Investigating Sequence Variation Effects on Shine Dalgarno in Tau

A Major Qualifying Project

Submitted to the Faculty of

Worcester Polytechnic Institute

In partial fulfillment of the requirements for the

Degree in Bachelor of Science

In

Biology and Biotechnology

By

Jillian Comeau

Date: March 26th, 2020

WPI Project Advisor: Lou Roberts, PhD

AbbVie Project Advisors: Keenan Taylor, PhD & Andrew DiChiara, PhD

Abstract

The microtubule binding protein Tau is the major component of brain lesions seen in patients with Alzheimer's Disease. Because of its role in neurodegeneration, the protein is a major focus for researchers eager to understand what drives Tau to aggregate and disrupt neurological functioning. To fulfill the need for Tau, the recombinant forms of the protein are expressed in *E. coli*. Using a bacterial host to produce Tau subjects the gene to bacterial cell translational machinery and risks the expression of suboptimal protein forms, like truncated variants. The presented research investigates the sequence features of a recombinant Tau construct that strengthen a Shine Dalgarno sequence to initiate translation at a cryptic start site within the coding region. The results suggest that the initiation of translation at the cryptic start site is dependent on the sequence upstream of the Shine Dalgarno region.

Introduction

Alzheimer's Disease (AD) is a devastating disorder of the brain that typically starts presenting symptoms in adults in their mid-sixties. The Alzheimer's Association estimates that 5.8 million Americans are currently living with AD, a number which is only expected to increase in future years (Alzheimer's Association, 2019). AD is described as the progressive loss of memory and decline of cognitive function, leading to the onset of dementia symptoms. Insoluble fibers accumulate within the neurons of the brain, causing lesions (Goedert, 1993). Plaques and neurofibrillary tangles are fibrous aggregates seen in AD; however, plaques have been determined to be a downstream component of Alzheimer's rather than a root cause (Drachman, 2014). The lesions that develop within nerve cells are composed of paired helical filaments (PHFs), which are further comprised of the microtubule binding protein Tau (MAPT) (Goedert 1993). There are six isoforms of Tau produced from alternative mRNA splicing of the *MAPT*

gene. The isoforms have varying repeat regions in their microtubule binding domains, leading to variation in tauopathies, which are neurodegenerative disorders characterized by abnormal tau protein deposits in the brain (Berriman et al., 2003).

When functioning appropriately, Tau stabilizes microtubules by allowing polymerization of microtubule neurons (Woerman et al., 2016). In AD, Tau is found to be hyperphosphorylated, likely causing the reduced binding capacity of the protein to microtubules coincident with the disease (Brunden et al., 2009; Giustiniani, 2014). The hyperphosphorylation of Tau is also believed to play a role in the misfolding of the protein, and with abnormal folding patterns and Tau aggregation come the PHFs seen in AD. Within the central nervous system, Tau is soluble and has no defined structure, thus classifying it as intrinsically disordered. Tau aggregates when it becomes insoluble due to the misfolding of the protein, and the aggregates spread from their point of origin through neuronal connections; the connectivity of the brain is a key factor of Tau's ability to propagate across the brain (Goedert, 2017).

Researchers are eager to learn more about what drives aggregation of Tau. Identifying ways to inhibit or even reverse the effects of Tau aggregation is of considerable interest in the biopharmaceutical industry (Paranjape et. al., 2015.) Along with the demand for a treatment of tauopathies comes the immense need for pure biological Tau protein for tests and assays. For example, in order to test azaphilone derivatives to determine their effects on Tau aggregate assembly *in vitro*, Paranjape et al. required full length 2N4R Tau protein for their methodological approach. This material was obtained through *E. coli* cell culture. *E. coli* is the most common host of choice for mass production of Tau, because recombinant forms of the protein can be expressed at will to achieve quantities of any of the six possible isoforms (Krishnakumar and Gupta, 2017). The six different isoforms of Tau contain variations of the protein's repeat region,

which make up Tau's microtubule binding units (Goedert and Jakes, 1990). However, though the expression of Tau in *E. coli* is preferred for research purposes, it is not without limitations. Protein expression in bacterial hosts subjects the genetic material to translational machinery governed by the bacterial cell. Such regulation could result in the desired protein being translated in less-than-optimal forms, such as truncated variants.

There are sequences that exist within prokaryotic and eukaryotic DNA that help to promote translation. In eukaryotic DNA (including DNA in human cells) these regions are called Kozak sequences. The Kozak sequence extends upstream of the start codon in eukaryotes and varies in length and composition among organisms (Li et al., 2017).

In prokaryotic DNA, such as in bacterial cells like *E. coli*, similar conserved regions are known as Shine Dalgarno (SD) sequences. SD sequences typically exist in the mRNA close to the initiation codon (Li, Oh, and Weissman, 2012). The SD consensus sequence AGGAGG interacts with the anti-SD sequence at the 3' end of the 16S rRNA, anchoring the translation initiation region of the mRNA to the 30S ribosomal subunit to form a complex. It is this complex that allows the start codon to be selected and initiation of translation to begin (Malys, 2012). The SD sequence is that the longer and better match to the consensus the SD sequence is, the more effective it initiates translation (Malys, 2012). The spacing between the SD sequence and the initiation codon is also important for creating optimal initiation conditions. Usually, approximately 7-9 base pairs between the SD sequence and the start codon is best for initiating translation (Malys, 2012).

Though SD sequences are usually only found prior to the start codon in the untranslated region, it is entirely possible that an AGGAGG type sequence can be present further downstream and within the defined coding region. Without a corresponding start codon, such a sequence

would be harmless and most likely not affect translation of the protein. However, if there exists a downstream cryptic start codon along with an AGGAGG sequence, it is possible that initiation could also be prompted at the downstream site as well (Slobodskaya, 1996). In this case, a cryptic start site, if selected, would result in a truncated form of the protein being produced.

The presence of truncated protein introduces challenges into the processes of protein expression and purification. Truncated protein is not only a waste of cellular energy and machinery, but also presents the risk of altered activity or interference with full length protein production (Whitaker et al., 2014). Furthermore, truncated protein reduces the purity of expressed proteins as it can be challenging to remove through purification due to the similarity in size, charge, and other parameters compared to the full-length version of the protein. In the case of Tau, which even at full-length is challenging to purify due to its lack of structure and unpredictable behavior, the presence of a truncated protein product could greatly impede the production of pure protein for research and development purposes.

The presented research investigates the phenomena of protein truncation seen in Tau produced in *E. coli* BL21 DE3 cells. The Tau construct showing truncation had a single amino acid change at position 301 from proline to leucine (P301L), which was designed to produce the pathological mutant of Tau believed to lead to severe tauopathies and frontotemporal dementia (Whitaker et al 2014; Giustiniani, 2014). This construct was codon-optimized in the sense that the Tau gene was engineered with synonymous codon changes in order to increase protein production while maintaining amino acid identity (Mauro and Chappell. 2014). The conducted research eliminates this amino acid mutation as a cause of truncation. It was hypothesized that the synonymous codon changes at the nucleotide level included in the Tau mutant increased the strength of the SD region to initiate translation at an alternative start site within the coding

region. It was determined that the ability of the SD sequence to initiate translation is dependent on the sequence upstream of the SD region and alternative start codon.

Materials and Methods

Designing Constructs and Primers

All constructs were developed using Geneious Prime™ bioinformatics software for sequence analysis. The reference sequence for all constructs was either Wild Type Tau Original (Wild Type Tau) or Codon-Optimized Tau P301L (CO Tau P301L). Strategy 1 designed DNA was ordered from Genewiz. Primers for site directed mutagenesis were designed using Agilent Primer Design software and ordered through Integrated DNA Technologies (IDT). 150 nucleotide G blocks were also ordered through IDT. These designed G blocks and primers are shown in Appendix A. Individual nucleotide mutations are denoted by listing the original nucleotide, then the position of the nucleotide, followed by what the nucleotide was mutated to. For example, an adenine (A) to guanine (G) mutation at position 88 is A88G.

Transformations

50ng of DNA was added to 50 µL of competent cells (BL21 DE3 *E. coli* cells from New England Biolabs® or OneShot® Top10 cells from Invitrogen™) on ice for 30 minutes. The cells were heat shocked in a 42°C water bath for 30 seconds, then chilled on ice for 2 minutes. 450 µL of 2xYT nutrient rich medium was added to each sample tube and then shook at 37°C for 1 hour. 100 µL of each sample was plated onto a LB+Kanamycin (Kan) agar plate, spread with sterile glass beads, and incubated overnight at 37°C.

Colony Selection and Culturing

14 mL culture tubes were prepared by adding 5 mL 2xYT supplemented with 50 µg/mL Kanamycin (2xYT+Kan). A pipette tip was used to select one isolated colony from each sample, and was used to inoculate cultures, which were then incubated shaking at 200 rpm and 37°C overnight.

Inoculation and Induction

Cultures were removed from the incubator and each sample was added to LB+Kan in a bevel-bottomed 500 mL culture flask at a 1:50 inoculation ratio. The total culture volume typically used was 100 mL. The culture was incubated at 37°C at 200 rpm for 1.5 hours or until the optical density (OD600) of the samples was between 0.6-0.8. The samples were then shifted to 18°C while still shaking at 200 rpm for 20 minutes to allow samples to acclimate. Finally, 1 mL of pre-induced samples of each culture were taken for analysis by SDS PAGE protein gel electrophoresis, and the flasks were induced with 1mM of IPTG and incubated at 18°C and 200 rpm for 20-24 hours.

Harvest

Final OD600 readings were taken for all samples prior to harvesting samples to ensure that culture OD values were between 1.6 and 2.0. 1 mL samples of each culture were taken for analysis by SDS-PAGE. Cultures were then poured from flasks into 250mL conical centrifuge tubes and spun at 3000 rpm for 30 minutes at 4°C. Supernatant was discarded and pellets were frozen at -80°C.

Lysis

1 mL samples of cultures in 1.5 mL Eppendorf tubes were spun at 6000 rpm for 5 minutes on benchtop minicentrifuge. Samples were then placed on ice and supernatant was carefully removed via pipette and discarded. Remaining pellets were resuspended in 0.5 mL BugBuster™ Protein Extraction Reagent lysis detergent from Thermo Fisher Scientific™ to permeate the cell wall of *E. coli* without denaturing protein. Resuspended cells tubes were vortexed and chilled on ice for 10 minutes, and then the process was repeated one more time. Samples were then spun at 14,000 rpm for 5 minutes at 4°C. Supernatant (the soluble fraction)

was separated from the insoluble pellet. The insoluble pellet was resuspended with 0.5 mL BugBuster™, then all lysed samples were prepared for gel electrophoresis.

SDS-PAGE and gel staining

10 µL (~5-10 ug) of each sample was utilized. 10 µL of 100 mM DTT was added to samples as a reducing agent. 6X blue loading dye was added to samples, which were then boiled for 5 minutes. Each sample was added to lanes of NuPAGE 1D SDS-PAGE 4-12% gradient gels (Invitrogen™). 5µL of SeeBlue® Plus 2 Protein Marker from Invitrogen™ was used as size standard. Gels were then run in SDS PAGE tris-glycine electrophoresis buffer at 150 V for 1 hour, then developed either by Coomassie staining or Western Blotting. Gels were rinsed with water for 20 minutes while shaking gently, and the proteins were stained with Coomassie SimplyBlue™ Safe Stain from Invitrogen™.

Western Blotting

Gels were rinsed with sterile water, and protein transfer was performed using iBlot2™ Dry Blotting System from Thermo Fisher Scientific™ onto a nitrocellulose membrane per manufacturer's instructions. After transfer was complete, iBind™ Western System from Thermo Fisher Scientific™ was used to bind primary (1:4000 anti-Tau monoclonal) and secondary (1:2000 Anti-Mouse Alkaline Phosphatase Conjugate) antibodies to the membrane per manufacturer's instructions. The membrane was then developed using alkaline phosphatase substrate until bands appeared, then rinsed immediately with sterile water for imaging with document scanner.

Partial Purification: Barghorn Boiling Protocol

Due to the intrinsically disordered nature of Tau, it can survive boiling while other bacterial proteins are denatured. This method effectively cleans up the sample enough for

analysis without requiring the use of more expensive and time-consuming chromatographic techniques (Barghorn, Biernat, and Mandelkow, 2005). After centrifugation at 3000 rpm for 30 minutes in 250 mL conical tubes, the supernatant was discarded and the pellets were resuspended in 25 mL of a buffer solution containing 20 mM MES pH 6.8, 500 mM NaCl, 1 mM MgCl, 1 mM EGTA, and 2 mM DTT. Resuspended samples were kept on ice until boiling for 20 minutes, then placed on ice. The samples were then transferred to 50 mL centrifuge tubes and spun for 30 minutes at 30,000 x g. Following centrifugation, supernatant was decanted and saved for analysis by Coomassie staining and Western Blotting; pellets were discarded.

Plasmid Isolation

Plasmid isolation was completed using the E.Z.N.A.® Plasmid DNA Mini Kit II from Omega Biotek. 1 mL harvested sample pellets were resuspended with 250 µL resuspension solution with RNAase A, then the solution was transferred to a 1.5 mL Eppendorf tube. The resuspended samples were then mixed with 250 µL of high pH detergent lysis solution to solubilize and denature proteins in order to obtain a clear lysate. Then 350 µL of neutralizing acetic acid solution was added to the lysate and the tube was inverted until a flocculent white precipitate formed. The tube was then spun at 14,000 rpm for 10 minutes at room temperature. While the centrifugation was underway, the HiBind® DNA Mini Column was prepped by placing a column into a provided 2 mL collection tube. 700 µL of lysate was transferred from the centrifuged tube to the HiBind® column, while carefully avoiding transferring any of the white pellet formed. The column with the lysate was then spun at 14,000 rpm for 1 minute. The flow through was discarded from the column and the remaining lysate was added to the column and the spin was repeated, discarding the flow through after centrifugation again. 500 µL of HBC guanidine hydrochloride buffer with 100% Isopropanol was added to the column to wash the

sample bound to the column. The column with the wash was spun at 14,000 rpm for 1 minute, and flow through was discarded. Next, 700 µL of DNA wash buffer with 100% Ethanol, then spun with the same parameters again. The wash and spin were repeated once more, and flow through was discarded. Finally, the empty column was spun with the same parameters to ensure that all Ethanol was spun off the column, and flow through was discarded. The last step was to place the column into a 1.5 mL Eppendorf tube and elute the sample with 50 µL ddH2O, spinning at 14,000 rpm for 1 minute. The flow through in the tube was saved and labeled as final purified plasmid. The concentration of the plasmid was measured using a NanoDrop.

Seven Base Pair Variation Construct Designs- Strategy 1

To explore whether the seven nucleotide region between the SD sequence and the AUG cryptic start site effects ribosomal binding and truncation of the expressed protein, variations of the Wild Type Tau DNA sequence were made using Geneious Prime™ Software and ordered through Genewiz. The following variations were made:

1. ***WT-CO Tau*.** Wild Type Tau sequence with single amino acid mutation at position 301 changing proline to leucine. Construct designed as a control to show that the amino acid change is not solely responsible for the observed truncation of CO Tau P301L.
2. ***A84C-A90T***. Wild Type Tau sequence with A84C and A90T.
3. ***T87C*- A90T.** Wild Type Tau sequence with T87C and A90T.
4. ***A84V-T90C-A90T.*** Wild Type Tau sequence with A84C, T87C, and A90T.
5. ***A84T.*** Wild Type Tau sequence with A84T.
6. ***T87C.*** Wild Type Tau sequence with T87C.
7. ***A84C-T87C.*** Wild Type Tau sequence with A84C and T87C.

Site Directed Mutagenesis- Strategy 2

Using the Wild Type Tau sequence as the reference (or template) DNA, site-directed mutagenesis (SDM) was initially performed for three purposes: to create the "perfect" SD sequence, to "destroy" the SD sequence, and to change the codon identity of cryptic start site within the WT Tau vector. The primers were designed using the Agilent QuickChange® II Primer Design Program and were ordered through Genewiz. Four different primers were created to introduce the described changes into the WT Tau vector. The primers were each 18 nucleotides (6 codons) in length, identifying the first codon at position 76-78 containing the AGGAGG of the SD region, up to the cryptic start site ATG at nucleotides 91-93. The following mutations were integrated into each distinct primer to achieve the desired results:

1. ***"Perfect":*** primer was designed for the purpose of trying to create the SD sequence that matches the 16S rRNA as closely as possible. This primer changed three nucleotides within 18 nucleotide sequence from nucleotide region 76-93. The specific mutations included A84T, T85G, and A88C.
2. ***"Perfect Destroyed"*:** SD primer created to remove the AGGAGG of the SD region but still maintain similarity to the 16S rRNA as the "Perfect" primer. The AGGAGG was changed to a GGGGGG, introducing five total nucleotide changes within the 18 nucleotide region to remove the SD sequence. The specific mutations to be introduced included A78G, A81G, A84T, T85G, and A88C.
3. ***Met-Val*:** The first primer designed to change the cryptic start site included a mutation that changed the ATG to GTG, but otherwise maintain similarity to the 16S rRNA as in the "Perfect" primer. This primer, called "Met-Val" because it changed the cryptic methionine to valine, included the following mutations: A84T, T85G, A88C, and A91G.

4. ***Met-Leu*:** The second primer designed to change the cryptic start site included a mutation that would change the ATG to TTG, but otherwise maintain similarity to the 16S rRNA as the "Perfect" primer did. This primer, called "Met-Leu" because it changed the cryptic methionine to a leucine, included the following mutations: An A to T change at position 84, a T to G change at nucleotide 85, an A to C change at nucleotide 88, and an A to T change at nucleotide 91.

All actual primer sequences are seen in Table 1.

The Agilent QuickChange® II Site Directed Mutagenesis Kit was used. The kit functions to denature the parental vector with high temperature treatment at 95°C to allow for the designed forward and reverse primer to anneal to the denatured parent vector. The primer annealed to the strands then was extended via *Pfu* polymerase. The cycle of denaturing, annealing, and extending continues throughout the Polymerase Chain Reaction (PCR), amplifying the mutant DNA produced with each cycle. The reaction was run for 18 cycles to produce a sufficient quantity of mutant DNA. After PCR was complete, the amplified material was digested using DpnI, which digests methylated parental DNA. The resultant material containing only the mutant DNA was then transformed into XL1-Blue Super-competent cells.

Primer Name	Primer Description	Parental DNA	Forward Primer	Reverse Primer
Perfect	Close match to 16s rRNA	WT Tau	5'-ctcttggtcctgatgcattggatcacctccttgatccttacgatc-3'	5'-gatcgtaaggatcaaggaggtgatccaatgcatcaggaccaagag-3'
Destroyed	Close match to 16s rRNA without SD sequence AGGAGG	WT Tau	5'-attggatcacccccctgatccttacgatcgcccagg-3'	5'-cctgggcgatcgtaaggatcagggggtgatccaat-3'
Met-Val	Close match to 16s rRNA with alternative start codon Valine	WT Tau	5'-tcgccctcttggtcctgatgcactggatcacctccttgatccttacgatc-3'	5'-gatcgtaaggatcaaggaggtgatccagtgcatcaggaccaagagggcga-3'
Met-Leu	Close match to 16s rRNA with alternative start codon Leucine	WT Tau	5'-gatcgtaaggatcaaggaggtgatccattgcatcaggacctaagagggcga-3'	5'-cgccctcttggtcctgatgcaatggatcacctccttgatccttacgatc-3'

Table 2. Site Directed Mutagenesis Primer Designs. All primers designed using Agilent Primer Design Software.

Seven additional Site Directed Mutagenesis primers were designed based on the CO Tau P301L that showed truncation. These constructs were designed using the same Agilent Primer Design Program as described above. One of the seven primers designed in this set mutated the Leucine at amino acid position 301 back to a Proline (L301P) in order to achieve a sequence that had all the same characteristics as the CO Tau but without the amino acid mutation. The remaining six primers were designed to change one or two nucleotides within the 18 nucleotide range of the SD region to try and prevent truncation in combination with the P301L mutation. However, because SDM was not successful with the initial four primers, the remaining seven primers were not utilized.

150 G Block Construct Designs- Strategy 3

To test if mutations in the local region of the SD sequence effect the ability of the sequence to initiate ribosomal binding, constructs were designed using Geneious Prime™ Software and ordered through Genewiz. For the purposes of this experiment, "local region" was defined as the first 150 base pairs of the Codon-Optimized (CO) P301L DNA construct, starting at the primary start codon. Individual base pair mutations in four regional combinations were done using Geneious Prime™ Software. The 150 base pair constructs (G blocks) were then inserted into the Wild Type Tau vector for expression, so all base pairs downstream of the 150 nucleotide long G block inserted were of the Wild Type Tau DNA sequence. The first 150 base pairs of the CO sequence were changed in the following capacities:

1. ***CO Tau Mimic.*** The first 150 base pairs were left unchanged in order to provide a baseline of ribosomal binding which was expected to cause truncation.

2. ***CO After SD Tau.*** The nucleotides before the AGGGAGG of the SD sequence were changed to match the corresponding nucleotides of the Wild Type Tau sequence. From after the AGGAGG to the end of the 150 base pairs inserted was CO Tau nucleotides.
3. ***CO After/In SD Tau.*** The nucleotides before the ATG cryptic start site were changed to match the corresponding nucleotides of the Wild Type Tau sequence, with the exception of the AGGAGG of the SD sequence. From after the AGGAGG to the end of the 150 base pairs inserted was CO Tau nucleotides.
4. ***CO Before SD Tau.*** The nucleotides before the AGGAGG of the SD sequence remained CO Tau nucleotides. After the AGGAGG, base pairs were changed to match the corresponding nucleotides from Wild Type Tau up to the end of the 150 base pair inserted sequence.
5. ***CO Before/In SD Tau.*** The nucleotides before the AGGAGG of the SD sequence and within the AGGAGG and cryptic start site remained CO Tau nucleotides. After the cryptic ATG, base pairs were changed to match the corresponding nucleotides from Wild Type Tau up to the end of the 150 base pair inserted sequence.

Restriction Enzyme Digestion of DNA Templates and Constructs

First, the Wild Type Tau DNA Vector template and 150 base pair DNA constructs were digested with restriction enzymes. 100 ng of the template DNA was treated with 1 µL of NdeI and 1 µL of PstI, both from New England BioLabs®. 10X Cutsmart® buffer was added to the sample and incubated for 30 minutes at 37C. The digested DNA was then purified with the E.Z.N.A.® Cycle Pure Kit from Omega Biotek and eluted with 50 µL of ddH2O (150 bp insert constructs eluted with 40 µL to give a final concentration of 40 ng/µL).

Dephosphorylation of Digested DNA Templates

Dephosphorylation of the digested WT Tau DNA Vector was necessary to avoid re-ligation of the cut DNA into the vector. Dephosphorylation was completed using Antarctic Phosphatase from New England Biolabs®. 1 pmol vector was added to 2 µL of 10X Antarctic Phosphatase Reaction Buffer and 5 units of Antarctic Phosphatase. ddH2O was then added to a final volume of 20uL. The mixture was incubated at 37C for 30 minutes to ensure complete dephosphorylation of the cut ends. The ligated samples were then transformed into Top10 Cells.

Ligation of Digested DNA Samples

Ligation of digested Tau DNA insert constructs and vector was completed using the Quick Ligation Kit from New England Biolabs®. 50 ng of the digested and dephosphorylated WT Tau DNA vector was mixed with 37.5 ng of the digested Tau DNA insert constructs, along with 2X Quick Ligase Reaction Buffer and 1 µL of the Quick Ligase Enzyme. The entire mixture was brought to 20uL with ddH2O. The mixture was then incubated for 10 minutes at room temperature before being chilled on ice. The ligated DNA was then used to transform BL21 DE3 cells.

Results and Discussion

Wildtype Tau vs P301L (Truncating) Mutant Problem

The problem of truncation was originally seen in Codon-Optimized Tau with a Proline to Leucine mutation at amino acid 301 (CO Tau P301L). In this Tau construct, all other amino acids were identical to the Wild Type Tau. However, the nucleotide composition between the two constructs were different due to the codon-optimization algorithms used by the DNA supplier, Genewiz. Wild Type Tau and CO Tau P301L were expressed in BL21 DE3 cells and partially purified by boiling. There was an observed difference between the two proteins. CO Tau P301L yielded a protein band with slightly smaller molecular weight (~55 kDa), in addition to the expected full-length protein (~62 kDa). This phenomenon is seen in both Coomassie staining and Western blotting (Figure 1).

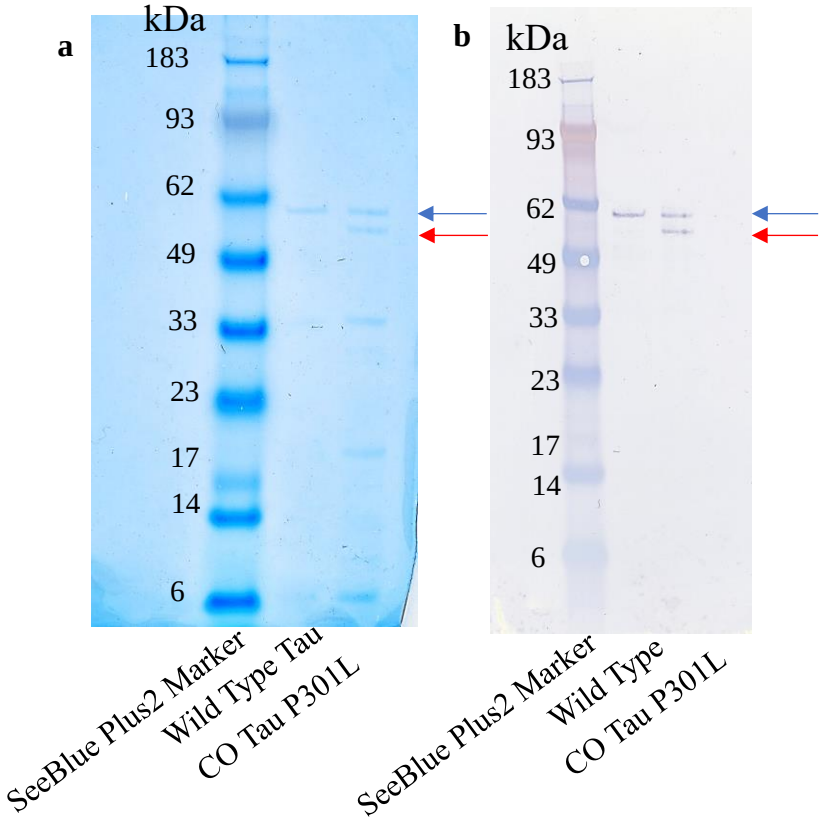

Figure 1. Comparison of Wild Type Tau to CO Tau P301L. Coomassie stained gel (a) and Western Blot (b) with SeeBlue® Plus2 Pre-stained Standard Marker from Invitrogen™. 10 uL of each sample loaded. Samples were partially purified by boiling. Single band of Wild Type Tau indicated by blue arrow. Truncation product seen directly below full-length Tau band in CO Tau P301L lane (red arrow), seen in approximately equal proportion.

To further investigate this phenomenon, the nucleotide sequences of both the wildtype and the mutant constructs were compared using Geneious Prime™ Sequence Alignment Software. There were found to be 282 individual nucleotide differences between the two constructs as a result of codon optimization when DNA sequences were purchased (Figure 2). Based on the information gathered from the qualitative gels and the nucleotide analysis, it was hypothesized that the nucleotide differences between the constructs was likely the cause of the truncation.

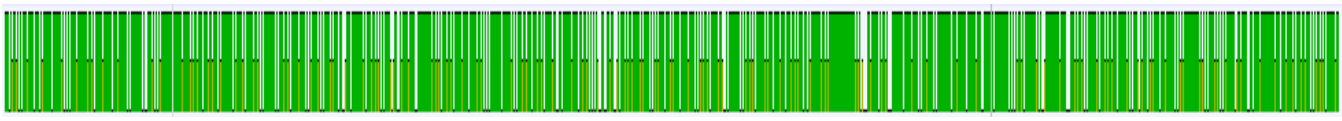

Figure 2. Wild Type Tau to CO Tau P301L Sequence Comparison. Green segments indicate matches between the Wild Type Tau nucleotide sequence and CO Tau P301L. White breaks indicate differences between the sequences. In total, there are 282 nucleotide differences. Only three of these from the single amino acid change at position 301 in CO Tau P301L compared to Wild Type Tau. Remaining 279 differences are the result of codon optimization.

Shine Dalgarno and Cryptic Start Site

To begin to identify which aspects of the CO Tau P301L construct were responsible for producing the truncated product, the Wild Type Tau and CO Tau P301L constructs were compared more closely using Geneious Prime™ Sequence Alignment Software. A SD sequence (AGGAGG) was identified seven base pairs upstream of methionine at nucleotide position 31 (Met31) in both the Wild Type and CO Tau P301L constructs. However, only in the CO Tau P301L construct did Met31 serve as a cryptic start site that resulted in the truncated protein, as seen in Figure 1. When examining the two constructs using Geneious Prime™ Software, there was found to be three nucleotide differences between the Wild Type Tau and CO Tau P301L in the region between the AGGAGG sequence and Met31 (Figure 3).

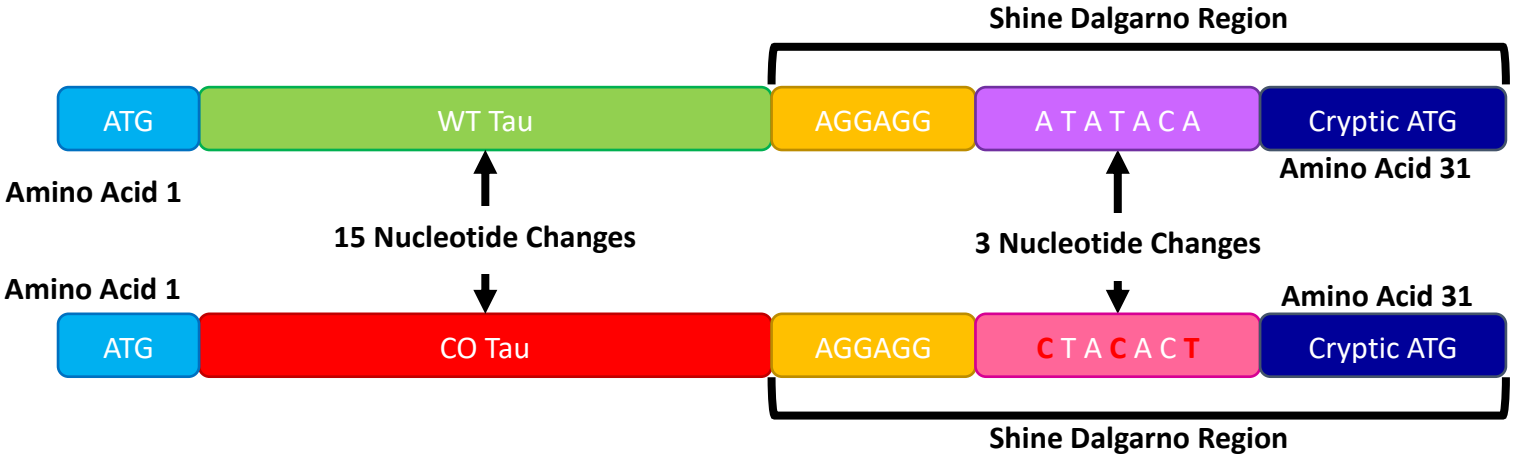

Figure 3. Wild Type Tau vs CO Tau P301L Shine Dalgarno Region schematic. Differences upstream of Met31 are detailed. Specific nucleotide differences between the Shine Dalgarno AGGAGG sequence and the Cryptic ATG are shown in red. Note that the AGGAGG SD sequence is conserved between Wild Type Tau and CO Tau P301L.

The observed differences between the two constructs served as the basis for the next hypothesis: the combination of nucleotides between the Shine-Dalgarno AGGAGG sequence and the Methionine 31 cryptic start site effect ribosomal binding, leading to translation at Met31 only in CO Tau P301L. To test this hypothesis, nine mutant Tau DNA sequences were developed based on the Wild Type Tau sequence. These constructs included a Wild Type Tau construct for a negative control, which was not expected to truncate. A Wild Type- CO Tau mutant was also designed, which was identical to the Wild Type Tau sequence on a nucleotide level but included the proline to leucine mutation at amino acid 301, just as the CO Tau P301L construct did. This construct was designed to determine if the single amino acid mutation had any effect on creating the truncated product. The seven other mutants included single, double, and triple mutations within the nucleotides between the AGGAGG SD sequence and the Met31 cryptic start site (Figure 4). All the mutants designed were identical to the Wild Type Tau construct on a

nucleotide level with the exception of the introduced mutations.

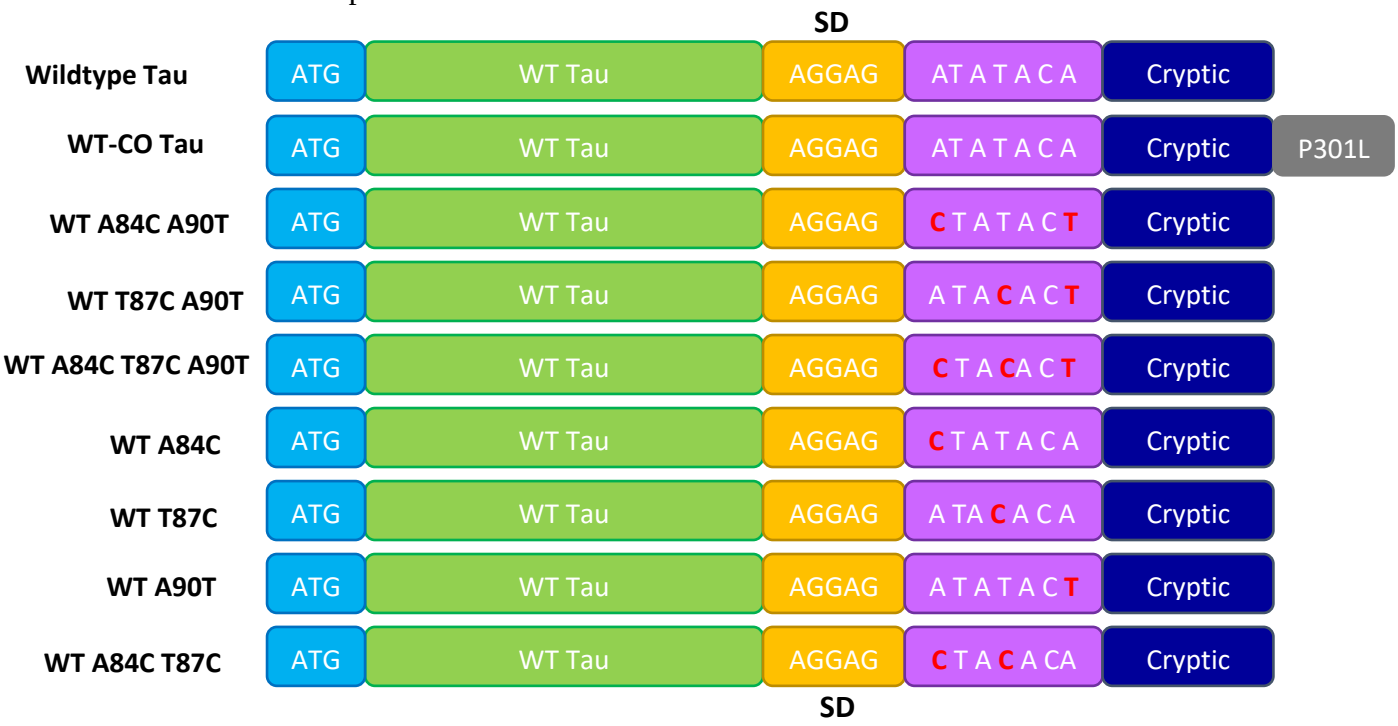

Figure 4. Shine Dalgarno Mutant Construct Designs. Designs of mutant 7 base pairs between SD AGGAGG and cryptic Met31 within Shine Dalgarno region. Constructs designed for the purpose of testing whether the identity of the base pairs in this region have effect on the strength of the Shine Dalgarno region for ribosomal binding and thus creation of the truncated product. Mutations made based on Wild Type Tau parental sequence, indicated in red. Note that WT-CO Tau is the Wild Type Tau sequence with the single proline to leucine amino acid change at amino acid position 301. All other construct names indicate that nucleotide change made at its respective position. (i.e. WT A90T has an A to T change at nucleotide position 90 within the Wild Type Tau sequence.)

Shine Dalgarno Mutant Constructs

The nine constructs designed were transformed into BL21 DE3 cells and the expressed protein was partially purified by boiling. The resulting purified lysates were then compared via Coomassie staining and Western Blot using an Anti-Tau monoclonal antibody. Included in this comparison were three controls: original Wild Type Tau, CO Tau P301L, and a repaired version of CO Tau P301L that had the cryptic Met31start site removed (Fixed CO Tau) and thus no longer truncated. None of the designed constructs showed truncation (Figures 5 and 6). The only

construct showing truncation was CO Tau P301L, as expected as the positive control. This indicated that the initiation of translation at an alternative start site is dependent on more than just the region between the Shine-Dalgarno AGGAGG and the Met31 start codon. Using this information, a second strategy was developed using site directed mutagenesis to try and induce truncation of Tau.

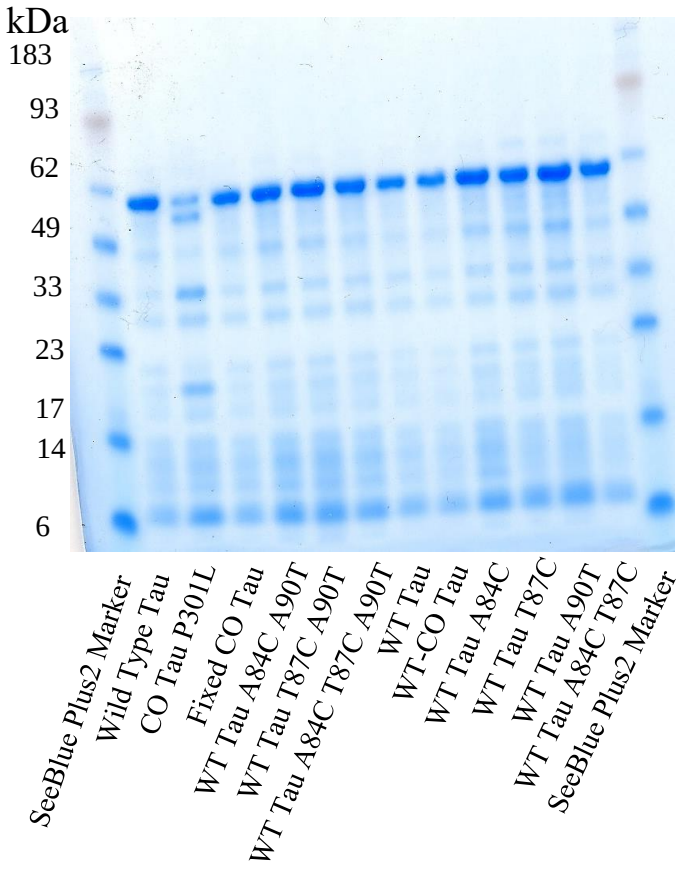

Figure 5. Shine Dalgarno Mutant Construct Results. Coomassie stained gel. Comparison of all designed mutant SD region constructs and Wild Type Tau, CO Tau P301L, and Fixed CO Tau (with cryptic start site removed) as controls. Wild Type Tau and Fixed CO Tau were not expected to truncate. CO Tau P301L was expected to truncate. No truncation is observed in any of the designed mutant constructs.

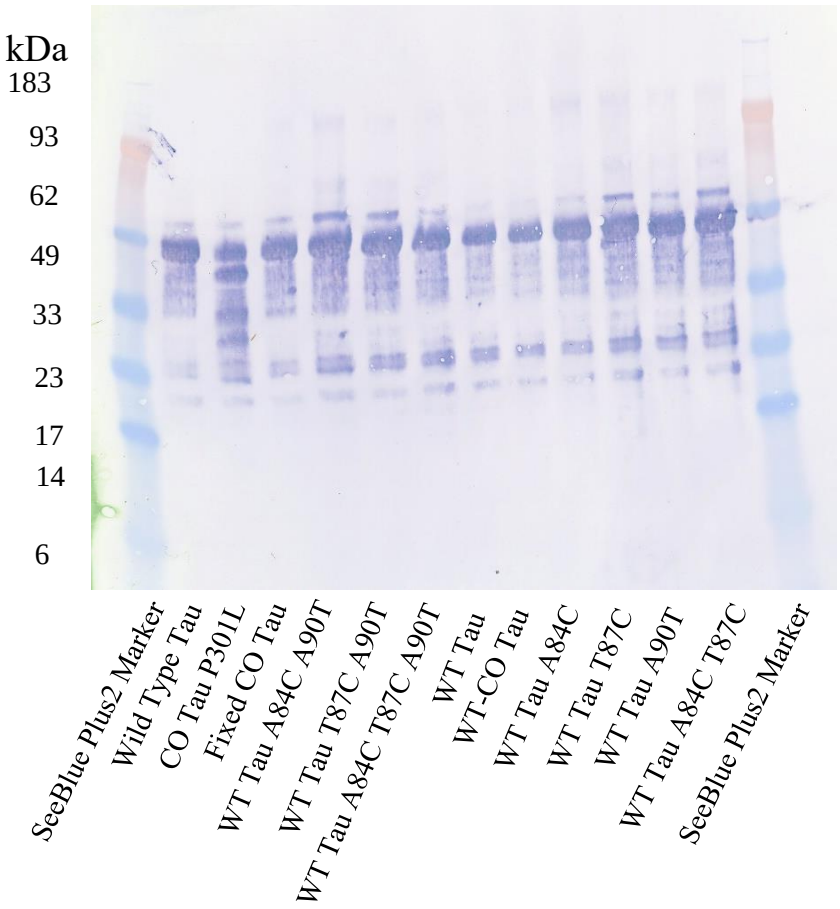

Figure 6. Shine Dalgarno Mutant Construct Results. Western Blot. Comparison of all designed mutant SD region constructs and Wild Type Tau, CO Tau P301L, and Fixed CO Tau (with cryptic start site removed) as controls. Wild Type Tau and Fixed CO Tau were not expected to truncate. CO Tau P301L was expected to truncate. No truncation is observed in any of the designed mutant constructs.

Site Directed Mutagenesis (SDM)

Four oligonucleotide primers were designed based on the Wild Type Tau DNA sequence. The four primers, Perfect, Perfect Destroyed, Met-Val, and Met-Leu (Table 1) were designed and the Agilent QuickChange® II Site Directed Mutagenesis Kit was used to perform SDM on the with the Wild Type Tau sequence as the parental vector. The mechanism of the reaction for SDM is described in Figure 7. None of the performed transformations yielded colonies for any of the samples. The experiment was repeated three times. Each time transformations were repeated twice with 50 uL of PCR product and 100 uL of PCR product for each sample. None of the repetitions yielded any colonies.

It was hypothesized that the reason the SDM reaction was failing was because the primers were self-annealing instead of annealing to the vector. Parameters of the SDM reaction were optimized in a final attempt to yield successful primer to vector annealing. The amount of primer was reduced, decreasing the concentration from 12.5 pmol to 1.25 pmol. The time spent in the annealing stage of SDM was also reduced by 50% from 30 seconds to 15 seconds. Lastly, the Wild Type Tau vector plasmid concentration was increased from 50 ng to 100 ng in attempt to push the annealing of oligonucleotide primers to the vector. After running SDM with the optimized parameters, still no colonies were successfully transformed into the XL1-Blue Supercompetent cells. It was assumed that the primers were self-annealing or creating hairpin structures inhibiting the function of the polymerase. However, these theories were neither proved nor denied through the presented research.

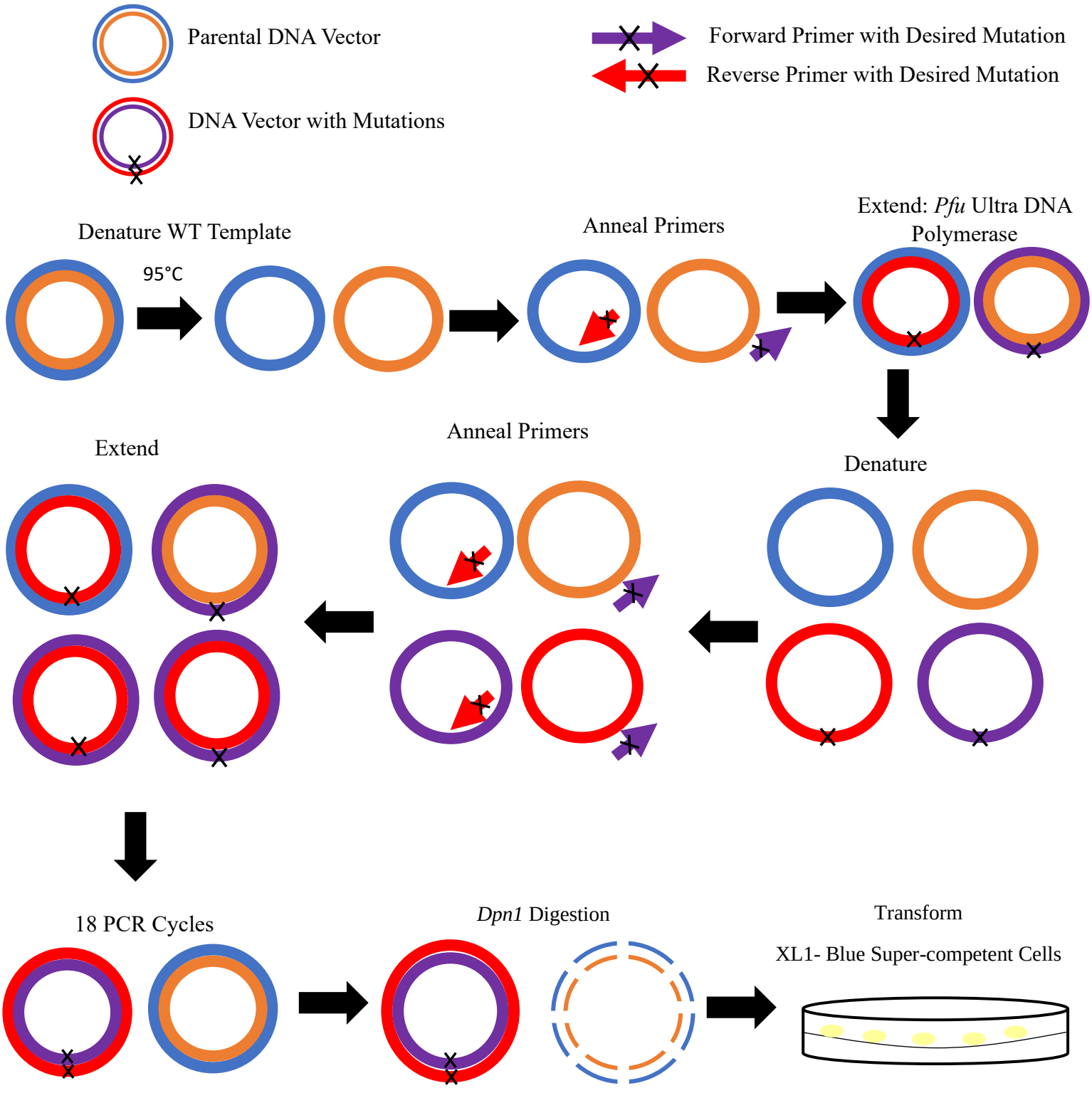

Figure 7. Site Directed Mutagenesis Experiment Schematic.

150 Nucleotide G Block Insert Constructs

Due to the inability of site directed mutagenesis to produce the desired mutations, a third strategy was devised. As shown in Figure 2, there are 282 nucleotide differences in the coding sequence between the Wildtype and CO Tau P301L constructs. 15 of these differences occur upstream of the Shine Dalgarno AGGAGG sequence, and three occur between the SD sequence and the Met31 cryptic start site. There are remaining 264 differences occur downstream of the cryptic start site (Figure 5A). Using the original CO Tau P301L construct as a base, five constructs were designed. Each construct was designed as a 150 nucleotide G block, including the region upstream of the SD sequence and a portion of the downstream region from the cryptic start site. Note that from the cryptic Met31 to the 150th nucleotide, there are 10 nucleotide differences between the CO Tau P301L construct and the Wild Type Tau construct (Figure 8).

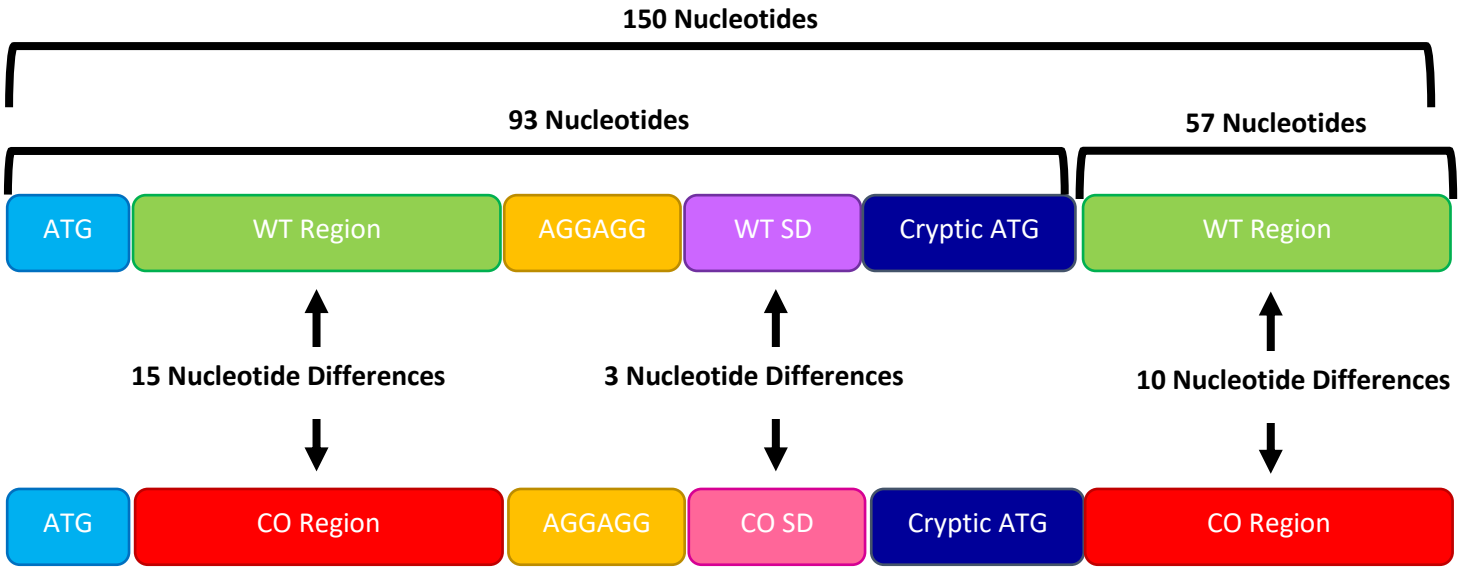

Figure 8. Comparison of the first 150 nucleotides of the Wild Type Tau and CO Tau P301L coding sequences. There are 15 differences upstream of the AGGAGG Shine Dalgarno region, 3 differences between the AGGAGG and Met31 Cryptic start site, and 10 differences downstream of the cryptic ATG.

Five mutants were designed as described (Figure 9):

1. ***CO Tau Mimic:*** First 150 base pairs of P301L sequence inserted into Wild Type Tau sequence in place of WT first 150 bp, expected to truncate the same as the original P301L sequence
2. ***CO After/In SD Mutant:*** sequence leading up to AGGAGG of SD region is the Wild Type Tau sequence, everything from AGGAGG onward is CO Tau P301L sequence
3. ***CO After SD Mutant:*** sequence through SD region up to ATG cryptic start site is the Wild Type Tau sequence, everything from ATG cryptic site onward is CO Tau P301L sequence
4. ***CO Before/In SD Mutant:*** sequence after ATG cryptic start site is the WT sequence, everything before ATG cryptic is P301L sequence
5. ***CO Before SD Mutant:*** sequence after AGGAGG of SD region is the WT sequence, everything before AGGAGG is P301L sequence

The CO Tau Mimic construct served as a positive control and was expected to truncate. Each of the designed constructs were inserted into the Wild Type Tau vector.

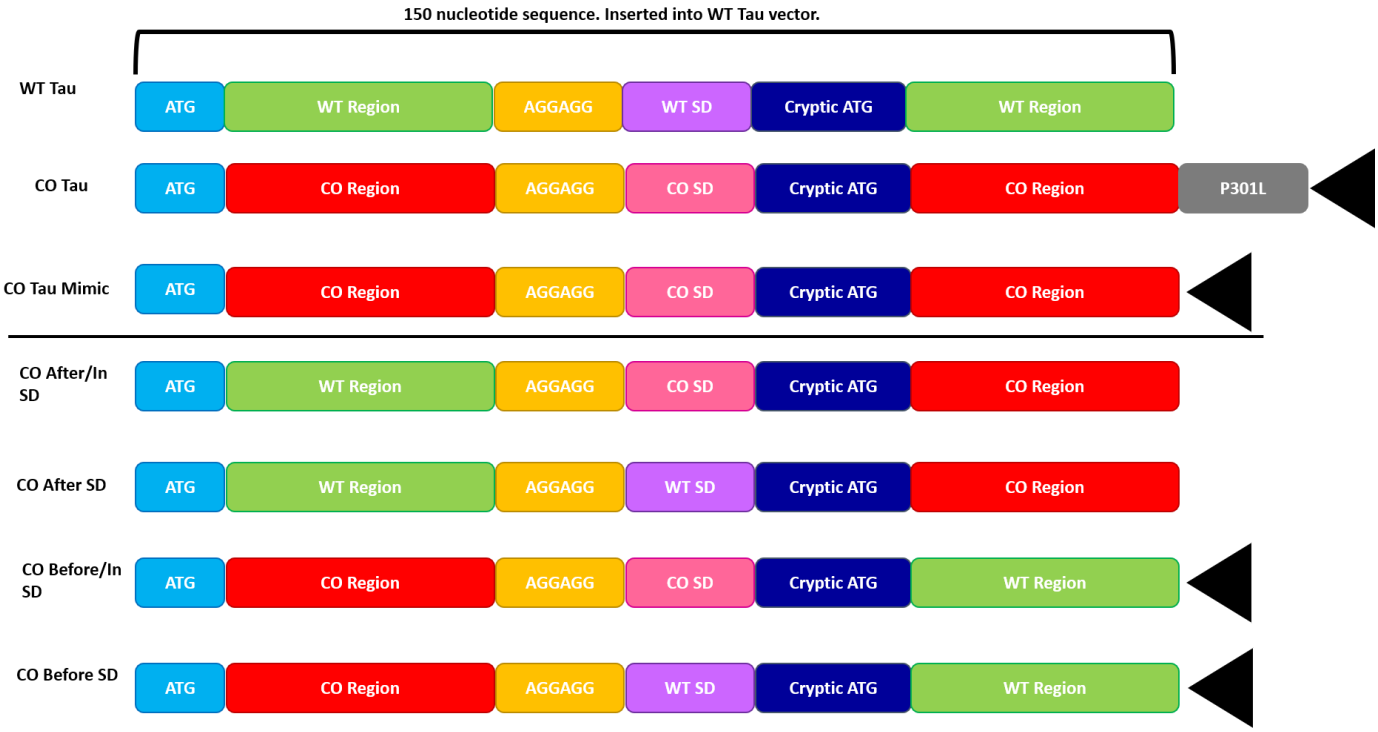

Figure 9. 150 Nucleotide G Block Designs. Wild Type Tau, CO Tau P301L, and CO Tau Mimic were all included as control constructs (above the black line). The four constructs below the black line are the experimental G Blocks introducing regional changes. Black arrows indicate truncation observed by Coomassie and Western Blots (Figure 12).

The experimental strategy (Figure 10) developed for insertion of the 150 nucleotide G Blocks into the Wild Type Tau vector utilized vector digestion by two restriction enzymes, ligation, and transformation. A sample of the digested Wild Type Tau vector was reserved as a negative control. The control was taken through the process of insertion and transformation into Top10 cells but did not receive a G block. This was done to ensure that the vector was successfully treated with phosphatase and that the ends of the vector did not re-ligate. No colonies were observed after this negative control transformation, indicating successful phosphatase treatment of the vector (Figure 11D-a). All G block plasmids were successfully isolated and resulted in colony growth following transformation into BL21 DE3 cells (Figure 11D-b).

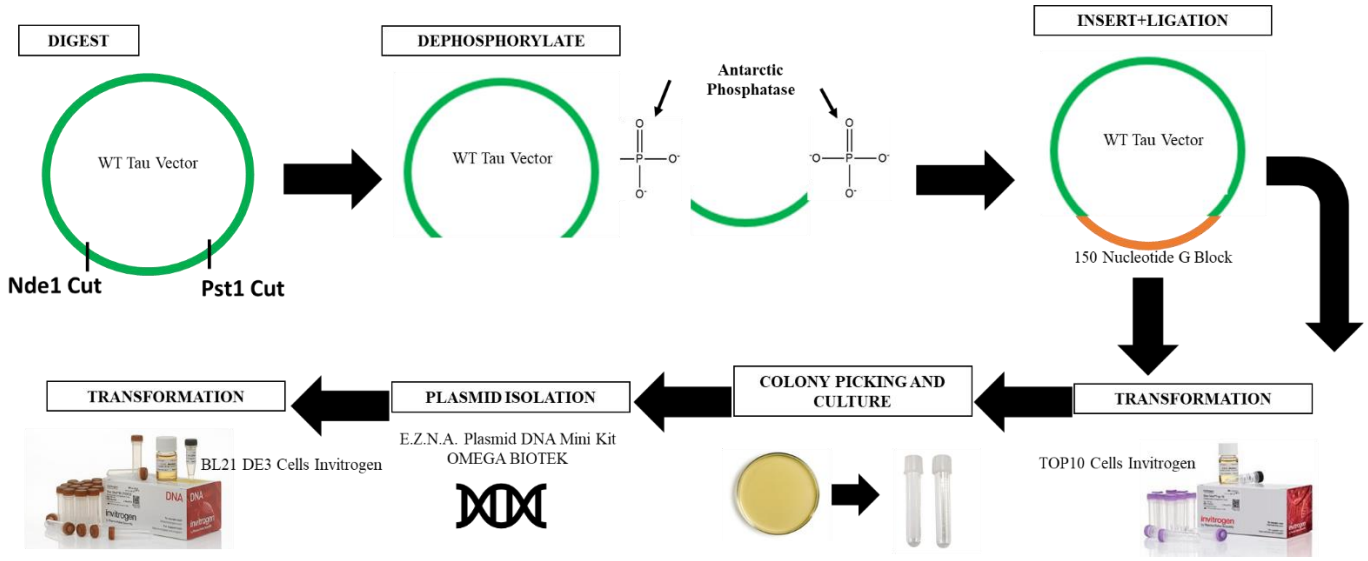

Figure 10. Restriction Enzyme Method Schematic for Insertion of G Blocks.

a)

b)

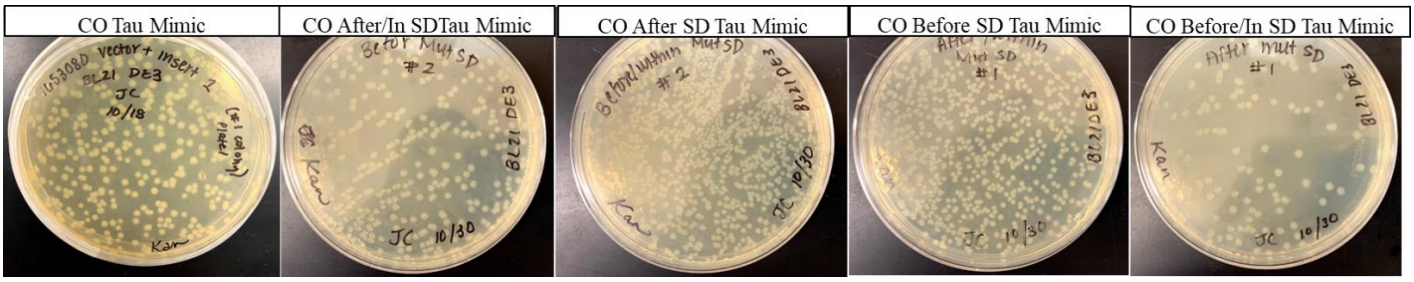

Figure 11D. Control and Colony Growth After Restriction Enzyme Digestion, G Block Insertion, and Ligation. a) *Negative Control.* Wild Type Tau vector plasmid digested, dephosphorylated, and ligated without inserted DNA. Cut portion was not re-inserted into vector and vector ends were not ligated together, so vector remained linear and thus no colonies grew. b) *G Block Insert Mutants*. Digested and dephosphorylated vector was mixed with respective 150 nucleotide G blocks, then ligated and transformed into Top10 Cells. Colonies from Top10 transformation were picked and cultured. Plasmids were isolated from their culture, which were then transformed into BL21 DE3 cells. All mutants yielded sufficient colonies, indicating successful insertion and ligation of G blocks.

G Block Insert Mutant Results

The G block transformed colonies were cultured and partially purified via the Barghorn Boiling protocol. After purification, samples of each G block lysate were analyzed by Coomassie Staining and Western Blotting with Tau monoclonal antibody. Wild Type Tau and CO Tau P301L were both included as controls on the gel; Wild Type being a negative control expected not to show truncation and CO Tau P301L as a positive control showing truncation. The CO Tau Mimic was included as a secondary positive control and was expected to truncate as well. A sample of Fixed CO Tau was also included as a negative control showing that the removal of the cryptic start site disables truncation.

As can be seen in Figure 12 in both the Coomassie gel and the Western Blot, the positive controls CO Tau P301L and CO Tau Mimic both truncated as expected. The CO After SD Tau Mutant and CO After/In SD Tau Mutant both showed no signs of truncation by either analysis method. The CO Before SD Tau Mutant and CO Before/In SD Tau Mutant both did show truncation, however. In the Coomassie gel, the truncation was more apparent in CO Before SD Tau Mutant that the Before/In Mutant. But in the Western Blot, the truncation of both constructs is relatively the same from visual analysis. The fact that the constructs that truncated included nucleotides from the originally truncating construct (CO Tau P301L) upstream of the cryptic start site suggests that ribosomal binding and initiation of translation at the cryptic start site is more dependent on upstream nucleotide combinations than downstream nucleotide composition. This suggests the upstream region seems to have more of an effect on the strength of the SD sequence to initiate translation than the downstream region.

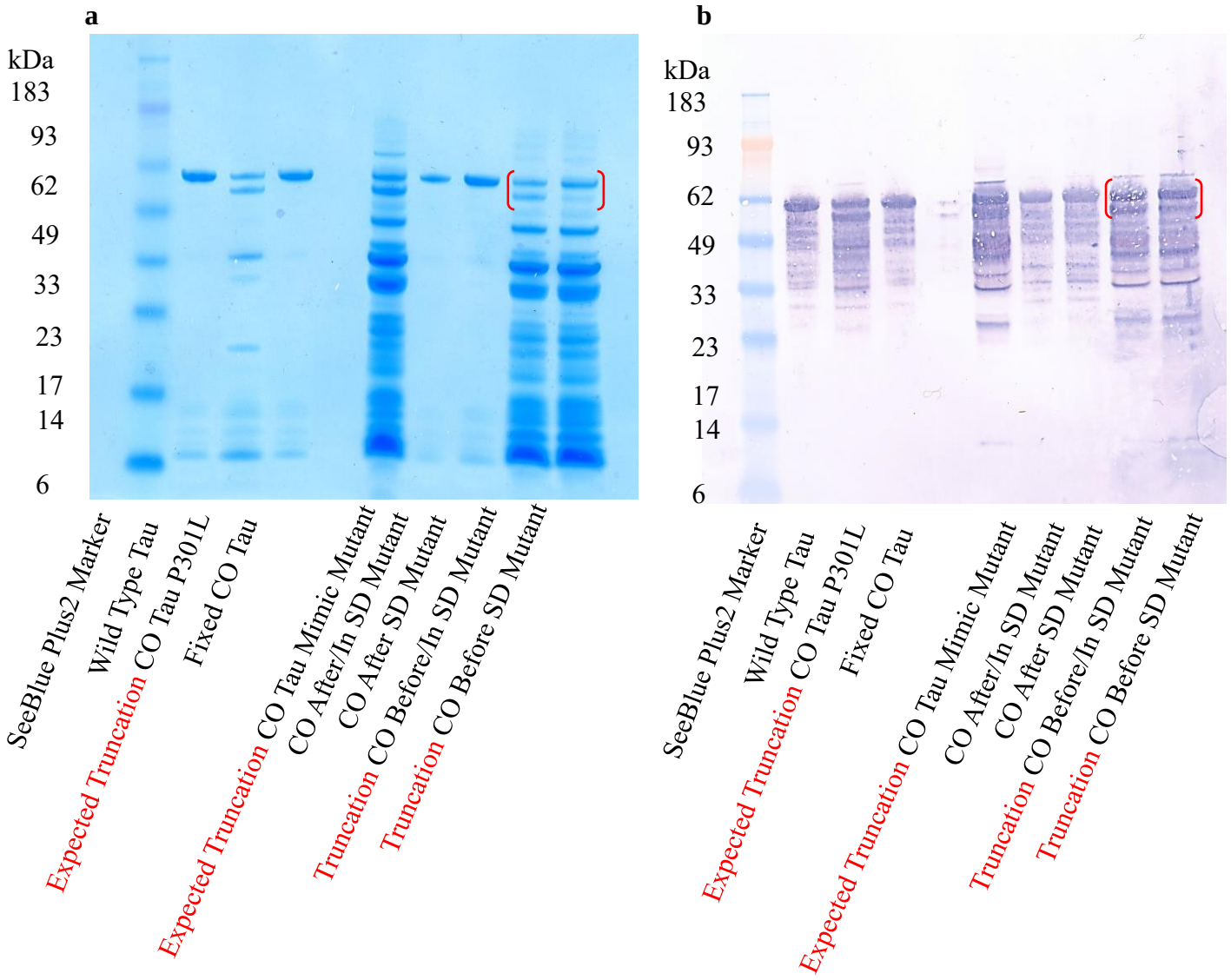

Figure 12. G Block Insert Mutant Results

Coomassie Stained Gel (a) and Western Blot (b). Truncation seen in CO Tau P301L as expected. Truncation observed in Before and Before/In SD Mutants (red brackets). All truncating constructs contain CO Tau sequence upstream of the SD site.

Conclusions

The presented research provides evidence that the initiation of translation at Met31 in CO Tau P301L is dependent on the sequence upstream of the Shine-Dalgarno AGGAGG region. Nucleotide changes within the AGGAGG and Met31 did not initiate translation at Met31 in Wild Type Tau, indicating that factors beyond the immediate Shine-Dalgarno region effected truncation. Nucleotide alterations downstream of Met31 did not appear to have any effect on the initiation of translation at Met31, as truncation was not observed under such circumstances. The presented research further provides evidence that the P301L mutation is not a causal factor of truncation. The inclusion of the P301L mutation in Wild Type Tau in Strategy 1 did not result in truncation, nor did the omission of P301L in Strategy 3 prevent truncation.

Though this project does provide evidence suggesting that the sequence upstream of the SD region effects initiation of translation, it does not provide exact locations and nucleotide identities that contribute to the phenomena of truncation. To further the presented research, the 74 nucleotides between the primary start codon and the AGGAGG Shine-Dalgarno sequence could be varied on a nucleotide-by-nucleotide basis without altering the amino acid sequence of CO Tau P301L. Identifying specific locations/nucleotides that effect the ability of translation to occur at the cryptic start site would further the understanding of structural and sequential factors that strengthen the SD sequence. Future research could also include relative quantitative analysis by enhanced chemiluminescence blotting. This method would allow blotted proteins to be compared quantitatively by comparison of intensity. Analysis by mass spectrometry could also be used to investigate if different nucleotide combinations effect the extent of truncation by quantifying the amount of full length Tau compared to truncated protein.

The presented research was purely investigational, as the truncation problem observed in CO Tau P301L was solved by removal of the cryptic Met31 in Fixed CO Tau. However, the research conducted in this project offers a deeper understanding of Shine-Dalgarno sequences as well as how sequence differences can arise through codon optimization. This understanding has the potential to influence construct design in future endeavors.

References

Alzheimer's Association. (2019). Alzheimer's Disease Facts and Figures. *Alzheimer's Dement 2019*. 15(3):321-87.

Barghorn, S., Biernat, J., and Mandelkow, E. (2005). Purification of Recombinant Tau Protein and Preparation of Alzheimer-Paired Helical Filaments in Vitro. *Methods in Molecular Biology.* 299: 35-51.

Berriman, J., Serpell, L.C., Oberg, K.A., Fink, A.L., Goedert, M., Crowther, R.A. (2003). Tau filaments from human brain and from *in vitro* assembly of recombinant protein show cross-beta structure. *Proceedings of the National Academy of Sciences of the United States of America.* 100(15): 9034-9038.

Brunden, K.R., Trojanowski, J.Q., Lee, V.M.Y. (2009). Advances in tau-focused drug discovery for Alzheimer's disease and related tauopathies. *Nature Reviews Drug Discovery.* 8:783-793.

Drachman, D.A. (2014). The amyloid hypothesis, time to move on: Amyloid is the downstream result, not cause, of Alzheimer's disease. *Alzheimer's and Dementia*. 10(3): 372-380.

Giustiniani, J. et al. (2014). Immunophilin FKBP52 induces Tau-P301L filamentous assembly in vitro and modulates its activity in a model of tauopathy. Proceedings of the National Academy of Sciences of the United States of America. 111(12): 4584-4589

Goedert, M. 1993.Tau protein and the neurofibrillary pathology of Alzheimer's disease. *Trends in Neuroscience*. 16(11): 460-465.

Goedert, M., Eisenberg, D.S., Crowther, R.A. 2017. Propagation of Tau Aggregates and Neurodegeneration. *The Annual Review of Neuroscience*. 40: 189-210.

Goedert, M. and Jakes, R. (1990). Expression of separate isoforms of human tau protein: correlation with the tau patterns in brain and effects on tubulin polymerization. *The EMBO Journal*. 9(13): 4225-4230.

KrishnaKumar, V.G. and Gupta, S. (2017). Simplified method to obtain enhanced expression of tau protein from E. coli and one-step purification by direct boiling. *Preparative Biochemistry and Biotechnology*. 47:5, 530-538

Li, J., Liang, Q., Song, W., and Marchisio, M.A. (2017). Nucleotides upstream of the Kozak sequence strongly influence gene expression in the yeast S. cerevisiae. *Journal of Biological Engineering*. 11(25).

Li, G-W., Oh, E., and Weissman, J.S. (2012). The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*. 484: 538-543.

Malys, N. (2012). Shine-Dalgarno sequence of bacteriophage T4: GAGG prevails in early genes. *Molecular Biology Reports*. 39(1): 33-39.

Mauro, V.P. and Chappell, S.A. (2014). A critical analysis of codon optimization in human therapeutics. *Trends in Molecular Medicine*. 20(11): 604-613.

Paranjape, S.R., Riley, A.P., Somoza, A.D., Oakley, C.E., Wang, C., Prisinzano, T.E., Oakley, B.R., and Gamblin, T.C.(2015). Azaphilones inhibit tau aggregation and dissolve tau aggregates in vitro. *American Chemical Society: Chemical Neuroscience*. 6: 751-760.

Slobodskaya, O.R., Gmyl, A.P., Maslova, S.V., Tolskaya, E.A., Viktorova, E.G., Agol, V.I. (1996). Poliovirus Neurovirulence Correlates with the Presence of a Cryptic AUG Upstream of the Initiator Codon. *Virology*. 221(1): 141-150.

Whitaker, W.R., Lee, H., Arkin, P.A., Dueber, J.E. (2014). Avoidance of truncated proteins from unintended ribosome binding sites within heterologous protein coding sequences. *ACS Synthetic Biology*. 4: 249-257.

Woerman, A.L., Aoyagi, A., Patel, S., Kazmi, S.A., Loback, I., Grinberg, L.T., McKee, A.C., Seeley, W.W., Olson, S.H., Prusiner, S. B. (2016). Tau proteins from Alzheimer's disease and chronic traumatic encephalopathy patients propagate in cultured cells. *Proceedings of the National Academy of Sciences of the United States of America.* 113(50): E8187-E8196.

Appendices

Appendix A. Summary of designed DNA constructs and primers. All DNA constructs designed with Geneious Prime™ Software. All DNA primers designed with Agilent Primer Design Software.

Strategy	Construct Name	Description	Truncation?
All	WT Tau	Original Tau construct, not codon-optimized or mutated in any way	No
All	CO Tau P301L	Codon-optimized Tau construct with single amino acid mutation at position 301 changing proline to leucine	Yes
All	Fixed CO Tau	Codon-optimized Tau construct with single amino acid mutation at position 301 changing proline to leucine, but with cryptic Met31 removed	No
1. Mutations within 7 nucleotide sequence upstream of ATG	WT-CO Tau	WT Tau sequence with single amino acid mutation at position 301 changing proline to leucine to prove amino acid change is not responsible for truncation	No
1	A84C-A90T	WT Tau sequence with A to C change at nucleotide 84 and A to T change at nucleotide 90	No
1	T87C A90T	WT Tau sequence with T to C change at nucleotide 87 and A to T change at nucleotide 90	No
1	A84C T87C A90T	WT Tau sequence with A to C change at nucleotide 84, T to C change at nucleotide 87, and A to T change at nucleotide 90	No
1	A84T	WT Tau sequence with A to C change at nucleotide 84	No
1	T87C	WT Tau sequence with T to C change at nucleotide 87	No
1	A90T	WT Tau sequence with A to T change at nucleotide 90	No

1	A84C T87C	WT Tau sequence with A to C change at nucleotide 84 and T to C change at nucleotide 87	No
2. Site Directed Mutagenesis	Perfect	Primer designed to match 16s rRNA as closely as possible. Changed 3 nucleotides within nucleotides 76-93. A to T change at nucleotide 84, T to G change at nucleotide 85, and A to C change at nucleotide 88.	N/A
2	Perfect Destroyed	Primer created to remove the AGGAGG of the Shine Dalgarno region but maintain similarity to the 16s rRNA as the "Perfect" primer. The AGGAGG was to be changed to a GGGGGG, introducing 5 total nucleotide changes within the 18 nucleotide region to remove the SD sequence. A to G change at nucleotide 78, A to G change at nucleotide 81, A to T change at position 84, T to G change at nucleotide 85, and A to C change at nucleotide 88.	N/A
2	Met-Val	Primer designed to change the cryptic start site from ATG to GTG, but otherwise maintain similarity to 16s rRNA as the "Perfect" primer. A to T change at nucleotide 84, a T to G change at nucleotide 85, an A to C change at nucleotide 88, A to G change at nucleotide 91.	N/A
2	Met-Leu	Primer designed to change the cryptic start site from ATG to TTG, but otherwise maintain similarity to the 16s rRNA as the "Perfect" primer. A to T change at position 84, T to G change at nucleotide 85, A to	N/A

		C change at nucleotide 88, A to T change at nucleotide 91.	
3. 150 Nucleotide G Block Insertion	CO Tau Mimic	First 150 nucleotides of CO Tau inserted into WT Tau sequence	Yes
3	CO After SD Tau	150 nucleotides total. WT sequence up until AGGAGG of SD region, remainder of 150 nucleotides are CO Tau sequence. Inserted into WT Tau vector.	No
3	CO After/In SD Tau	150 nucleotides total. WT sequence up until ATG cryptic start site of SD region, remainder of 150 nucleotides are CO Tau sequence. Inserted into WT Tau vector.	No
3	CO Before SD Tau	150 nucleotides total. CO sequence up until AGGAGG of SD region, remainder of 150 nucleotides are WT Tau sequence. Inserted into WT Tau vector.	Yes
3	CO Before/In SD Tau	150 nucleotides total. CO sequence up until ATG cryptic start site of SD region, remainder of 150 nucleotides are WT Tau sequence. Inserted into WT Tau vector.	Yes