Fall 11-12-2020

Pyrococcus horikoshii RadA Intein Site Selection

Noah Stallins

Murray State University Honors College

HONORS THESIS

Certificate of Approval

Pyrococcus horikoshii RadA Intein Site Selection

Noah Reed Stallins

November 2020

Approved to fulfill the
requirements of HON 437
Professor

_____
Dr. Christopher Lennon, Assistant Professor
Biology

Approved to fulfill the
Honors Thesis requirement
Of the Murray State Honors
College

_____
Dr. Warren Edminster, Executive Director
Honors College

Examination Approval Page

Author: Noah Reed Stallins

Project Title: *Pyrococcus horikoshii* RadA Intein Site Selection

Department: Biology

Date of Defense: 11/12/2020

Approval by Examining Committee:

_____ _____
(Dr. Christopher Lennon, Advisor) (Date)

_____ _____
(Dr. Gary ZeRuth, Committee Member) (Date)

_____ _____
(Dr. Terry Derting, Committee Member) (Date)

Pyrococcus horikoshii RadA Intein Site Selection

Submitted in partial fulfillment

Of the requirements

Of the Murray State University Honors Diploma

Noah Reed Stallins

November 2020

ABSTRACT

Inteins are protein segments interrupting polypeptides with the unique ability to excise from the host protein and link flanking protein fragments (exteins) to form a functional protein. Many of these proteins are involved in crucial processes such as DNA replication, DNA recombination, and DNA repair. The functionality of these elements in nature and applications in biotechnology necessitates the study of inteins. *Pyrococcus horikoshii* (*Pho*) provides an array of inteins to study, with a total of nine different proteins possessing inteins. Among the nine total inteins, the DNA repair and recombination protein RadA gained interest due to the speed and efficiency of splicing when flanked by non-native exteins. Curiously, RadA splices poorly within the native exteins in the absence of catalysts due to intein-extein interactions. Single-stranded DNA (ssDNA), which binds to RadA on the c-extein, vastly improves the speed and accuracy of splicing. ssDNA is not only a substrate of the RadA protein but also a signal indicating DNA damage and the need for recombinase activity. Given the observation that intein-extein interactions form to block splicing, we wondered how alternative intein insertion sites within the host protein would influence splicing. Observation and comparison of splicing speed and efficiency in natural and variant RadA proteins may help us further understand the phenomena of intein splicing, explain intein-extein relationships, and give insight on practical applications.

Table of Contents

Index

Introduction

Intervening proteins, or inteins, are polypeptides that interrupt host proteins and are removed through protein splicing. The first intein discovered was within *Saccharomyces cerevisiae*, a yeast species commonly used in fermenting alcoholic beverages and breadmaking (Basturea, 2020). While once believed to be biological anomalies, inteins are found in organisms from all three domains of life, as well as in viruses and phages. Most commonly found in archaea (47%) and bacteria (24%), but also present in single-celled eukaryotes (approximately 1%), inteins are widespread throughout the tree of life (Novikova et al., 2016).

Inteins are most commonly present in DNA-binding proteins involved in replication, recombination, and repair of DNA. Within the bacterial and archaeal domains, around 60% of known inteins are present in replication, recombination, and repair proteins (Novikova et al., 2016). Inteins were originally suspected to be parasitic, providing no benefit to the host protein. However, the presence of inteins in non-orthologous helicases of bacteria and archaea suggests that inteins can be selectively retained when beneficial to the host protein. Inteins are commonly found clustered within functional regions of DNA-binding proteins. Roughly 70% of inteins are located in ATP-binding domains, blocking the binding site of ATP (Novikova et al., 2016). Without ATP binding the protein lacks ATPase function. Therefore, if the intein splices, ATP can bind and the protein will function normally. The intein splicing reaction can be conditional on environmental stress or the introduction of the host protein substrate. (Lennon et al., 2017). The conditions required for intein splicing paired with the strategic clustering of inteins in the ATP-binding domains help create an effective post-translational protein regulation mechanism.

Proteins that contain these self-splicing inteins are precursor polypeptides. The host protein sequence that precedes the intein in the polypeptide chain is referred to as the N-extein while the protein sequence immediately following the intein is referred to as the C-extein. Once the intein undergoes autocatalytic excision, the two extein bodies are ligated to form a fully mature host protein.

The majority of inteins splice in a four-step mechanism with the intein autocatalytically excising itself from the host proteins (Figure 1). The first step in intein splicing is a nucleophilic (cysteine or serine) attack on the peptide bond preceding the intein. This attack creates a thioester linkage. Next, a nucleophile at the first position of the C-extein (cysteine, serine, or threonine) attacks the thioester, creating a branched intermediate. The third step begins when the branched intermediate releases the free intein from the exteins through the cyclization of the terminal intein residue, an asparagine. The reaction is completed by the thioester rearrangement, connecting the two exteins (N-extein and C-extein) to form a new peptide bond and a mature host protein. Off-pathway cleavage of the N- or C-extein can also occur, a reaction that can be conditional on the environment as well as useful in protein engineering applications. Approximately 150 amino acids contribute to the splicing of inteins, folding into a horseshoe shape and positioning the N-extein and C-extein in proximity to the excision site.

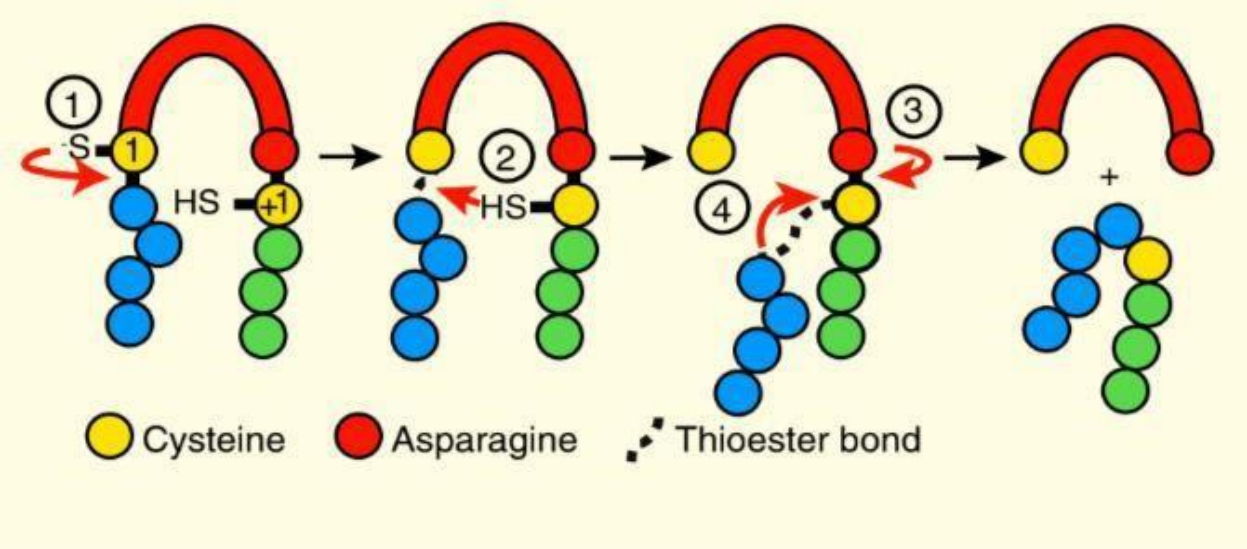

Figure 1: Splicing mechanism reaction of Class I Inteins. Adapted from Lennon and Belfort, 2017 *Current Biology*. Pictured is the four-step splicing mechanism of inteins with color-coordinated regions to identify the different components and reactions within the mechanism.

This mechanism of self-excision and post-splicing extein ligation is intriguing and has several applications across different fields of science. In protein engineering, intein cleavage and splicing reactions are used to mediate protein purification and protein labeling. The intein reaction places an affinity tag on the protein intended to be purified and can be removed via induced cleavage. Inteins can also be used to modify proteins by rearranging the present peptide bonds. A mutated intein can be moved to a non-native location and splicing can be induced, causing backbone cyclization, site specific labeling, or proteolysis, generating covalent modifications to the desired protein (Wood and Camarero, 2014).

One example of the utility of inteins is modifications to the CRISPR/Cas9 system. The CRISPR/Cas9 system is the primary method of genome editing via insertion, deletion, or

modification of DNA. Changes in the genome can cause changes in the organism such as physical traits or reduced disease risks. The CRISPR/Cas9 system has limitations though due to the size of Cas9, impeding efficient delivery of Cas9. Inteins can be split and expressed as two polypeptides that splice upon reassembly, which can be used to deliver Cas9 as fragments that can reassemble. Intein-mediated split-Cas9 can be packaged, delivered and its nuclease activity reconstituted efficiently, improving the efficiency of the CRISPR/Cas9 system (Truong et. al, 2015). Inteins also have biomedical applications. As inteins are absent in humans, and present in essential genes of pathogens such as *Mycobacterium tuberculosis*, the causative agent of tuberculosis, specific protein splicing inhibitors could serve as novel antibacterials.

Pyrococcus horikoshii is an anaerobic, hyperthermophilic archaeon species first isolated from hydrothermal vents from the East China sea. The discovery of the RadA intein was a revolutionary find for understanding intein-based regulation as it possesses the first intein that is vulnerable to substrate-induced splicing. As a recombinase, RadA can detect DNA damage and initiate repair of damaged DNA through homologous recombination. In the natural extein context, extein-intein interactions block splicing. However, in foreign exteins, the RadA intein is able to excise rapidly (Topilina et al, 2015). This post-translational regulation mechanism allows the RadA intein to be synthesized and rest in an inactive state. Remarkably, introduction of single-stranded DNA (ssDNA), a substrate of RadA and signal that recombinase activity is needed by the cell, disrupts the intein-extein interactions, allowing the intein to self-excise with tremendous speed and accuracy (Lennon et al., 2016, 2018, 2019).

These findings suggest that RadA coevolved with the intein as means to respond to DNA damage, representing an exciting new form of post-translational regulation.

The objective of this study was to examine the splicing behavior of *Pho* RadA intein in several different non-native locations within native exteins, which is a topic that has yet to be explored. By relocating the intein to neighboring threonine residues, we determined if other threonines of RadA can be utilized as the +1 nucleophile to facilitate splicing. By further researching the *Pho* RadA system, new knowledge of post-translational regulation by inteins and major factors controlling intein site selection in natural and non-native systems may be discovered.

Experimental Design

The *Pho* RadA intein requires a specific amino acid comtext in order to properly splice. RadA must have a nucleophile upstream in the +1 position which is the first amino acid located within the c-extein. In the case of *Pho* RadA, this is a threonine (Oeemig et al., 2012). While several other internal and external factors have been attributed to the speed and accuracy of RadA splicing, the only necessary requirement for RadA splicing is the presence of a threonine as a +1 nucleophile (Topilina et al. 2015). To test the hypothesis of coevolution of *Pho* RadA and the intein as a means of post-translational protein regulation and further study site selection for the intein, the intein was moved from the native location to similar threonine containing sites within the RadA gene. By examining the splicing activity of the intein in alternate sites, the importance of intein location within the RadA protein can be

assessed. Variants of *Pho* RadA gene with the intein artificially inserted at various positions in the DNA sequence, were previously cloned into a pET vector for their expression (Figure 2).

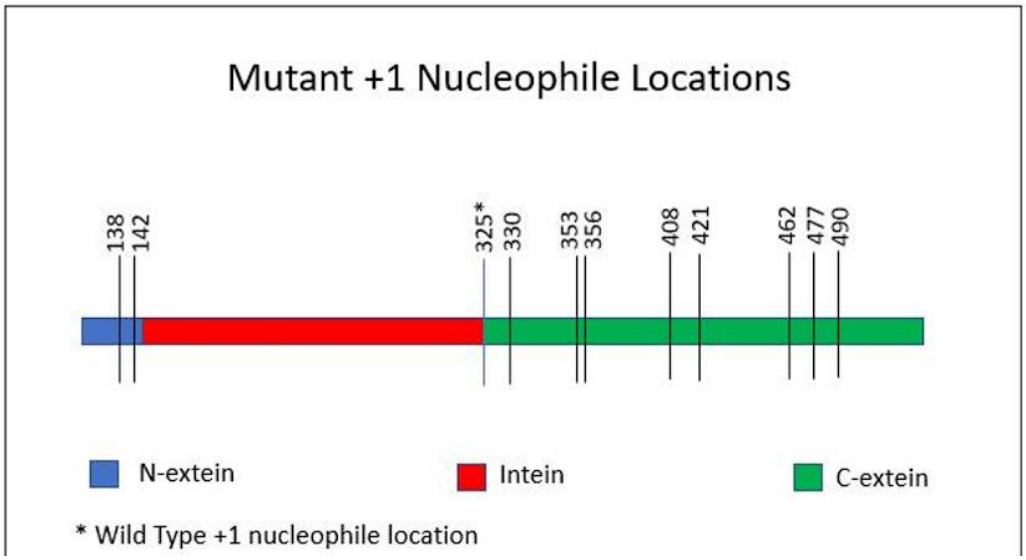

Figure 2: Locations of *Pho* RadA threonines (+1 nucleophile) within the primary protein sequence. Each section represents the location of an alternate threonine, which can act as a +1 nucleophile in the case of intein relocation. The threonines are labeled T138, T142, T330, T353, T356, T408, T421, T462, T477, and T490, respectively. The blue section indicates the wild type N-extein, the red portion indicates the natural (wild type) location of the intein, while the green section indicates the wild type C-extein.

The variant RadA proteins were overexpressed from plasmids in *Escherichia coli* and purified by nickel affinity chromatography. Post purification, the splicing competency was tested. The pET vector used was a lab-designed plasmid that promotes protein expression specifically at the cloning region containing the target *Pho* RadA gene. The expression of the RadA gene is regulated by the Lac repressor, which can be induced via IPTG (Isopropyl β- d-1-thiogalactopyranoside) and other lac operon regulatory mechanisms, allowing for RadA gene expression to occur when necessary. Each pET vector also included an antibiotic

resistance gene, β-Lactamase, allowing us to select only *E. coli* cells that retained the plasmid following transformation. The colonies were grown on ampicillin agar plates causing the death of any *E. coli* not containing the β-Lactamase antibiotic resistance gene. The pET vectors also contained genetic coding for a chain of 6 histidines upstream of the N-extein, which allows protein purification as it has high affinity for nickel resin.

Materials and Methods

TRANSFORMATION

In order to run different tests on the RadA variants, each plasmid must be transformed into a strain of *E. coli*, BL21(DE3). The purpose of transformation is to introduce the foreign DNA plasmid into the bacteria so that subsequently the RadA proteins can be expressed. Transformation begins with the thawing of BL21 (DE3) competent *E. coli* cells, which are stored at -80°C, on ice. Once thawed, 1µl of plasmid DNA (50 ng) was added to cell mixture and flicked to ensure proper mixing of cells and DNA. The mixture was then replaced on ice, heat shocked at 42°C, then replaced on ice again. 1mL of LB broth was added to the mixture as an outgrowth solution. The solution was then incubated at 37°C and shook vigorously (250 rpm) for approximately an hour. After incubation, the solution was pipetted onto Luria-Bertani (LB) agar plates with ampicillin and spread via the copacabana method to give isolated colonies of transformed *E. coli.* The plates were placed in the incubator at 37°C overnight to allow for proper growth. The transformed plasmid contained β-lactamase, an antibiotic-resistance gene.

The ampicillin present in the agar plates acted as a bacterial selector and killed any *E. coli* bacteria that did not retain the plasmid from the transformation.

INDUCTION AND CELL GROWTH

Prior to continuing the experiment, the transformed cells must be grown at a high concentration. A 5 mL starter culture was inoculated with a colony from our transformation plate and incubated overnight. Using sterile technique, the starter culture was added to the growth media containing ampicillin at a 1:100 volume ratio (400mL total), and the flask was placed inside a shaking incubator set to 37°C, 250rpm.

Optical density was recorded using a spectrophotometer to measure cell growth. The spectrophotometer readings were recorded until the cell mixture reached an optical density of 0.5, indicating the middle of log phase growth. The stage of bacterial growth known as log phase indicates exponential growth of the bacteria. During the log phase, the bacteria contain peak ribosome levels. The ribosomes are a key aspect in protein synthesis and overexpression of the specific intein carrying target protein. Once the log phase was reached, Isopropyl β- d-1-thiogalactopyranoside (IPTG) was added to the mixture to induce the overexpression of the target protein.

PROTEIN EXPRESSION AND PURIFICATION

Once overexpression of our target protein within the *E. coli* was completed, protein purification must occur. The process of protein purification involves separating the *E.*

coli cells containing our overexpressed protein from the growth media they were growing in. The first step requires separating the cells from the growth media is centrifuging. By placing the media mixture into a centrifuge tube and centrifuging at 3000xG, the cells form a pellet at the bottom of the tube, allowing for clean separation from the media. After decanting the media from the cell pellets, 20 mM Tris pH 8, 500 mM NaCl, and 10 mM Imidazole (Nickel Buffer A) was used to resuspend the pellet. The pellet was froze at -20°C overnight.

The separated cell cultures were removed from the freezer and placed on ice to thaw. The ice allows the cells to thaw but prevents denaturation due to excess heat produced by the sonicator. The sonicator probe is a device used to produce sound energy that agitates particles, causing the lysis of cells. The sonicator probe was placed into our solution and ran for intermittent periods of thirty seconds on, thirty seconds off. After sonication was completed, the lysate was once again centrifuged to separate the cell debris from the protein of interest. After centrifuging at 20,000xG, the supernatant was removed from cell debris via decanting.

The next stage of the purification process is to use a magnetic nickel resin. Nickel resin is used for immobilized metal affinity chromatography (IMAC) for the purification of recombinant proteins with a polyhistidine tag. Using a magnetic rack and the magnetic nickel beads, each RadA lysate was applied and allowed a short time period (approximately five minutes) to ensure proper binding. The polyhistidine tag present on the target protein binds to the nickel in the resin, allowing us to wash away any residual

proteins or cell matter with a buffer washing solution. Since the target protein contains an affinity His-tag, the RadA containing proteins remained bound. To dislodge the bound RadA proteins after washing, an eluate solution with a higher affinity for nickel was used. Due to the higher affinity of imidazole, the imidazole-based solution removed the RadA protein from the column and allowed for collection of the purified target protein.

SPLICING ASSAY

After expression and purification of each of the target RadA proteins, a splicing assay was performed to determine the necessary conditions for the inteins to undergo splicing. Two variables were used in this experiment: temperature and single-stranded DNA (ssDNA) addition. Each mutant RadA protein sample was incubated at 63°C, 75°C, and 85°C for fifteen minutes with and without 188ng/μl of single-stranded DNA.

SDS-PAGE

The final stage necessary to analyze the results of the experiment is to perform SDS PAGE. By loading our samples onto a gel and introducing an electric field the proteins can be separated by molecular weight. When proteins are separated in the gel by electrophoresis, the smaller proteins move faster due to less resistance in the gel. In order to ensure size was the only main contributor to the rate of migration, sodium dodecyl sulfate (SDS) is used. Sodium dodecyl sulfate (SDS) is a detergent that has a strong protein-denaturing effect which binds to the protein backbone at a constant molar ratio regardless of protein size. The polymerized acrylamide gel forms a mesh-like matrix suitable for the

separation of proteins which is durable and easy to handle. When the proteins are exposed to the SDS and reducing agent, the proteins unfold into chains with negative charge proportional to the polypeptide chain length. A Tris-Glycine-SDS buffer was used in the electrophoresis chamber. A size marker ladder and samples were loaded into the gel cassette. Using a micropipette, twenty μl of each sample was loaded into their respective reservoir wells. The electrophoresis unit was connected to a power supply at 200 volts. Upon completion of electrophoresis, the gel cassette was opened and the gel was rinsed with distilled water and stained using coomassie dye. Protein staining allows visualization and imaging of the proteins present in the gel. The staining process was completed using an eStain L1 Protein Staining System, which provided fast, automatic, efficient, and reliable stains. After the staining process was completed, the stained gels were imaged using epi-illumination.

Results

In order to evaluate the efficiency of intein splicing in non-native location within the host protein, the intein was relocated to use different threonine residues as the +1 nucleophile. The wild type intein was used as a base control to quantify splicing within the mutant *Pho* RadA inteins. The relocation of the intein to the threonines at T138, T477, T353, T408 was examined, with T325 as the natural position for reference. Each variant of the *Pho* RadA was initially monitored for splicing ability without the addition of heat or single-stranded DNA during purification, as some splicing should occur during this step. There were many characteristics noted for each variant prior to beginning the experiment, including the threonine position, -1 residue, and the secondary structure of where the intein is inserted (Table 1).

Table 1: Table containing information regarding the threonine used as a +1 residue, the -1 residue, predicted -1 splicing efficiency based on NMR and crystal structures of the *Pyrococcus horikoshii* RadA intein (Oeemig et al.,), and the secondary structure of the intein. T325 is marked as the native intein +1 nucleophile.

Threonine as +1 Nucleophile	-1 Residue	Predicted -1 Residue Splicing Efficiency	Secondary Structure of Insertion
138	Glutamic acid	40%	Loop
477	Proline	23%	Unstructured
353	Asparagine	95%	Loop
408	Asparagine	95%	Loop
325 (*)	Lysine	97%	Loop-Native Site

Analysis of the proteins through electrophoresis showed that the control wildtype *Pho* RadA spliced under all conditions, however, none of the mutants displayed splicing competence. (Figures 4 & 5). Furthermore, the T353 mutant, which had the highest expected splicing efficiency due to the loop positioning and asparagine at the -1 residue also showed no splicing (Figure 3).

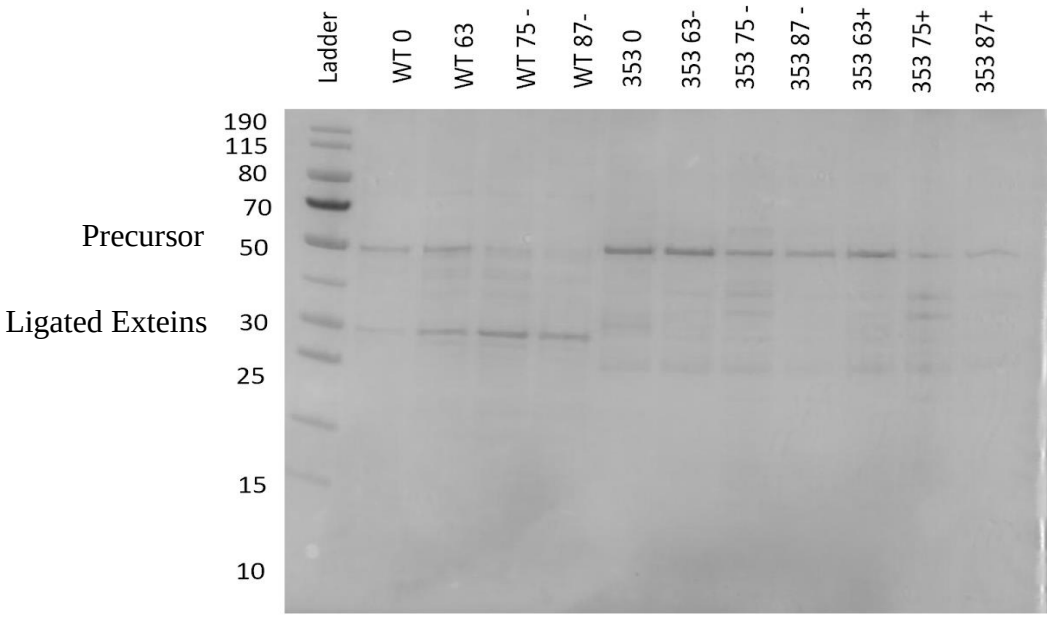

Figure 3: Splicing assays for wildtype (WT) *Pho* RadA and T353 mutant. This gel compares the splicing capability of Wild Type (WT) RadA and the 353 T+1 mutant under stated conditions. The ladder in the first column serves as reference for amino acid chain size.

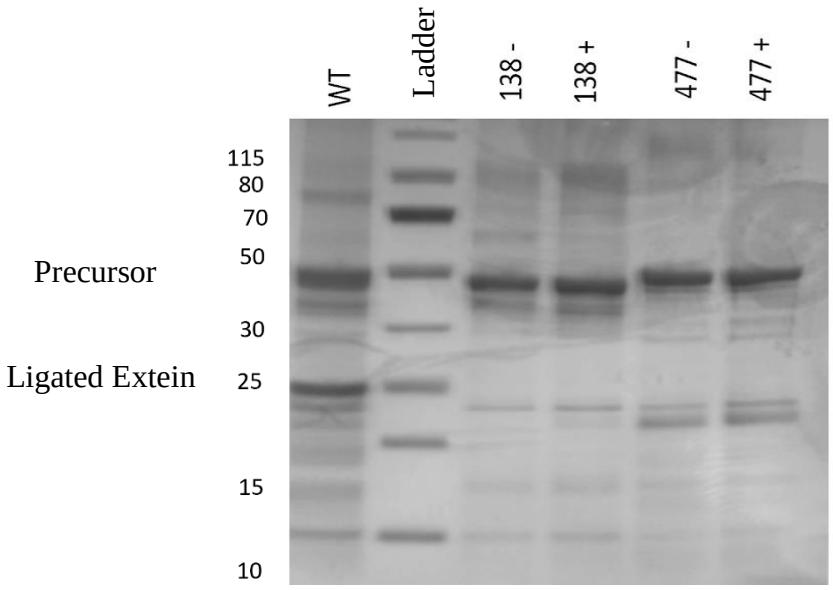

Figure 4: Splicing assay for T138 and T477 mutants compared to the wildtype (WT) *Pho* RadA at 75℃ for fifteen minutes. No splicing occurred in mutants, but splicing occurred for wildtype.

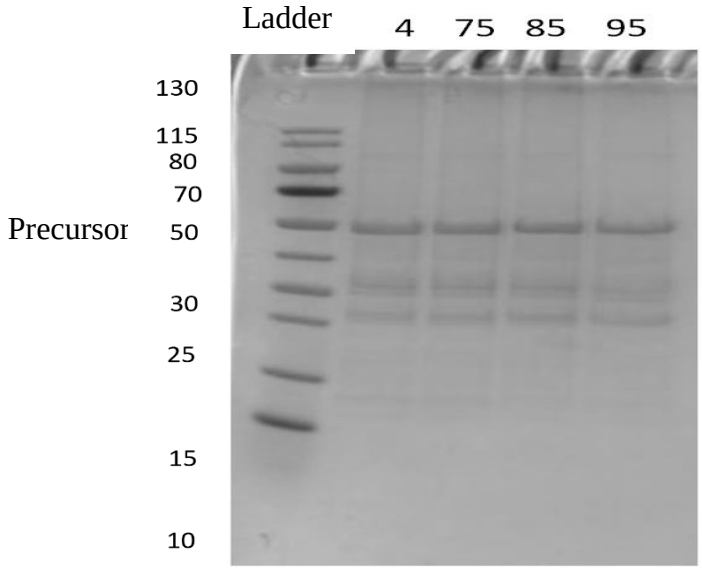

Figure 5: Splicing assay for T408 mutant at indicated temperature for fifteen minutes. No splicing occurred.

Preliminary results show that the wild type *Pho* RadA spliced under all conditions tested while the variants did not splice under any conditions, including the temperature changes and the addition of ssDNA. The equivalent threonine positions on Pho were modeled onto the structure of *Pyrococcus furiosus* RadA (Figures 6,7,8, and 9).

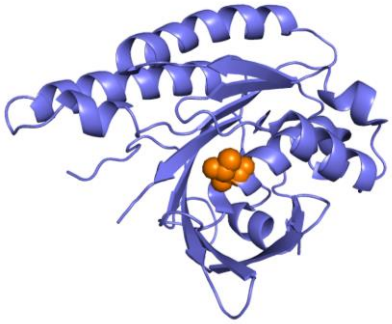

Figure 6: T325 Equivalent position on *Pyrococcus furiosus* RadA (Native +1).

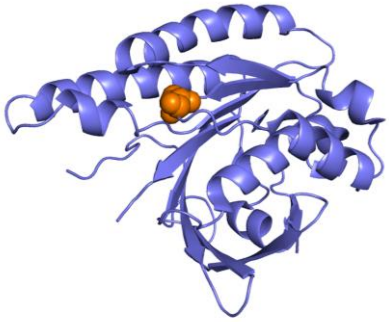

Figure 7: T353 Equivalent position on *Pyrococcus furiosus* RadA.

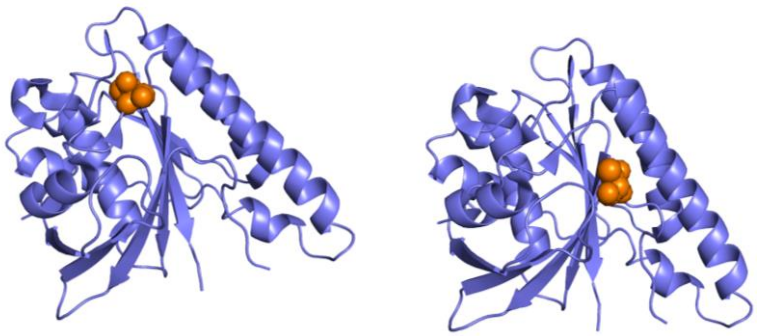

Figure 8: T138 and T477 Equivalent positions on *Pyrococcus furiosus* RadA, respectively.

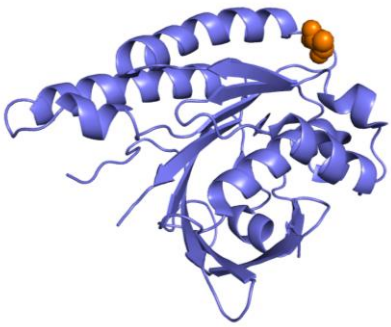

Figure 9: T408 Equivalent position on *Pyrococcus furiosus* RadA.

Discussion

Each variant protein was expressed using *E. coli* cells, purified using nickel resin pulldown chromatography, and tested under two additional treatment groups. The first treatment involved the addition of heat while the second treatment involved the addition of heat and ssDNA. Each sample group was heated and recorded at 63°C, 75°C, and 85°C, respectively. The overall results of the experiment were not expected. The splicing incompetence of each of the variant inteins is surprising, since all of the necessary conditions were met. Each variant of the *Pho* RadA intein had the same +1 nucleophile, and at least in two cases favorable residues in the -1 position.

The splicing incompetence in the variant inteins could possibly be explained by the secondary structure of the precursor protein surrounding the active splicing site. However, this conclusion is inconsistent due to the splicing incapability of inteins located in the loop and unstructured sites. The T353 variant of the *Pho* RadA intein was especially intriguing due to the secondary structure and the asparagine -1 residue. This intein had great potential to splice but did not.

Future directions to further understand the splicing behaviors of *Pho* RadA inteins and the variants include further research including changing the -1 residue to a lysine, the native -1. The electrophoresis gels showed stable precursors for all of the variants, but there were never ligated exteins, which is important. Unfortunately, due to the COVID-19 pandemic, we are currently unavailable to continue research regarding the splicing incompetence of the mutant *Pho* RadA inteins.

References

Basturea, G. (2020, October 19). Inteins. Retrieved November 02, 2020, from https://www.labome.com/method/Inteins.html

Daugelat, S., & Jacobs, W. R., Jr (1999). The Mycobacterium tuberculosis recA intein can be used in an ORFTRAP to select for open reading frames. *Protein science : a publication of the Protein Society*, *8*(3), 644–653. https://doi.org/10.1110/ps.8.3.644

Lennon, Christopher W., and Marlene Belfort. "Inteins." *Current Biology*, vol. 27, no. 6, 20 Mar. 2017, doi:10.1016/j.cub.2017.01.016.

Lennon Christopher W., Stanger M, Belfort M. Protein splicing of a recombinase intein induced by ssDNA and DNA damage. Genes Dev. 2016;30(24):2663-2668. doi:10.1101/gad.289280.116

Lennon Christopher W., Stanger MJ, Belfort M. Mechanism of Single-Stranded DNA Activation of Recombinase Intein Splicing. Biochemistry. 2019 Aug 6;58(31):3335-3339. doi: 10.1021/acs.biochem.9b00506. Epub 2019 Jul 23. PMID: 31318538.

Lennon, Christopher W., et al. "Protein Splicing of a Recombinase Intein Induced by SsDNA and DNA Damage." *Genes & Development*, vol. 30, no. 24, 2016, pp. 2663– 2668., doi:10.1101/gad.289280.116.

Oeemig, Jesper S., et al. "NMR and Crystal Structures of the Pyrococcus Horikoshii RadA Intein Guide a Strategy for Engineering a Highly Efficient and Promiscuous Intein." *Journal of Molecular Biology*, vol. 421, no. 1, 3 Aug. 2012, pp. 85–99., doi:10.1016/j.jmb.2012.04.029.

Novikova O, Jayachandran P, Kelley DS, et al. Intein Clustering Suggests Functional Importance in Different Domains of Life. *Mol Biol Evol*. 2016;33(3):783-799. doi:10.1093/molbev/msv271

Topilina, Natalya I., et al. “Post-Translational Environmental Switch of RadA Activity by Extein–Intein Interactions in Protein Splicing.” *Nucleic Acids Research*, vol. 43, no. 13, 2015, pp. 6631–6648., doi:10.1093/nar/gkv612.

Wood, David W., and Julio A. Camarero. “Intein Applications: From Protein Purification and Labeling to Metabolic Control Methods.” *Journal of Biological Chemistry*, vol. 289, no. 21, 2014, pp. 14512–14519., doi:10.1074/jbc.r114.552653.