_DHmtP58A	alnB_DHmtP58A-F	tcgtgttcGCCggcgcgggccacctcga
pET28a-his-AlnB_DHmtP58V	alnB_DHmtP58V-F	tcgtgttcGTCggcgcgggccacctcga
	alnB_DHmtS220-R	aagcacgcatcgagcagcgccgggt
pET28a-his-AlnB_DHmtS220A	alnB_DHmtS220A-F	ccagtcgctgGCCatcgcgatgggcaacg
pET28a-his-AlnB_DHmtS220V	alnB_DHmtS220V-F	ccagtcgctgGTCatcgcgatgggcaacg
pET28a-his-AlnB_DHmtS220C	alnB_DHmtS220C-F	ccagtcgctgTGCatcgcgatgggcaacg
pET28a-his-AlnB_DHmtS220T	alnB_DHmtS220T-F	ccagtcgctgACCatcgcgatgggcaacg
pET28a-his-AlnB_DHmtS220N	alnB_DHmtS220N-F	ccagtcgctgAACatcgcgatgggcaacg
pET28a-his-AlnB_DHmtS220I	alnB_DHmtS220I-F	ccagtcgctgATCatcgcgatgggcaacg
	alnB_DHmtH62-R	gggaacacgatcggcccctgcacgcgg
pET28a-his-AlnB_DHmtH62T	alnB_DHmtH62T-F	cggcgcgggcACCctcgatctcgtcgtcgg
pET28a-his-AlnB_DHmtH62Y	alnB_DHmtH62Y-F	cggcgcgggcTACctcgatctcgtcgtcgg
pET28a-his-AlnB_DHmtH62L	alnB_DHmtH62L-F	cggcgcgggcCTGctcgatctcgtcgtcgg
pET28a-his-AlnB_DHmtH62F	alnB_DHmtH62F-F	cggcgcgggcTTCctcgatctcgtcgtcgg

pET28a-his-alnB_ERv2	alnB_ERv2-F	tggacagcaaatgggtcgcg*gatcc*TCGGACGGGACCACG TTC
	alnB_ERv2-R	ggtgctcgagtgcggccgca*agctt*TTAATCGAATCCGACG ACGACC
pET28a-his-alnB_ERmtR46A	alnB_ERmtR46A-F2	GCcgacatcatgctgaccctcg
	alnB_ERmtR46A-R2	gaagttcaagccggtcgcc
pET28a-his-alnB_ERmtM129A	alnB_ERmtM129A-F	GACGCTGCCGgcGGCGTATCTGACCG
	alnB_ERmtM129A-R	GCGGCGTCAGCGTCGGAC
pET28a-his-alnB_KSATDH^0	AlnB_KSATDHwDHnul-F1	tggacagcaaatgggtcgcg*gatcc*ATGGTGAATGCTGCCC AC
	AlnB_KSATDHwDHnul-R1	gcacgcgggcGTCCTCGAGCCACGGGAA
	AlnB_KSATDHwDHnul-F2	gctcgaggacGCCCGCGTGCAGGGGCCG
	AlnB_KSATDHwDHnul-R2	ggtgctcgagtgcggccgca*agctt*TCACGTCCGCTCGCGT ACCGGC
pET28a-his-alnB_KSAT	alnB_KSATv3-F	tggacagcaaatgggtcgcggatccATGGTGAATGCTGCCC AC
	alnB_KSATv3-R	ggtgctcgagtgcggccgcaagcttTCACGTTTCCAGCCAG TGCTCTTCGC
pET28a-alnB_KSAT-his	alnB_KSATv4-F	actttaagaaggagatataccatggTGAATGCTGCCCACGG CT
	alnB_KSATv4-R	ggtgctcgagtgcggccgcaagcttCGTTTCCAGCCAGTGC TC
pET28a-his-alnB_KS	alnB_KS-F	tggacagcaaatgggtcgcggatccATGGTGAATGCTGCCC AC
	alnB_KS-R	ggtgctcgagtgcggccgcaagcttTTAGAAGGCGACATTG CTC
pET28a-alnB_KS-his	alnB_KS-F2	actttaagaaggagatataccatggATGCTGCCCACGGCTC CC
	alnB_KS-R2	ggtgctcgagtgcggccgcaagcttAGCTGCCACGATCAC GTGC
pET28a-his-alnB_ACPv2	alnB_ACP-F	tggacagcaaatgggtcgcg*gatcc*CTGCGGGCAGAACTCC TCGAACTGC
	alnB_ACP-R	ggtgctcgagtgcggccgca*agctt*TCACGCCGCGCCGGTG GA

In vitro reactions

Standard reaction for AlnB

Reagent	Stock conc.	Add volume, □l	Final conc.
H_2O		32	
Tris-HCl, pH 8.0	1 M	5	100 mM
TCEP, pH7.2	0.5 M	1	10 mM
Glucose	100 mM	1	2 mM
$NADP^+$	50 mM	1	1 mM
GDH, 40% glycerol	0.1 mM	1	2 □M
SAM	10 mM	1	200 □M
Acetyl-CoA	10 mM	1	200 □M
Malonyl-CoA	50 mM	2	2 mM
AlnB-his, 40% glycerol	23 □M	5	1.2 □M

Total		50	2.4% glycerol

Reaction was performed at 25°C overnight and then equal volume of (MeOH:DMSO 4:1) was added to stop the reaction, precipitate protein, and dissolve the products. After centrifugation at 13000×*g* for 5 min, 10 □l of supernatant was injected onto HPLC.

TCEP derivation assay

In the reaction buffer, 20 mM (final conc.) of TCEP was incubated with 600 □M (final conc.) of all four PANTs for 1 h, and then subjected to HPLC analysis.

Assays for enoyl reduction by AlnB_ER and its mutants

Reagent	Stock conc.	Add volume, □l	Final conc.
H_2O		18.5	
Tris-HCl, pH 8.0	1 M	2.5	50 mM
TCEP, pH 7.2	0.5 mM	0.5	10 □M
Glucose	100 mM	0.5	2 mM
$NADP^+$	50 mM	0.5	1 mM
GDH (glucose dehydrogenase in 50% glycerol)	0.1 mM	0.5	20 □M
□□□-PANT substrates	15 mM	1	600 □M
His-AlnB_ER or M129A mutant in 40% glycerol	500 □M	1	20 □M
Total		25	

Reaction was performed at 25°C for 1 h., and then equal volume of MeOH was added to stop the reaction, precipitate protein, and dissolve the products. After centrifugation at 13000×*g* for 5 min, 20 □l of supernatant was injected onto HPLC.

Assays for chain elongation reaction by AlnB_KSATDH^0

Reagent	Stock conc.	Add volume, □l	Final conc.
H_2O		80	
Tris-HCl, pH 8.0	1 M	10	100 mM
TCEP, pH 7.2	0.5 mM	2	10 □M
Oct-2-enoyl-PANT substrates	15 mM	1	150 □M
Oct-3-enoyl-PANT substrates	15 mM	1	150 □M
Hex-2-enoyl-PANT substrates	15 mM	1	150 □M
Hex-3-enoyl-PANT substrates	15 mM	1	150 □M
Malonyl-CoA	50 mM	2	1 mM
His- AlnB_ACPv2	400 □M	3	12 □M
His-AlnB_KSATDH^0	300 □M	1	3 □M

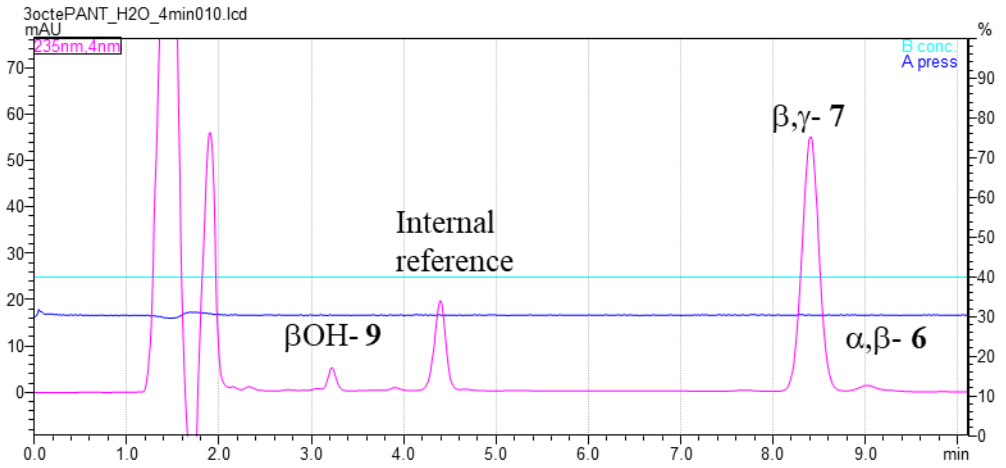
Total		100	

Reaction was performed at 25°C for 2 h., and then dialyzed against water for 2 h, finally injected 2 ul for LC-HRMS analysis.

Time-course experiment for AlnB_DH

Standard reaction condition was used for AlnB_DH time-course experiment. An isocratic elution program, 40% of ACN (0.1% TFA) holding for 10 min, was used for better separation of isomers and calculation. Reactions were stopped by immediately mixing with 2-fold volume of pre-cooled MeOH. Areas at 235 nm were calculated after background subtraction.

RT	**□□□□ 9**				**□□□□ 6**				**□□□□ 7**				***K***
min	#1	#2	aver	stdeva	#1	#2	aver	stdeva	#1	#2	aver	stdeva	□OH-/□□□-
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	527.2	509.3	518.3	12.7	
5	13.7	14.2	14.0	0.4	16.8	17.1	17.0	0.2	478.6	490.6	484.6	8.5	0.8
10	26.1	25.9	26.0	0.1	25.2	24.6	24.9	0.4	462.0	460.9	461.5	0.8	1.0
30	78.4	75.8	77.1	1.8	51.3	49.4	50.4	1.3	400.9	386.1	393.5	10.5	1.5
50	114.0	119.7	116.9	4.0	69.3	72.6	71.0	2.3	318.6	333.4	326.0	10.5	1.6



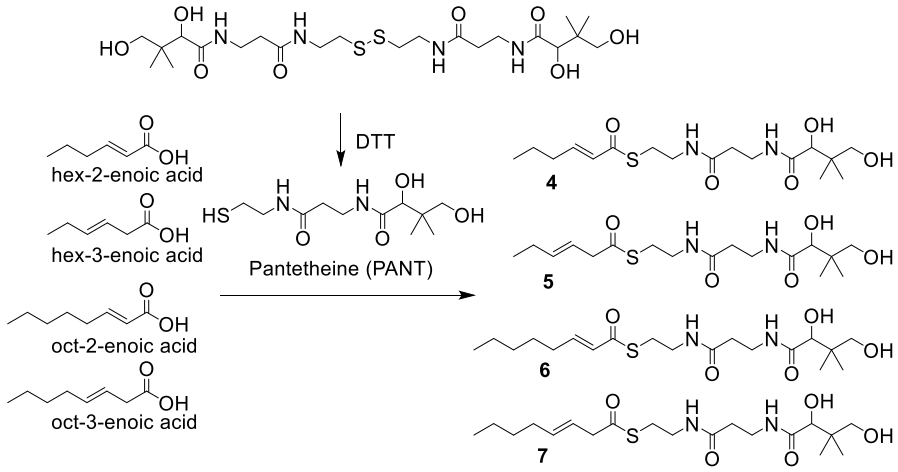
HPLC and LC-MS analyses

For analysis of allenomycin heptaketide (**1**) and octaketide (**2**), *E. coli* BAP1[82] was grown and induced overnight in 50 ml of LB medium. Fermentation broth was collected after centrifugation at 4000 rpm for 10 min and then extracted by equal volume of ethyl acetate after adjusting pH to 3.0. Organic solvent was removed by vacuum evaporation and the extract was dissolved in methanol. After centrifugation at maxi speed for 5 min, the crude extract was injected into HPLC.

HPLC analysis of reactions of AlnB holo enzyme and AlnB_ER was carried out on Agilent 1260, equipped with diode-array detection (DAD) detector and analytical column Kinetex SB-C18 (4.6 × 180 mm, 5 μm) at a flow rate of 1.0 mL/min. HPLC analysis of

AlnB_DH and all of its mutants were conducted on Shimadzu M20 equipped with PDA M40 DAD detector and analytic column Agilent ZORBAX Eclipse Plus (4.6 × 150 mm, 5 μm) at a flow rate of 1.0 mL/min. Wavelengths were set to 320 and 360 nm for □□□-polyene compound detection while set to 235 nm for detection of PANT substrates and products. Solvent A was water with 0.1% trifluoroacetic acid (TFA), and Solvent B was acetonitrile with 0.1% TFA. Same program was used: 20 to 100% B in 12 min, then held at 100% B for 2 min, 100 to 20% B in 0.5 min, and held 20% B for 3 min. For Liquid Chromatography–Mass Spectrometry (LC-MS) analysis, ESI positive ion mode (Bruker, Amazon SL Ion Trap) was used, equipped with a Kinetex 2.6-μm XB-C18 100 Å (Phenomenex).

Compound synthesis



Reagents	Mr. (g/mol)	Amount (mmol)	Mass (mg)	Vol. (□l)	Products
hex-2-enoic acid	142	1	-	157	**4**
hex-3-enoic acid			-		**5**
oct-2-enoic acid	170	1	-	188	**6**
oct-3-enoic acid			-		**7**
PANT	119	1.2		134	
EDC	192	2	392		
DMAP	122	0.5	62		

Briefly, carboxylic acid was dissolved in 5 ml of CH_2Cl_2, then EDC (3-(ethyliminomethyleneamino)-*N*,*N*-dimethylpropan-1-amine), PANT (R-pantetheine), and DMAP (4-dimethylaminopyridine) were added. After stirring at room temperature for overnight, the reaction mixture was dried and then loaded onto silica gel for column chromatography, finally eluted by mixtures of *n*-hexane/ethyl acetate. Fractions containing product were pooled and dried under vacuum and dissolved in DMSO.

(*E*)-hex-2-enoyl-PANT (**4**)
^{1}H NMR (600 MHz, $CDCl_3$): δ_H 0.85 (s, 3H), 0.87 (t, *J* = 7.3 Hz, 3H), 0.92 (s, 3H), 1.44 (m, 2H), 2.11 (t, *J* = 6.9 Hz, 2H), 2.34 (t, *J* = 5.9 Hz, 2H), 3.01 (m, 2H), 3.30-3.51 (m,

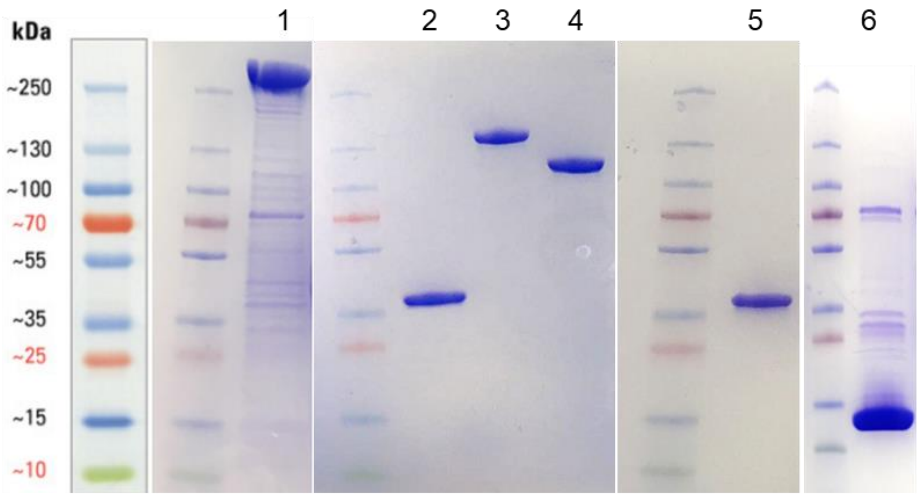
6H), 3.89 (d, *J* = 4.9 Hz, 1H), 4.03 (brs, 1H), 4.67 (d, *J* = 4.9 Hz, 1H), 6.05 (dt, *J* = 15.5, 1.5 Hz, 1H), 6.84 (dt, *J* = 15.5, 7.0 Hz, 1H), 7.38 (t, *J* = 5.3 Hz, 1H). ^{13}C NMR (150 MHz, $CDCl_3$): δ_C 13.5, 20.3, 21.0, 21.2, 27.9, 34.0, 35.0, 35.6, 39.1, 40.8, 70.7, 77.2, 128.3, 146.1, 171.4, 173.6, 189.7. The spectroscopic data matched well with those reported values[98].

(*E*)-hex-3-enoyl-PANT (**5**)
^{1}H NMR (600 MHz, $CDCl_3$): δ_H 0.85 (s, 3H), 0.92 (s, 3H), 0.93 (t, *J* = 7.4 Hz, 3H), 2.00 (m, 2H), 2.34 (t, *J* = 6.1 Hz, 2H), 2.94 (m, 2H), 3.18 (d, *J* = 7.0 Hz, 2H), 3.26-3.50 (m, 6H), 3.90 (d, *J* = 5.1 Hz, 1H), 4.02 (t, *J* = 5.7 Hz, 1H), 4.65 (d, *J* = 5.0 Hz, 1H), 5.42 (m, 1H), 5.60 (m, 1H), 6.77 (t, *J* = 5.0 Hz, 1H), 7.38 (t, *J* = 5.7 Hz, 1H). ^{13}C NMR (150 MHz, $CDCl_3$): δ_C 13.3, 20.5, 21.5, 25.5, 28.4, 35.2, 35.8, 39.3, 41.0, 47.5, 71.0, 77.5, 119.8, 138.2, 171.5, 173.7, 198.4. The spectroscopic data matched well with those reported values[99].

(*E*)-oct-2-enoyl-PANT (**6**)
^{1}H NMR (600 MHz, $CDCl_3$): δ_H 0.83 (t, *J* = 7.0 Hz, 3H), 0.85 (s, 3H), 0.93 (s, 3H), 1.19-1.29 (m, 4H), 1.41 (m, 2H), 2.14 (m, 2H), 2.35 (t, *J* = 6.0 Hz, 2H), 3.01 (m, 2H), 3.31-3.39 (m, 4H), 3.49 (m, 2H), 3.90 (d, *J* = 5.2 Hz, 1H), 4.06 (t, *J* = 5.8 Hz, 1H), 4.64 (d, *J* = 5.0 Hz, 1H), 6.05 (d, *J* = 15.5 Hz, 1H), 6.77 (t, *J* = 5.6 Hz, 1H), 6.86 (dt, *J* = 15.5, 7.0 Hz, 1H), 7.38 (t, *J* = 5.3 Hz, 1H). ^{13}C NMR (150 MHz, $CDCl_3$): δ_C 13.9, 20.5, 21.5, 22.4, 27.6, 28.1, 31.3, 32.2, 35.2, 35.8, 39.3, 39.5, 71.0, 77.5, 128.3, 146.7, 171.6, 173.7, 190.1. The spectroscopic data matched well with those reported values[98].

(*E*)-oct-3-enoyl-PANT (**7**)
^{1}H NMR (600 MHz, $CDCl_3$): δ_H 0.82 (t, *J* = 7.3 Hz, 3H), 0.85 (s, 3H), 0.92 (s, 3H), 1.22-1.32 (m, 4H), 1.98 (m, 2H), 2.93 (m, 2H), 3.17 (d, *J* = 7.0 Hz, 2H), 3.27-3.37 (m, 6H), 3.48 (m, 2H), 3.90 (d, *J* = 5.1 Hz, 1H), 4.06 (t, *J* = 5.7 Hz, 1H), 4.70 (d, *J* = 4.8 Hz, 1H), 5.42 (dt, *J* = 15.2, 7.0 Hz, 1H), 5.55 (dt, *J* = 15.2, 6.6 Hz, 1H), 6.83 (t, *J* = 5.6 Hz, 1H), 7.39 (t, *J* = 5.3 Hz, 1H). ^{13}C NMR (150 MHz, $CDCl_3$): δ_C 13.9, 20.4, 21.4, 22.1, 28.4, 31.1, 32.2, 35.2, 35.8, 39.3, 47.6, 50.4, 70.9, 77.5, 120.7, 136.8, 171.6, 173.8, 198.5.

Supplementary Figures



Lanes 1, AlnB-his, 308 kDa; 2, his-AlnB_DH, 40 kDa; 3, his-AlnB_KSATDH^{0}, 135 kDa; 4, his-AlnB_KSAT, 100 kDa; 5, his-AlnB_ERv2, 38 kDa; 6, his-AlnB_ACPv2, 14 kDa.

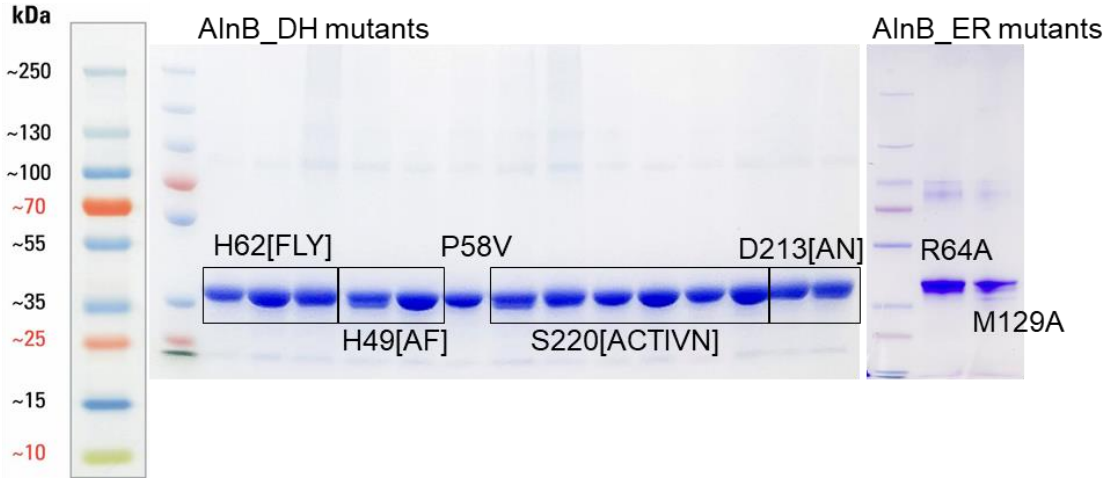


Figure S1. SDS-PAGE analysis of enzymes and mutants in this study.

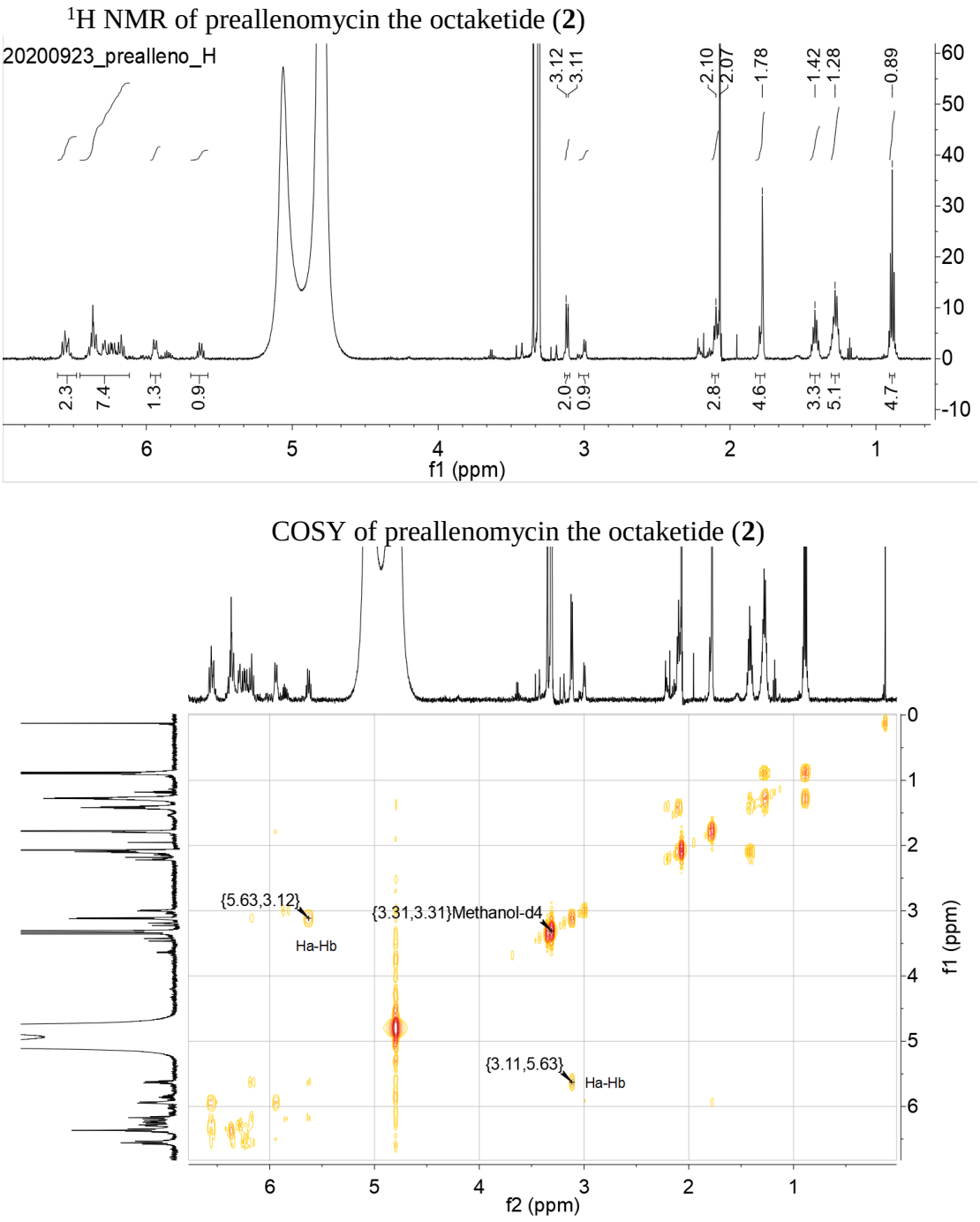


Figure S2. [1]H NMR and COSY of preallenomycin the octaketide (2).

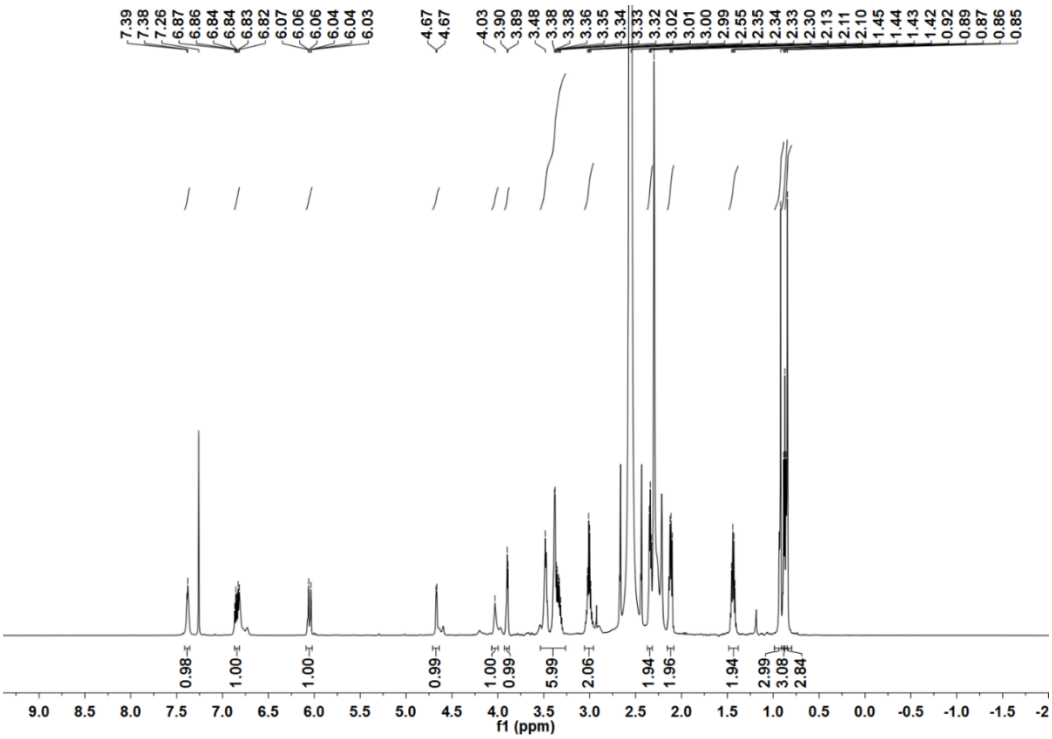


Figure S3. The ^{1}H NMR spectrum of (*E*)-hex-2-enoyl-PANT (4) in $CDCl_3$.

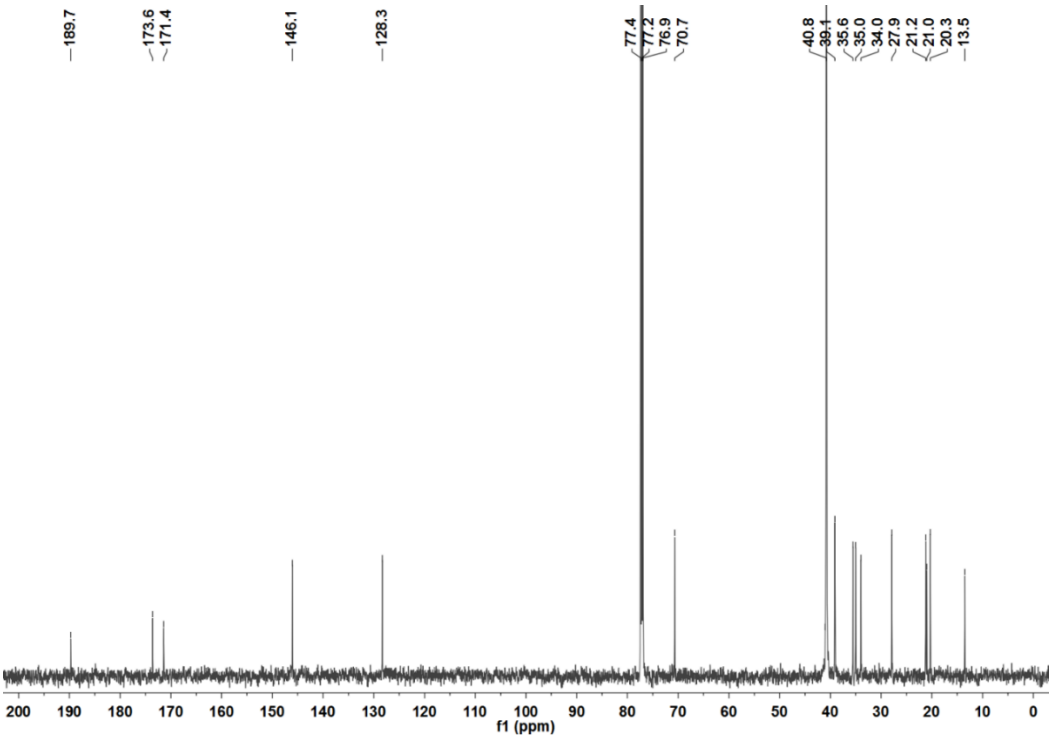


Figure S4. The ^{13}C NMR spectrum of (*E*)-hex-2-enoyl-PANT (4) in $CDCl_3$.

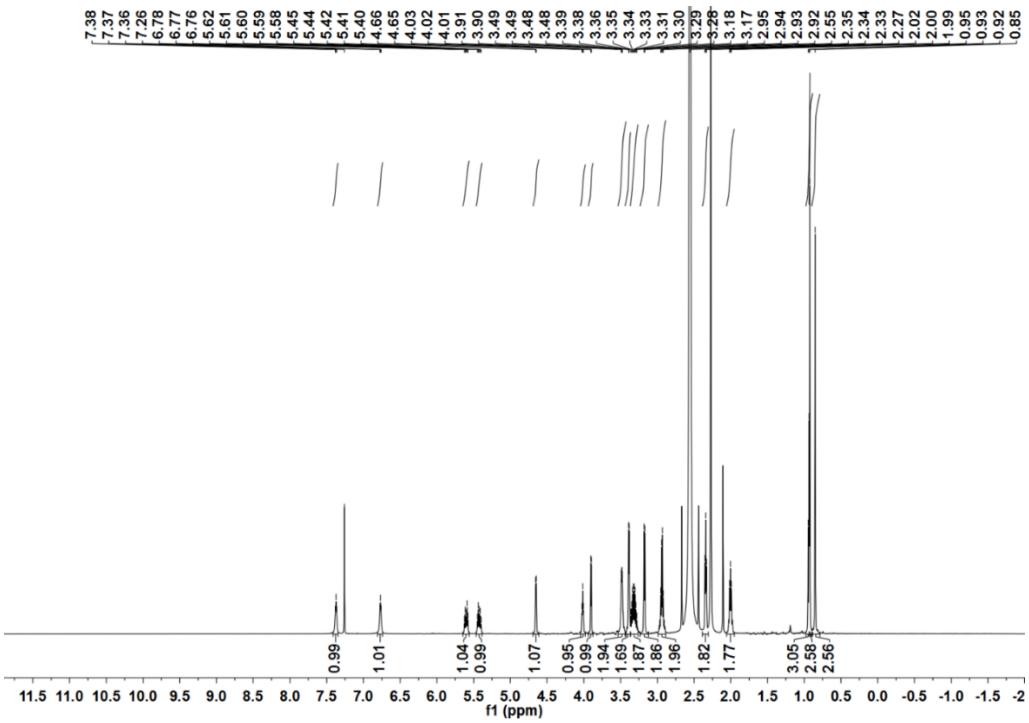


Figure S5. The ^{1}H NMR spectrum of (*E*)-hex-3-enoyl-PANT (5) in $CDCl_3$.

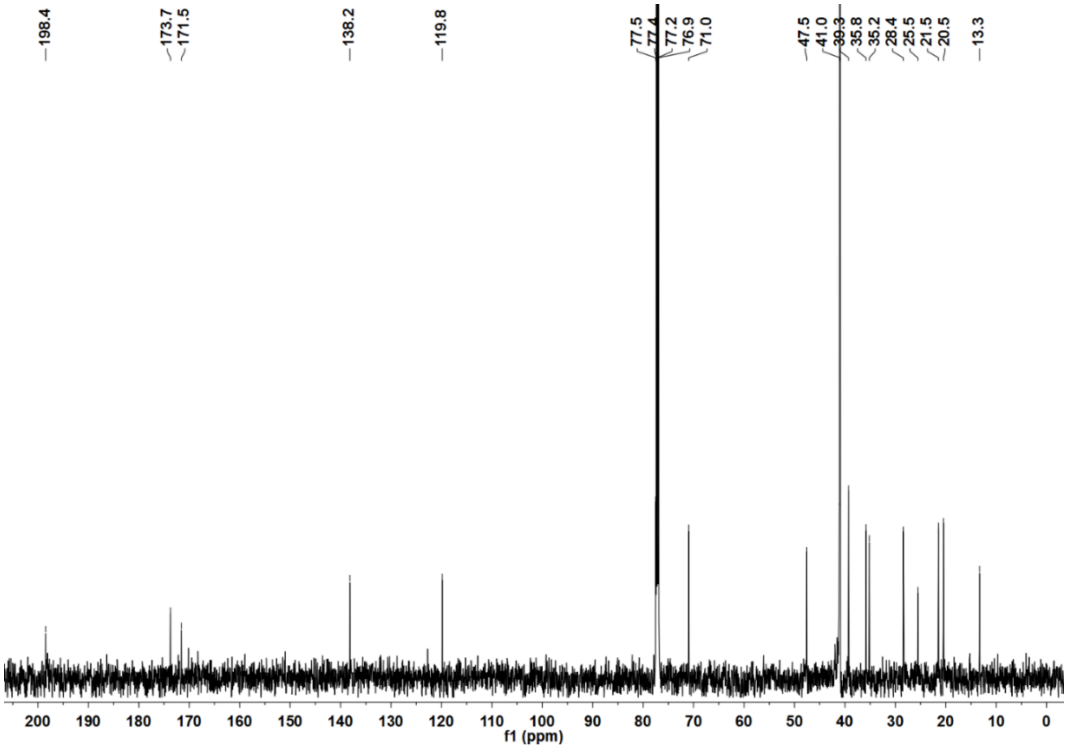


Figure S6. The ^{13}C NMR spectrum of (*E*)-hex-3-enoyl-PANT (5) in $CDCl_3$.

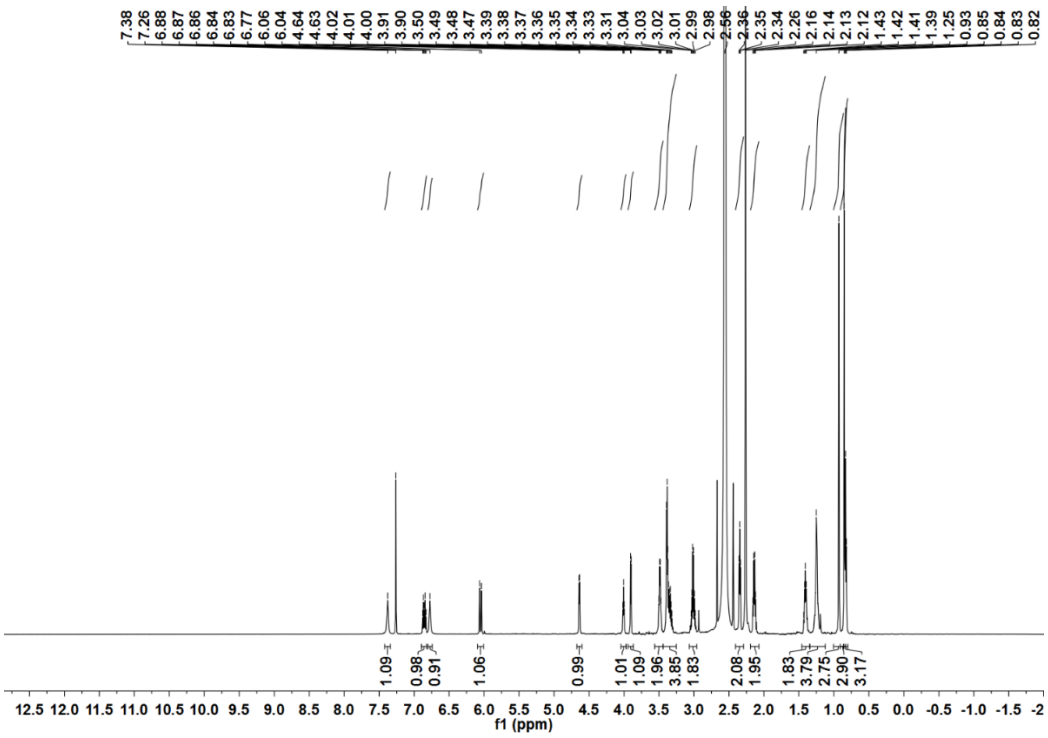


Figure S7. The ^{1}H NMR spectrum of (*E*)-oct-2-enoyl-PANT (6) in $CDCl_3$.

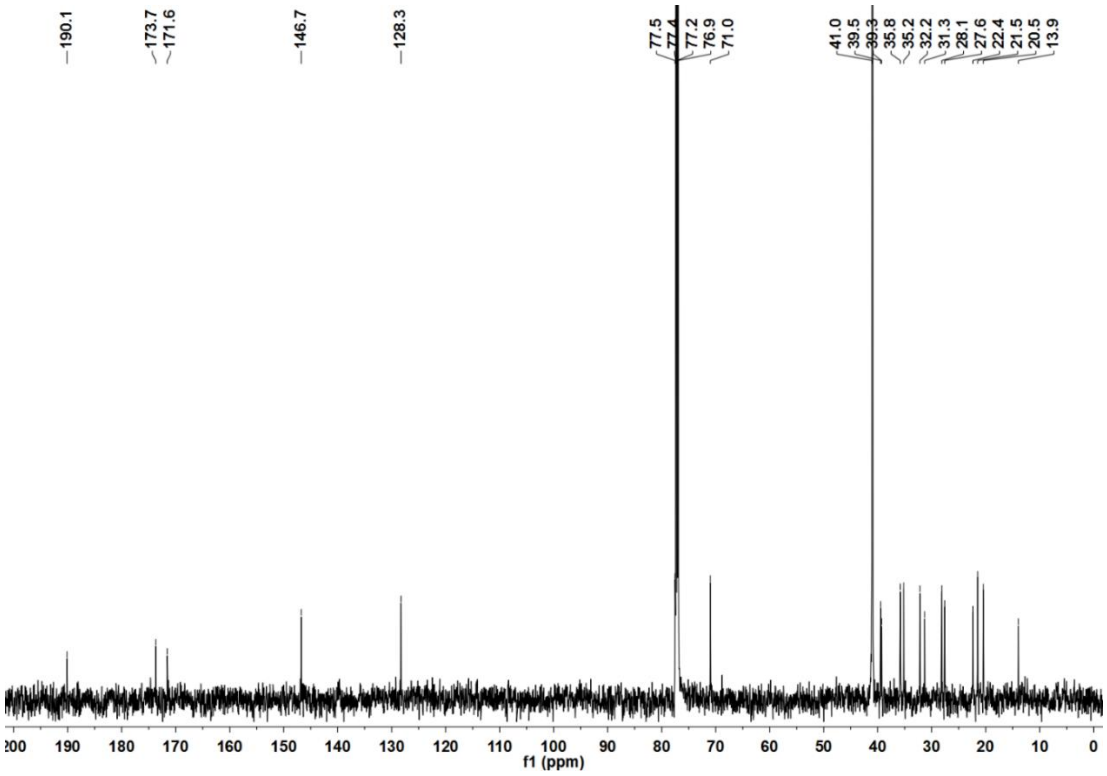


Figure S8. The ^{13}C NMR spectrum of (*E*)-oct-2-enoyl-PANT (6) in $CDCl_3$.

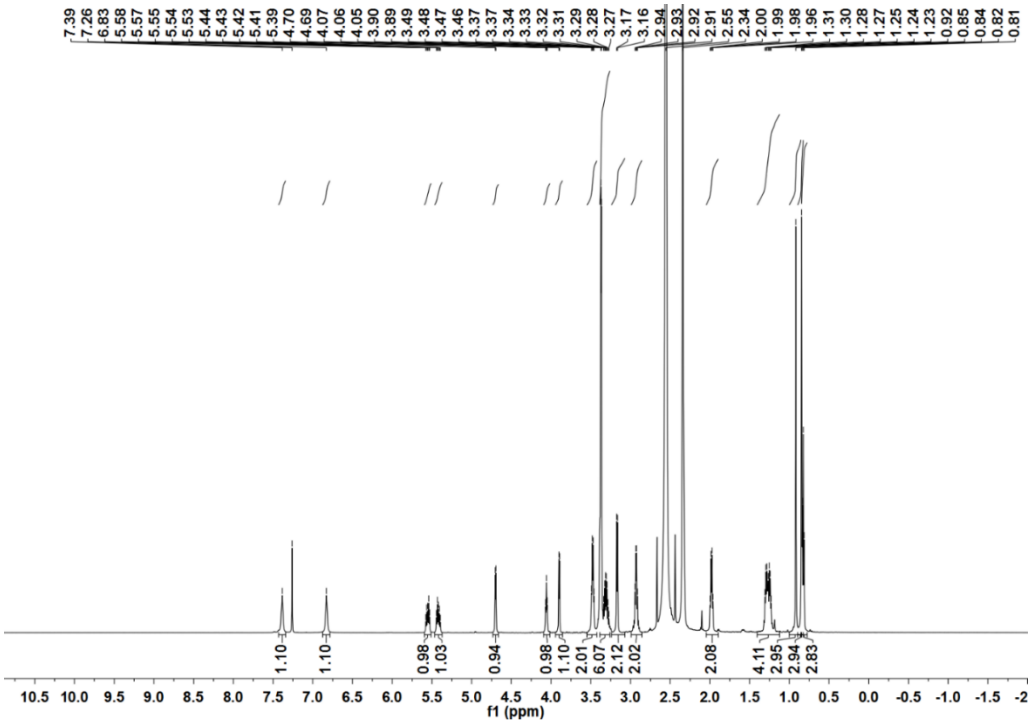


Figure S9. The ^{1}H NMR spectrum of (*E*)-oct-3-enoyl-PANT (7) in $CDCl_3$.

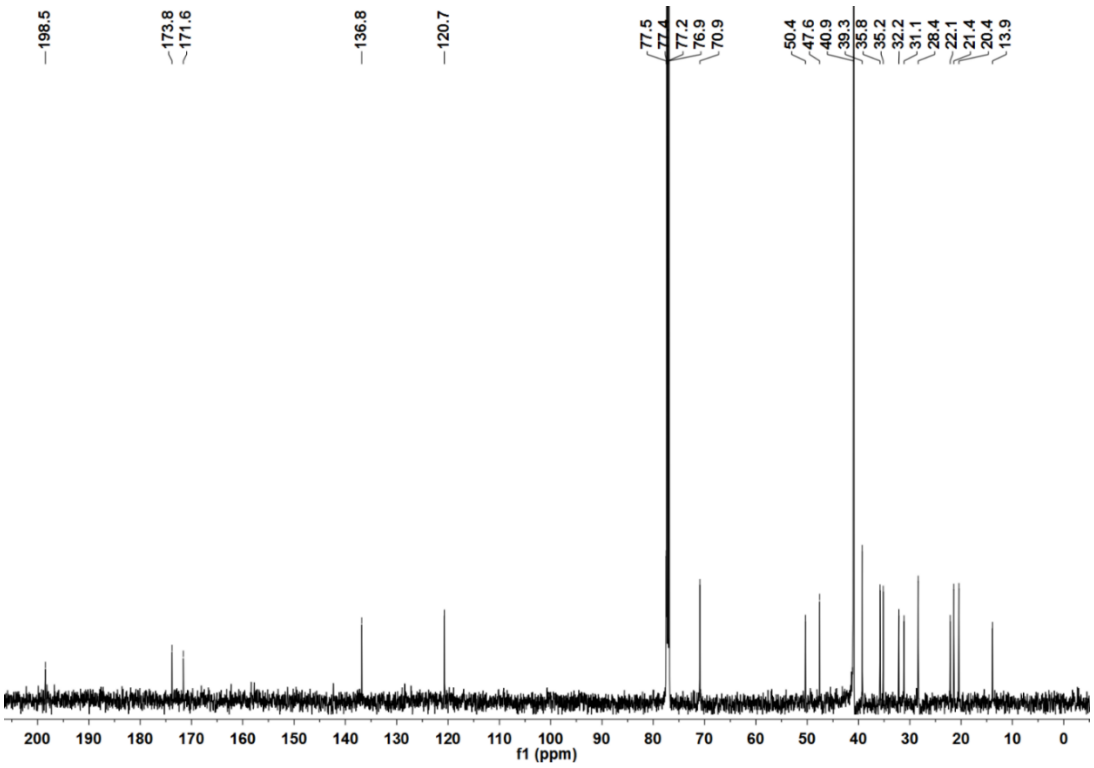


Figure S10. The ^{13}C NMR spectrum of (*E*)-oct-3-enoyl-PANT (7) in CDCl3.

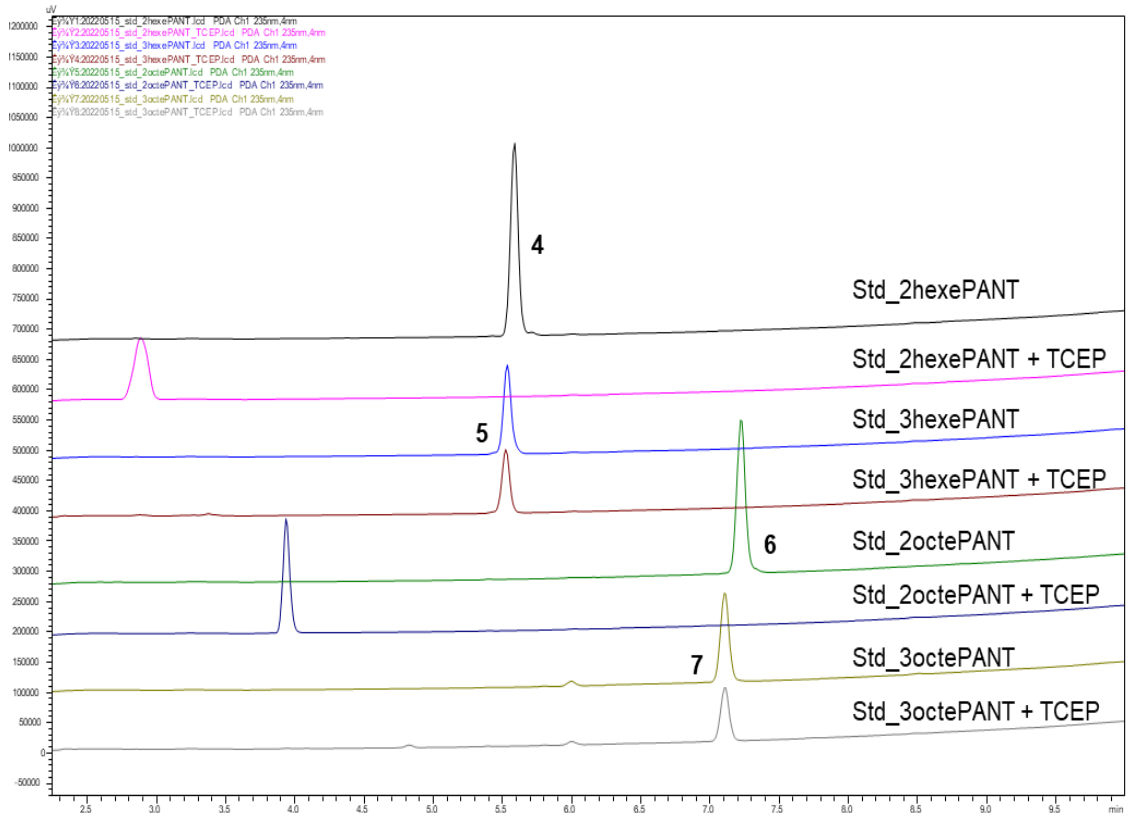


Figure S11. TCEP derivation assay. Due to the small shift in retention times for the a,b- and b,g- isomers, a TCEP assay was performed to verify the purity. 20 mM (final conc.) of TCEP was incubated with 600 □M (final conc.) of all four compounds in the reaction buffer for 1 h. This result indicates that all four compounds are very pure.

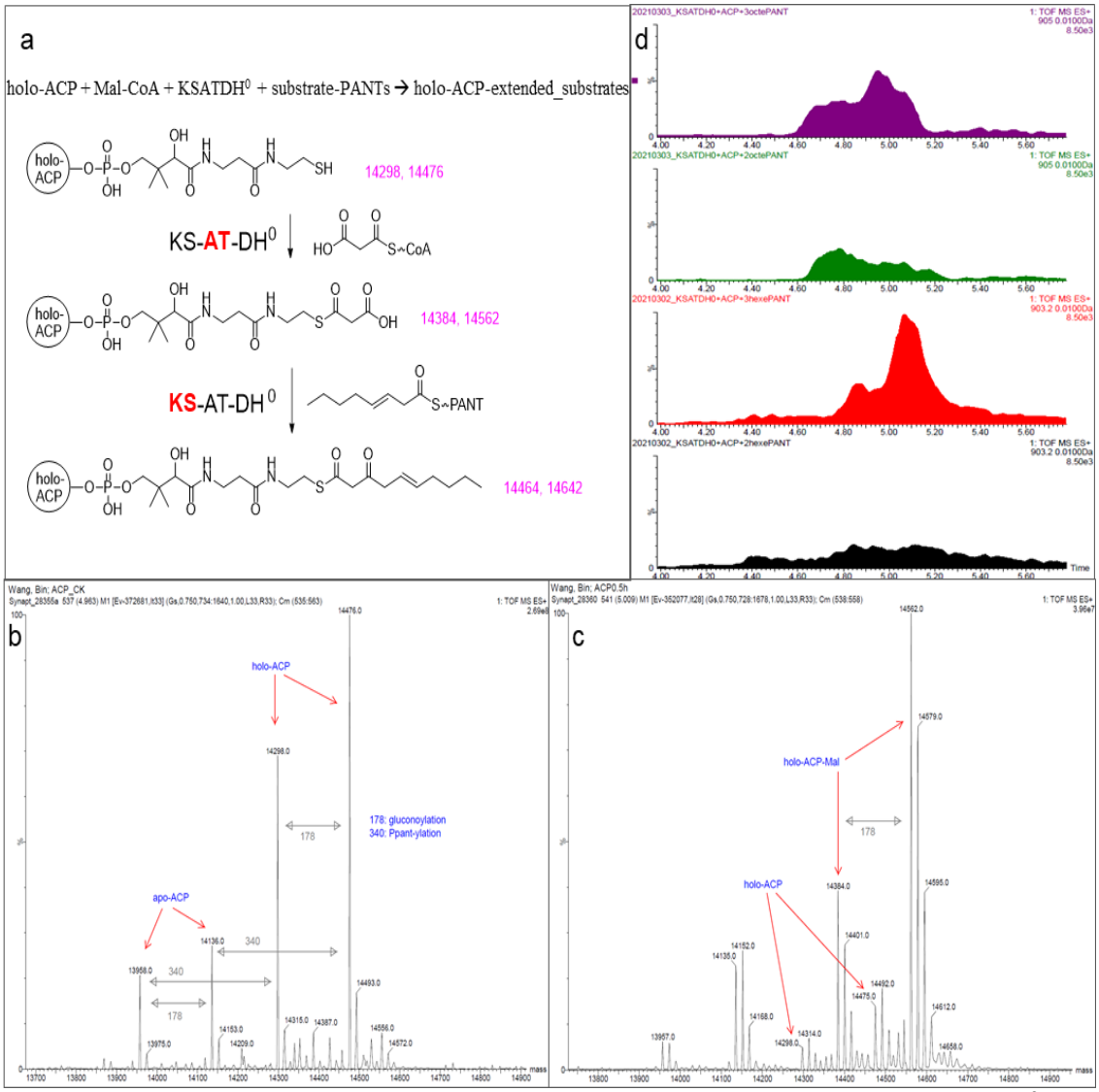


Figure S13. LC-MS analysis of reactions of AlnB_ACPv2 and AlnB_KSATDH^0. a) Reaction scheme. **b**) The holo-ACP. **c**) The mal-ACP catalyzed by AlnB_AT. **d**) Extracted ion chromatography of ACP products catalyzed by AlnB_KS.

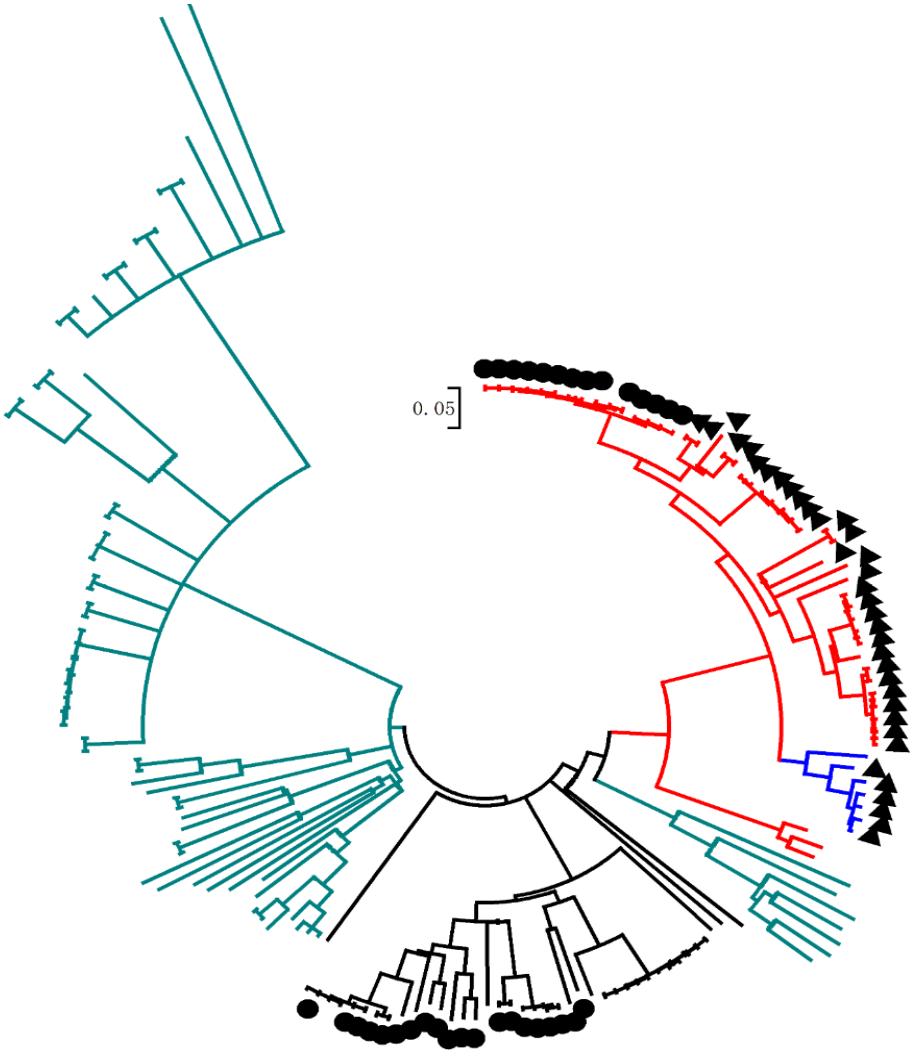


Figure S14. Phylogenetic analysis of the co-evolution of domains involved in the □□□-polyene biosynthesis and their distribution. Domains within AlnB, specifically the KS, DH, and ER that work collaboratively, are supposed to be co-evolved as well. For phylogenetic analysis, a BLASTp search was performed using AlnB_KS sequence as a query[32] against the NCBI non-redundant protein database with a cut-off identity of 50%. Meanwhile the corresponding full-length T1PKS sequences was collected for extraction of biomarkers by aligning to the specific query positions. Biomarkers are defined as key residues, H62 and S220 in AlnB_DH, and M129 in AlnB_ER, and H=[HY], S=[ST], M=[M], corresponding lower cases refer to any other residues. On the phylogenetic tree, clades in **red**, H-S-M; **blue**, H-s-M; **teal**, H-s-m; **black**, h-s-m; and ● *Streptomyces*, and ▲ *Nocardia*. Module organization was analyzed by NCBI Batch-CD search. Then the biomarker labels were projected onto the phylogenetic tree, and the result showed, 1) the BLASTp hits containing all three biomarkers are rather conserved (also conserved in the module composition, KS-AT-DH-MT-ER-KR-ACP-TE) and mainly present in *Streptomyces* and *Nocardia*, indicating a high degree of domain co-evolution in order to preserve the □□□-polyene biosynthetic capacity in the two genera of Actinobacteria—but what are the biological roles of the □□□-polyenes or the possibly

associated allenomycins? 2) A gradual gain of □□□-polyene biosynthetic capacity from possessing intrinsic □□□-DH activity to recruiting extrinsic driving forces. 3) the blue subclade buried in the red clade contains deviated DX_3T motif rather than the canonical DX_3[QH]. This motif is often replaced in EIs by **N**X_3[**VL**] which exhibits no dehydration activity. In these candidate DHs, the D and T should maintain their hydrogen-bonding as D and [QH], but the proton-transfer-network is surely not plausible. Since they are newly discovered by this strategy, their function is completely unknown. Whether they are DH or DH^0 remains to be verified. Perhaps, they are some newly evolved DH with special intrinsic programming.

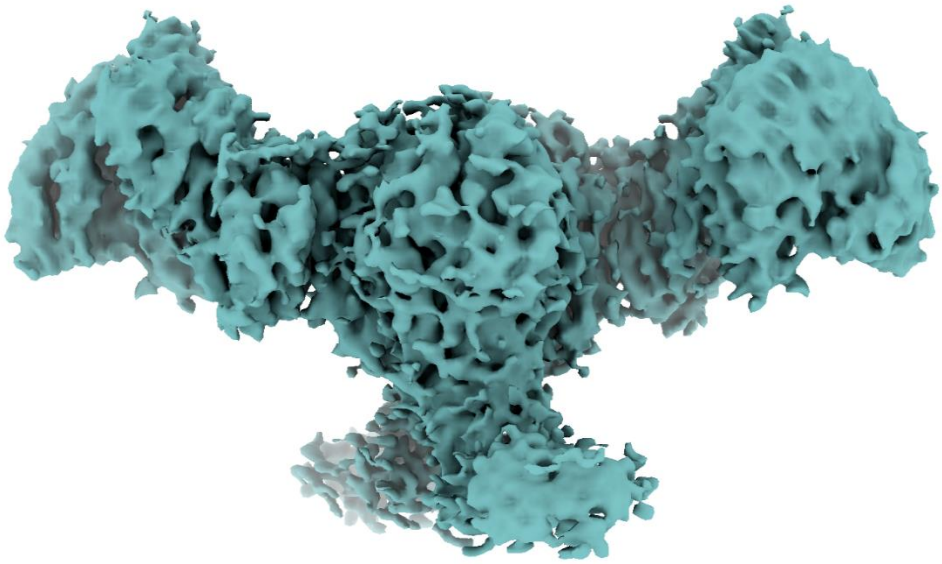


Figure S15: AlnB WT 3.8Å structure. This structure seems to be mostly the KS-AT domain and it is theorized that the remaining bulk of the wild-type AlnB is too disorganized to be correctly modeled by cryo-EM

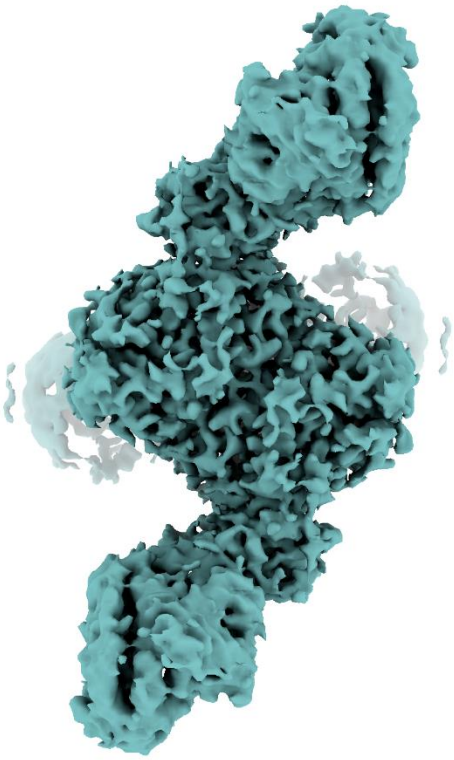


Figure S16: AlnB WT structure from “top-down” view with a focus on the KS-AT “bat-wing” domains

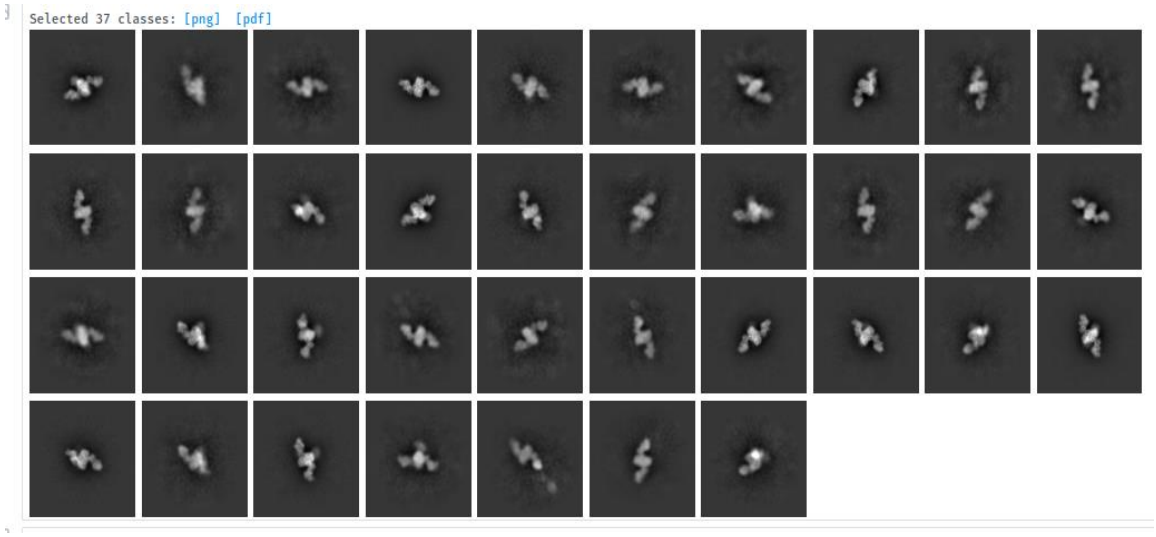


Figure S17: AlnB WT 2D classes from CryoSPARC data processing, as demonstration of the different conformations of AlnB during the imaging process

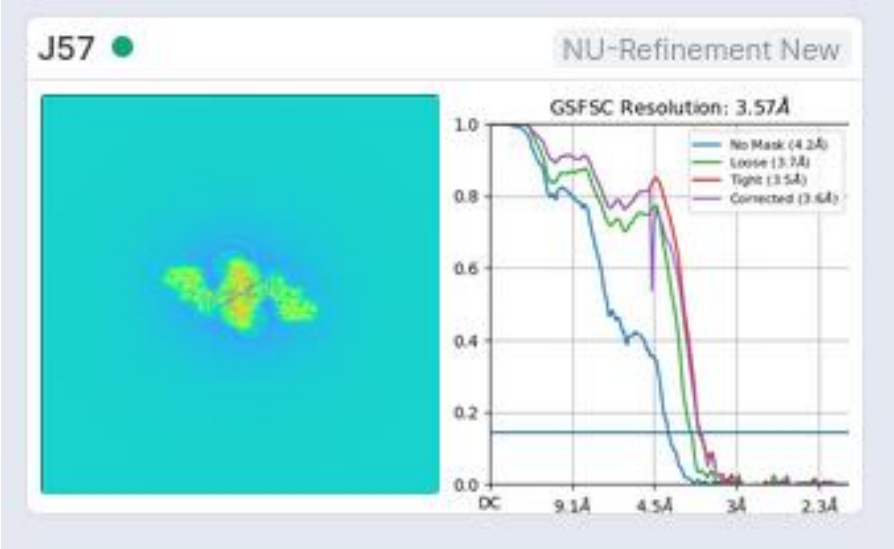


Figure S18: Final AlnB structural example in CryoSPARC

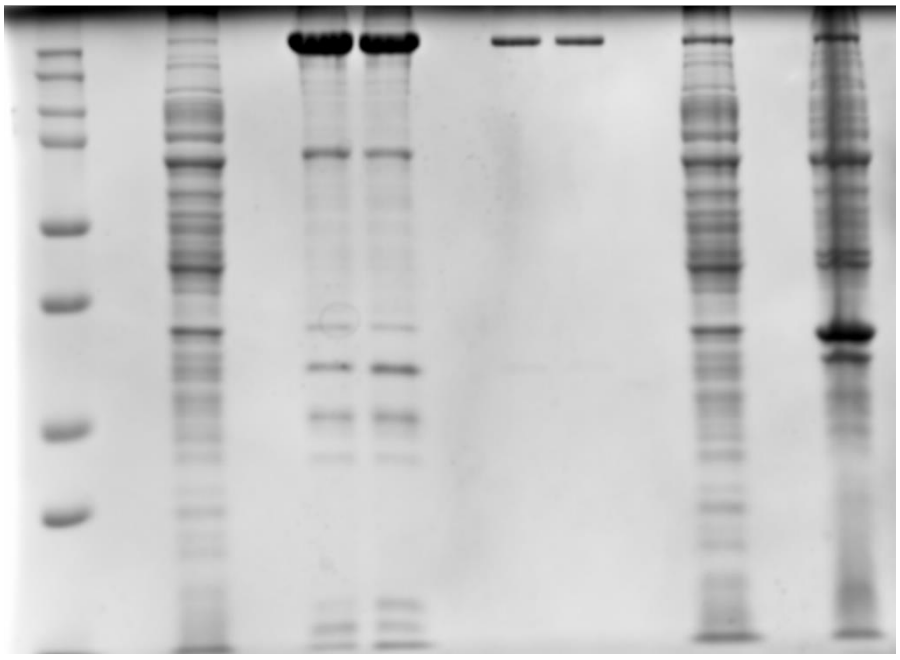


Figure S19: SDS-PAGE gel of AlnB WT purified by Ni-NTA and SEC. Lane 1 is a molecular mass standard, lane 2 is the flow through from an Ni-NTA column, lanes 3/4 are elutions from an Ni-NTA column, lanes 5/6 are final purifications from SEC. Lane 7 is a soluble fraction post sonication but pre purification, and lane 8 is an insoluble fraction post sonication but pre purification.

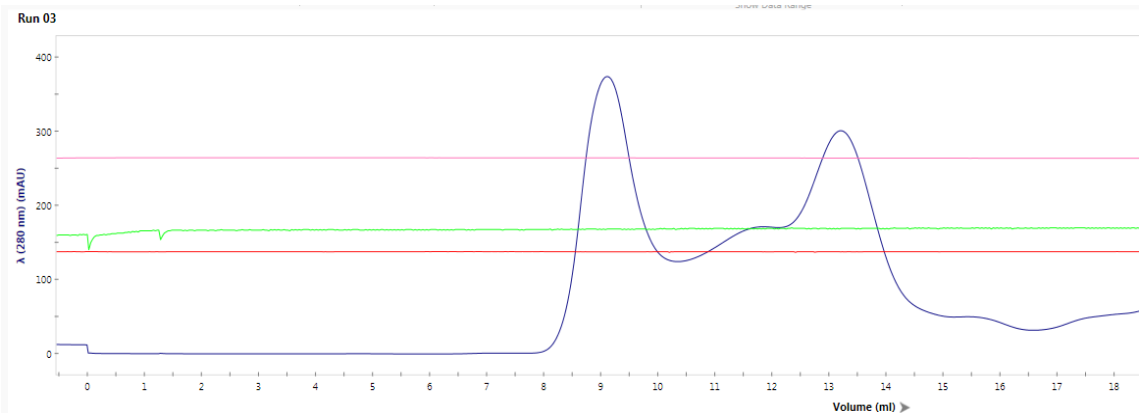


Figure S20: SEC trace of AlnB WT. The first major peak is aggregates and the void volume, while the second major peak is the full length AlnB WT. This is the purified fraction demonstrated in Figure S19.

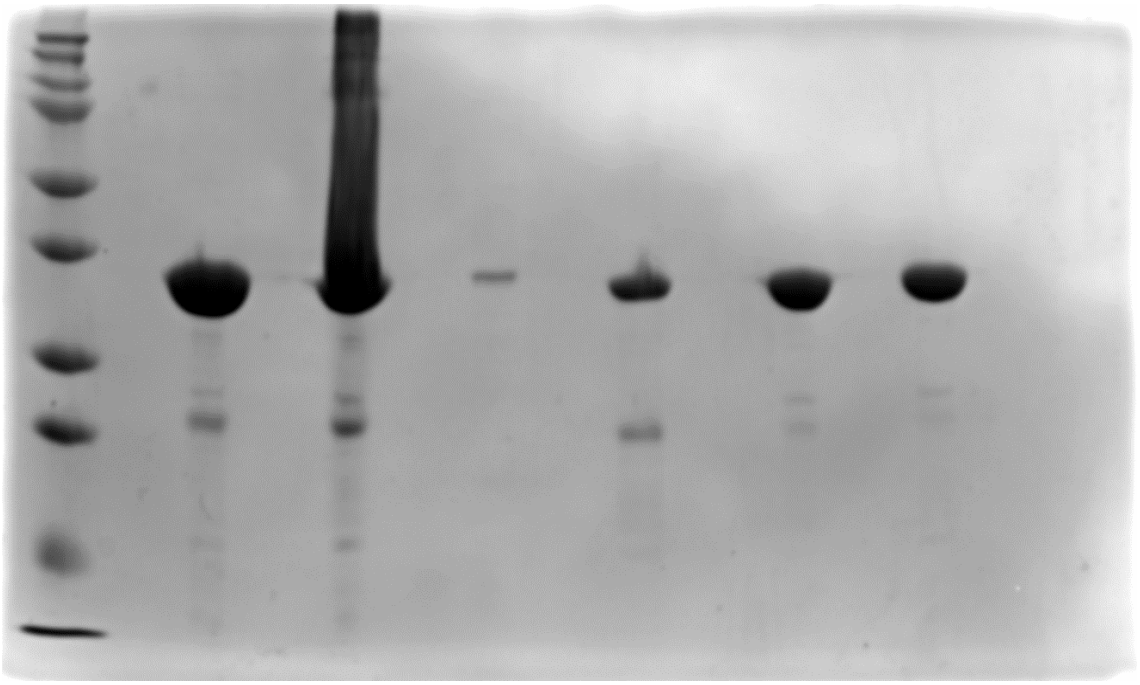


Figure S21: SDS-PAGE gel of AlnB_DH. Lane one is a molecular mass standard. Lanes 2/3 are elutions from an Ni-NTA column. Lane 4 is the void volume from an SEC run shown in figure S22. Lanes 5 is an unknown peptide with some AlnB_DH contamination, lanes 6/7 are elutions from the SEC run and contain purified AlnB_DH.

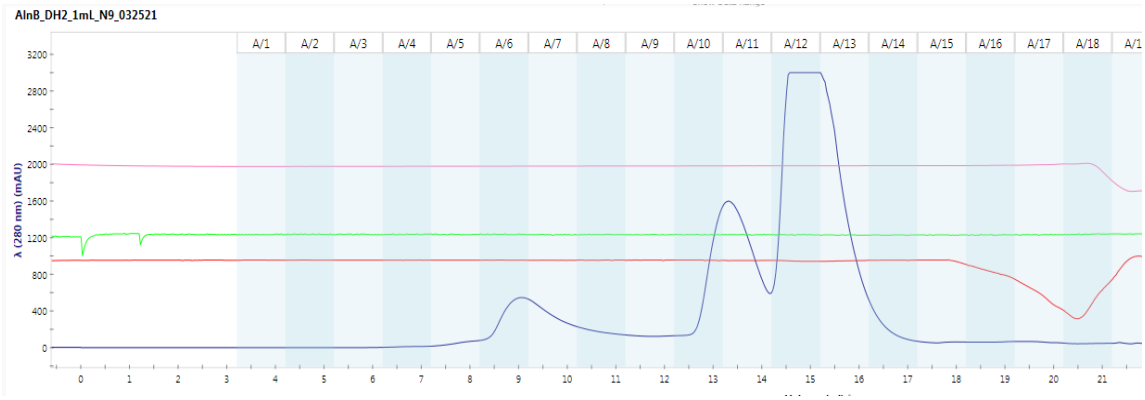


Figure S22: AlnB_DH SEC purification run. The first peak is aggregates and the void volume, the second peak is a larger, unknown peptide, and the third peak is AlnB_DH.

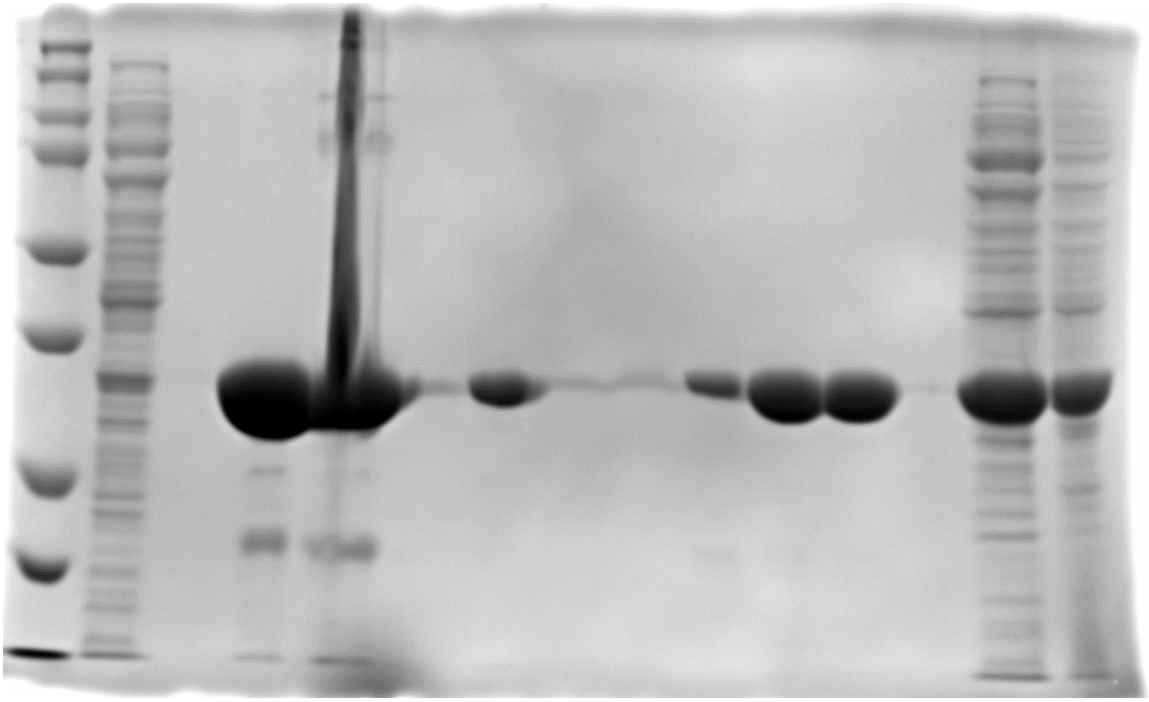


Figure S23: AlnB_DH Crys 3 SDS-PAGE. This construct was used to crystallize AlnB_DH in order to solve the structure. Lane 1 is molecular mass standards. Lane 2 is a flow-through post Ni-NTA to prove binding. Lanes 3/4 are bed volume elutions from the Ni-NTA column. Lane 5 is the void volume of the SEC run in Figure S24. Lanes 5-7 are elutions from the SEC run containing purified AlnB_DH Crys 3. Lane 8 is the soluble fraction post sonication but pre purification, and lane 9 is the insoluble fraction.

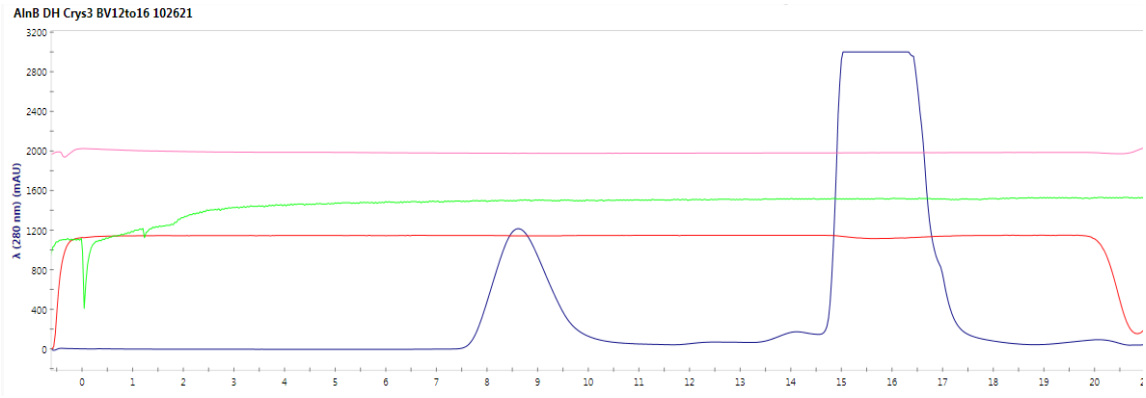


Figure S24: AlnB_DH Crys 3 SEC run. The first peak is aggregates and the void volume, while the second peak is AlnB_DH Crys 3.

Supplementary Tables

Table S1. LC-HRMS analysis of □□□-polyene compounds.

Compounds	Theoretical $[M+H]^+$	Measured $[M+H]^+$	□ (ppm)
heptaketide (**1**)	235.1693	235.1695	-0.9
octaketide (**2**)	261.1849	261.1849	0
nonaketide (**3**)	287.2006	287.2007	-0.3
octaketide (**2'**)	247.1693	247.1693	0
nonaketide (**3'**)	273.1849	273.1847	0.7
hex-2-enoyl-PANT (**4**)	375.1948	375.1952	-1.1
hex-3-enoyl-PANT (**5**)	375.1948	375.1947	0.3
oct-2-enoyl-PANT (**6**)	403.2261	403.2259	0.5
oct-3-enoyl-PANT (**7**)	403.2261	403.2260	0.2
3-hydroxyhexanoyl-PANT (**8**)	393.2054	393.2052	0.5
3-hydroxyoctanoyl-PANT (**9**)	421.2367	421.2368	-0.2
hexanoyl-PANT (**10**)	377.2105	377.2104	0.3
octanoyl-PANT (**11**)	405.2418	405.2420	-0.5