Sarah Lin
Haw Yang Lab
Lemma Pod

SRC 2019 Project Report

Abstract

Nonribosomal peptide synthetases (NRPSs) are multi-domain enzymes that incorporate various substrates to produce a diverse range of clinically valuable peptide products, such as antibiotics, antitumors, and antivirals. Despite their widespread utility, it has been difficult to elucidate the exact mechanisms of peptide bond formation involved in NRPS natural peptide synthesis because of the NRPS's modular nature and high levels of protein interactions. To better understand the condensation reaction, the research conducted uses single-molecule fluorescence spectroscopy as another angle with which to study the conformational changes of two NRPSs as they undergo the condensation reaction. The research conducted establishes the construct of the proteins and tests the enzymatic assay that characterizes the function of these proteins. In using the single-molecule fluorescence method to study NRPS, we can learn more about a fundamental biological process and later apply it to clinical products.

Project Context and Rationale

The most well-known peptide production pathway is through ribosomal synthesis. However, scientists are now investigating alternate pathways of peptide formation found in nature.[1] One of these alternate biosynthesis pathways is through the nonribosomal peptide synthetase (NRPS), a large enzyme (on the order of hundreds of kDa) mainly found in bacteria and fungi.[1] NRPS genes that encode a specific peptide are typically found in the same operon (prokaryote) or core of gene clusters (eukaryote).[1] Each NRPS synthesizes a unique secondary metabolite product.[2] By using this

nonribosomal pathway, microorganisms can synthesize structurally diverse peptide products with varying complexities.[2] Unlike ribosomes, these nonribosomal peptide synthetases can incorporate nonproteogenic amino acids, hydroxy acids, as well as the usual amino acids into complex secondary metabolite products, greatly contributing to the diversity of the peptide products.[3, 4] These nonribosomal peptide (NRP) products have numerous applications in environmental and health fields. In medicine, NRP core structures are commercially used as systemic and topical antibacterials, antitumor drugs, antifungals, animal feed additives, and immunosuppressants.[1] Environmentally, NRP structures are used for insecticides and biopesticides, plant-strengthening, and biocontrol agents to enhance growth and crop yields.[1] Understanding the mechanisms of NRPS would allow for pathway engineering of novel biological peptide products.[1]

Nonribosomal peptide synthetases can be thought of as collinear assembly-lines in which each module is a distinct protein domain with the catalytic capabilities for peptide synthesis.[1] Each NRPS contains a minimum of four domains.[3] First is the adenylation (A) domain for selection, activation, and loading of the amino acid onto T domain. The A domain is where the amino acid is activated through the formation of a reactive aminoacyl adenylate.[3] Second is the thiolation (T) domain, also known as the peptidyl carrier protein (PCP) domain.[3] The T domain bears a 4'-phosphopantetheine (Ppant) prosthetic group in its holoform that tethers the amino acid and transfers the covalently attached aminoacyl motif along the assembly line.[3] The next domain of the assembly line is the condensation (C) domain, where a peptide bond is formed with the upstream peptide chain. Since this is where the peptide bond is formed,[3] the condensation domain is catalytically necessary for NRP formation.[4] When the condensation domain of ProCAT (TycB) was deleted from the assay, no dipeptide was formed.[4] The last essential domain in the NRPS assembly-line is the thioesterase (TE) domain, where hydrolysis terminates the chain and the mature

oligopeptide is released from the NRPS machinery.[1] When attached to any of these domains, building blocks can be sent to optional protein domains such as epimerization (E), methylation (M), formylation (F), and many others.[1] For example, previous kinetics studies have shown that the epimerization domain has a major role in the selectivity of substrate.[2] The inclusion of these extra domains varies greatly and depends on the specific NRPS system.[1]

Though the basic functions of the domains are known, there is still much to be understood in terms of the kinetics and dynamics of NRPS mechanisms. Specifically, the exact mechanisms of the condensation reaction that creates the peptide bond are still unknown.[4] Structural biology studies have recently elucidated the general domain architecture of an NRPS module.[1] Research on NRPS has been based on 4 methods: chemoenzymatic (CHE), precursor-directed biosynthesis (PDB), mutasynthesis (MBS), and combinatorial biosynthesis (CBS).[1] Furthermore, the Yang Lab has recently established that it is possible to use multi-platform kinetics profiling to quantify the relative weights of on- and off-pathway reactions.[3] These kinetics assays include a mass spectroscopy kinetics assay (MS), a high-performance liquid chromatography (HPLC) assay, and a ATP-pyrophosphate (PPi) exchange assay.[3] Combining these kinetics assays with established enzyme-coupled byproduct-release assays has allowed for more detailed investigations of substrate-NRPS enzyme interactions.[3]

To study the kinetics and dynamics of the condensation reaction, there needs to be an efficient system to study the peptide bond formation without going through the whole reaction. Therefore, previous research has established Gramicidin S Synthetase as a model.[2] Gramicidin S is a well-established decapeptide with antibacterial activity against both gram-positive and gram-negative bacteria.[5] Iterative NRPSs are responsible for the biosynthesis of many antibiotics, such as Gramicidin S.[5] Using Gramicidin S Synthetase, an *in vitro* assay for studying the condensation

reaction has been established.[4] The assay simplifies the NRPS pathway to the single condensation event that occurs when the reaction between GrsA and TycB is initiated.[4] The gramicidin S synthetase A (GrsA) has an initiation module called PheATE, a three-domain subunit that recognizes the substrate L-phenylalanine (L-Phe).[2] In the assay, GrsA (initiation domain PheATE) and TycB (initiation domain ProCAT) are both recombinantly modified enzymes that have been converted to holomodules.[2] Each of these enzymes is attached to its substrate, D-Phe and L-Pro respectively.[2] The two modules are then combined and a fast reaction occurs in which the GrsA activates L-Phe and then transfers the molecule to GrsB.[3] Condensation occurs and the expected dipeptide product of D-Phe-L-Pro-S-enzyme is produced.[2] The paper also shows that GrsB can play the same role in accepting D-Phe from GrsA to form the D-Phe L-Pro product[2], which is why the project utilizes GrsB.

On a macro scale, the research I am conducting with the help of my mentor uses single-molecule fluorescence spectroscopy as another angle with which to study NRPS. The basic steps are protein purification of GrsA and GrsB, labelling the enzymes with fluorescent tags, and then viewing protein conformational changes with fluorescent microscopy when the two enzymes are placed into the previously described assay. One of the main challenges of this research is the purification and maintenance of the enzymes. Once they are successfully purified, the proteins need to be labeled with a fluorescent tag while maintaining their functionality and conformation. Finally, the proteins will be immobilized using a novel system of biotin-streptavidin binding. Based on my mentor's previous research, it is possible to apply single-molecule fluorescence spectroscopy in combination with computational modeling to capture images of proteinogenic biochemical reactions. Therefore, varying components of the GrsA-GrsB complex will be immobilized and the reaction will

be visualized with fluorescence microscopy with the goal of elucidating the dynamics of the condensation reaction that creates the peptide bonds in NRPS systems.

There is much to gain from the understanding of the NRPS system. To understand the biosynthesis pathway of NRPS systems is to be able to reprogram and redesign assembly lines and ultimately produce any peptide structure.[1] Through pathway engineering, scientists could feasibly use NRPS as an alternative to the standard synthesis of peptides.[2] Not only would this provide economic advantages, this would also create a new avenue to develop sustainable and environmentally-conscious biotechnology products.[1] Some of these new bioactive NRPs could include new anticancer drugs, stronger antibiotics, and agricultural products. Ultimately, elucidating the condensation reaction of NRPSs would lead to a deeper understanding of biosynthetic mechanisms.

Methodology

The methodology of this project can be broken three components: protein expression and purification, protein characterization, and probing and visualization.

Protein expression and purification. Expression and purification of GrsA and GrsB proteins were performed using established protocols.[6] *E. coli* cells were transformed with expression plasmids containing the gene encoding GrsA A-PCP (or GrsB A-PCP) as well as a C-terminal His_6-Tag. The GrsA and GrsB proteins were expressed in these cells at varying temperatures and varying times to test which conditions would produce the purest proteins. After being stored in -80 °C, the cells were lysed using a mechanical homogenizer, the French Press (EmulsiFlex-C3 homogenizer), to isolate the proteins from the rest of the cell. The proteins of interest were first isolated with multiple centrifugation washes. Then the proteins were run through a number of

columns in the FPLC (fast protein liquid chromatography) machine. A Cobalt-NTA column was used first to isolate the proteins with the His-tag. Following that column, a Mono Q anion exchange chromatography column was used to further purify the Grs proteins based on charge. Experiments were conducted with NRPS assay buffer and 20 mM TRIS buffer.

Protein characterization. After each step of protein purification, the NanoDrop Microvolume Spectrophotometer was used to quantify and assess the concentration and purity of the proteins. An extinction coefficient of 73,000 M^{-1} cm^{-1} at 280 nm wavelength was used to estimate the protein concentration. After each column, fractions containing significant amounts of protein were collected and run on a SDS-PAGE gel. These gels were visualized to confirm if the protein was fully purified.

Probing and visualization. [In progress] Once the proteins are successfully purified using the optimal combination of columns and expression conditions, the proteins will be labelled with fluorescent tags. Further testing will be done to ensure that structure and function have not been lost or changed after tagging. Finally, the condensation reaction will be initiated using the biotin-streptavidin system and under fluorescence spectroscopy, the reaction will be visualized by the fluorescent tags. Any protein conformation changes that occur during the condensation reaction will therefore be visualized and this data can be used to expand knowledge of how the NRPS forms peptide bonds.

Preliminary Results

Much of the work this summer has been trial-and-error for the first section of the project – protein purification and characterization. Gramicidin S Synthetase has been more difficult than anticipated to purify. Purification was measured through two metrics: when the fractions were run on

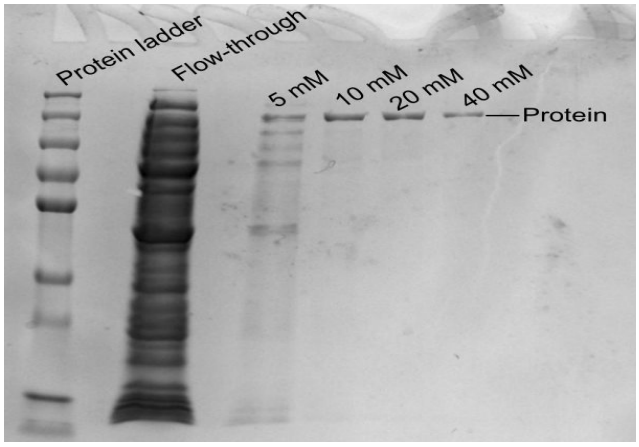
an SDS-PAGE gel, was a single protein band without impurity bands obtainable? when the fractions were combined and a mass spectrometry experiment was performed, did the experimental mass match the expected mass? To this end, a number of columns in the FPLC were trialed. The first separation technique was Size Exclusion Chromatography, but this did not successfully purify GrsA, which was being tested first because it is more stable than GrsB. The second separation technique tested was ion exchange chromatography with the Q-Sepharose Column, which did not successfully purify GrsA. The third separation technique tested was also ion exchange chromatography, but with the Mono-Q column, which is better suited for separation of high molecular mass proteins. This technique did not successfully purify the proteins. The final technique used was his-tag purification with a Cobalt-NTA Column. This technique was not successful at first, but when protein growth conditions were varied, the Cobalt-NTA Column was successful in purifying GrsA, as shown by the single protein band in the SDS-PAGE Gel (Figure 1). Once this was confirmed by the mass spectrometry experiment (Figure 2), we moved on to finding the optimal growth conditions. The optimal conditions for protein purification of GrsA were found to be 28ºC for 4 hours as well as 16ºC for 8 hours. With either of these conditions and the Cobalt-NTA Column, single-band protein purification was obtained. Next, we moved onto purification of GrsB. This is still a work in progress, as GrsB is more difficult to work with. The same expression conditions combined with the Cobalt-NTA Column did not purify GrsB, so experimentation continues for purification of this protein. At the same time, we are testing purification of GrsA with cysteines removed and GrsA with just one cysteine. In the future, fluorescent tags will be placed onto these cysteines, so it is essential to be able to purify modified GrsA with just one cysteine on each domain so that the conformational changes of each domain can be tracked. This is also proving to be very difficult; the Cobalt-NTA

Column not only did not purify the modified GrsA, but there was very little yield of the protein in the first place.

Aside from protein purification, experimentation has been conducted to make sure that the system of GrsA-GrsB works as expected. Using the HPLC, the two enzymes coupled with their respective amino acids were put together in solution and the expected dipeptide (DKP) was produced (Figure 3). The identity of the dipeptide was confirmed by a mass spectrometry experiment (Figure 4).

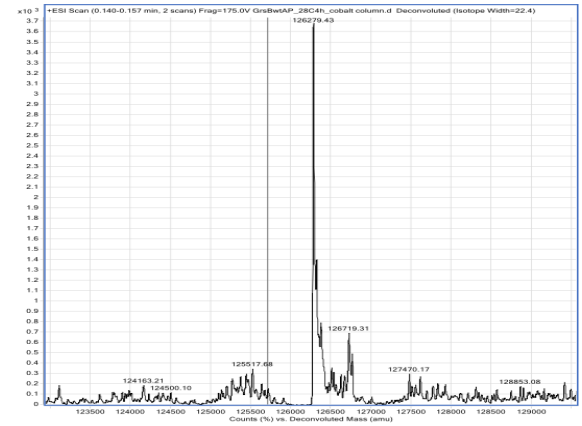
Figures

Figure 1: SDS-PAGE Gel of GrsA



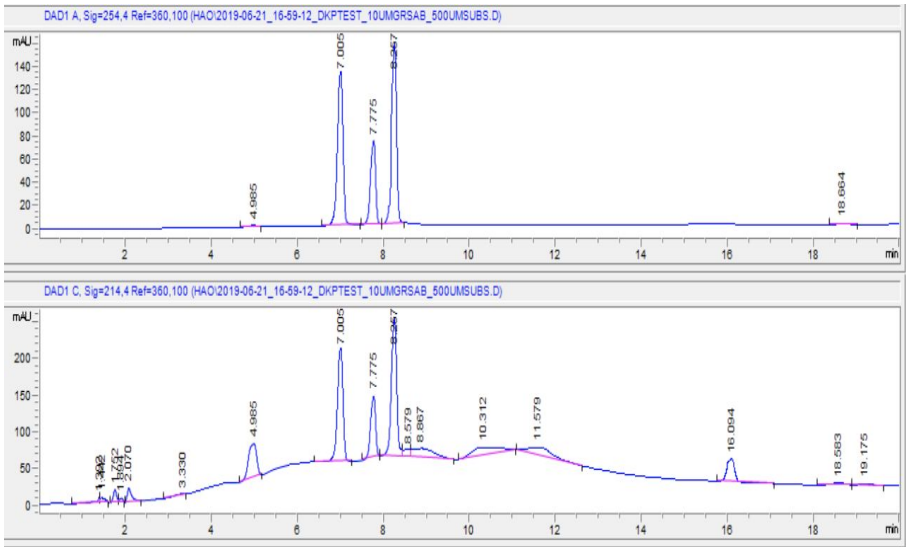
GrsA was grown at 28ºC for 4 hours and separated with the Cobalt-NTA Column. Single-band protein purity was obtained with imidazole step-elution.

Figure 2: Mass Spectrometry of Purified GrsA



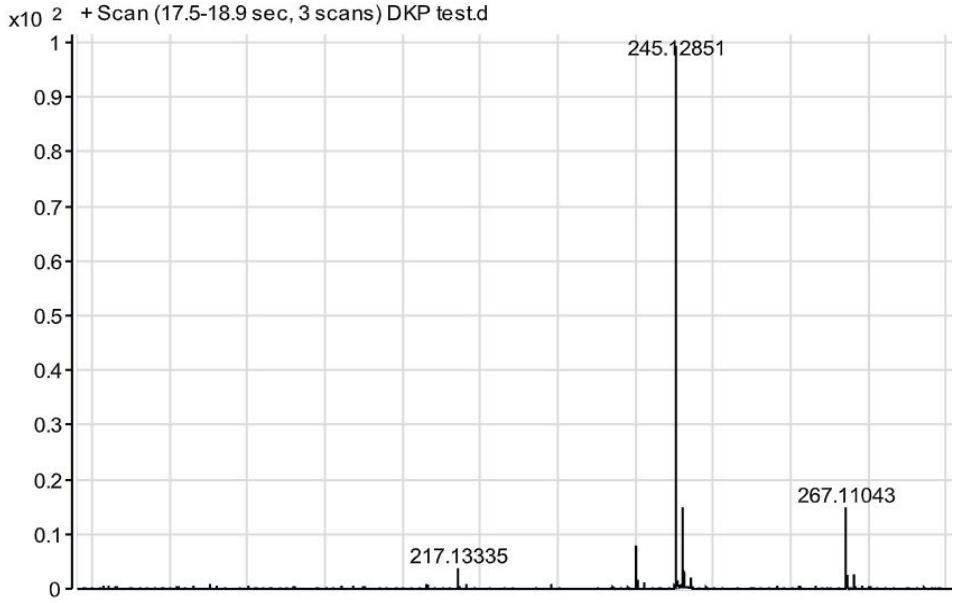
Confirmed that the mass of the purified GrsA matched the expected mass of GrsA

Figure 3: HPLC Profile of Dipeptide Product



The expected dipeptide product was formed from the reaction between GrsA and GrsB.

Figure 4: Mass Spectrometry of DKP Product



Confirmed that the product produced by the reaction was the expected DKP from literature.

Conclusions and Future Directions

There is still much to be done with this project. What can be concluded for now is that GrsA (native) can be successfully purified when expressed at 28ºC for 4 hours or 16ºC for 8 hours and then separated by a Cobalt-NTA His-Tag Purification Column. It is also confirmed that GrsA and GrsB interact to create the dipeptide DKP when GrsA transfers its amino acid, phenylalanine, to GrsB's amino acid, proline. In essence, the protein constructs have been established and the main enzymatic assay has been tested under controlled conditions. Further studies will be conducted to test what steps must be taken to purify GrsB, GrsA with no cysteines, GrsA with one cystine, and GrsB with the same variations. Then, experiments will be conducted to attach fluorescent tags to the cysteines on each of these enzymes and then confirm that the functionality does not change when the tags are added. Finally, using fluorescent microscopy, the GrsA-GrsB reaction will be put into place and the condensation reaction will be visualized through the movement of the fluorescent tags.

References

(1) Süssmuth, R. D.; Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angewandte Chemie International Edition* **2017**, *56*(14), 3770–3821.

(2) Luo, L.; Burkart, M. D.; Stachelhaus, T.; Walsh, C. T. Substrate Recognition and Selection by the Initiation Module PheATE of Gramicidin S Synthetase. *Journal of the American Chemical Society* **2001**, *123*(45), 11208–11218.

(3) Sun, X.; Li, H.; Alfermann, J.; Mootz, H. D.; Yang, H. Kinetics Profiling of Gramicidin S Synthetase A, a Member of Nonribosomal Peptide Synthetases. *Biochemistry* **2014**, *53*(50), 7983–7989.

(4) Stachelhaus, T.; Mootz, H. D.; Bergendahl, V.; Marahiel, M. A. Peptide Bond Formation in Nonribosomal Peptide Biosynthesis. *Journal of Biological Chemistry* **1998**, *273*(35), 22773–22781.

(5) Montavon, T. J.; Bruner, S. D. In *Comprehensive Natural Products II*; Elsevier Science; Vol. 5, pp 619–655.

(6) Zettler, J.; Mootz, H. D. Biochemical Evidence for Conformational Changes in the Cross-Talk between Adenylation and Peptidyl-Carrier Protein Domains of Nonribosomal Peptide Synthetases. *FEBS Journal* **2010**, *277*(5), 1159–1171.