5-2020

Recombinant Expression of Preptin Analogs for Alanine Scanning Mutagenesis of Residues 27-29

Tehgan N. Anguilm

COLUMBUS STATE UNIVERSITY

RECOMBINANT EXPRESSION OF PREPTIN ANALOGS FOR ALANINE SCANNING MUTAGENESIS OF RESIDUES 27-29

A THESIS SUBMITTED TO THE

HONORS COLLEGE

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR HONORS IN THE DEGREE OF

BACHELOR OF ARTS

DEPARTMENT OF CHEMISTRY

COLLEGE OF LETTERS AND SCIENCES

BY

TEHGAN N. ANGUILM

COLUMBUS, GEORGIA

2020

RECOMBINANT EXPRESSION OF PREPTIN ANALOGS FOR ALANINE SCANNING MUTAGENESIS OF RESIDUES 27-29

By

Tehgan N. Anguilm

A Thesis Submitted to the

HONORS COLLEGE

In Partial Fulfillment of the Requirements

for Honors in the Degree of

BACHELOR OF ARTS

CHEMISTRY

COLLEGE OF LETTERS & SCIENCES

Approved by

Dr. Jonathan Meyers, Committee Chair

Dr. Daniel Holley, Committee Member

Dr. Cindy Ticknor, Committee Member & Dean

ABSTRACT

The novel peptide preptin consists of 34 amino acid residues corresponding to Asp^{69}-Leu^{102} of pro-insulin-like growth factor E-peptide. Preptin has been isolated from pancreatic beta cells, and it has been found to exhibit both metabolic and mitogenic activities. This makes preptin an attractive candidate to treat insulin-independent diabetes and osteoporosis. Alanine scanning mutagenesis was utilized to substitute an alanine for residues Trp^{27}, Arg^{28}, and Gln^{29} of preptin in order to ultimately study the structure-activity relationship of these residues to preptin's metabolic activity.

ACKNOWLEDGEMENTS

This work is dedicated to my research mentor Dr. Jonathan Meyers and my research partner Ansley Felix, both of whom inspire me and encourage me in my current and future endeavors. I would like to thank the Honors College and the CSU Chemistry department for all of the great experiences and lasting memories from my undergraduate career.

TABLE OF CONTENTS

I. Introduction and Background Information 1
A. Diabetes and Cancer 1
B. Preptin 2
C. Alanine Scanning Mutagenesis 8
D. Cloning 8
E. Inverse Polymerase Chain Reaction Mutagenesis 9
F. Primer Design 13
G. Gel Electrophoresis 15
II. Methods and Materials 16
A. Plasmid Design 16
1. Rat Preptin Ala^{21} Gene 16
2. His^{6} Affinity Tag 16
3. GFP fusion protein 16
4. TEV Protease recognition site 16
5. Ampicillin Selection Marker 16
6. LAC Operon Inducible Promoter 16
B. Primer Design 18
C. Selection 19
D. Mutagenesis 20
E. Overnight Culture for Plasmid Isolation 23
F. Plasmid Isolation 23
G. Gel Electrophoresis 24
H. DNA Sequencing 24
III. Results and Discussion 25
A. Rationale 25
B. Results 25
C. Future Work 30
IV. References 32
V. Appendix 38
A. Plasmid Map and Sequence for pD444-SR:His6-GFP-TEV-Preptin Ala21 38

I. Introduction and Background Information

A. Diabetes and Cancer

The United States is experiencing an epidemic of chronic disease. Six in ten Americans have a chronic disease and four in ten have two or more.[1] On a global level, chronic diseases are the leading cause of death and disability.[2] The most prevalent forms include diabetes, cancer, heart disease, stroke, and chronic respiratory diseases.[2] In 2017, diabetes alone accounted for $327 billion in total healthcare costs, and every year, approximately 1.5 million Americans are diagnosed with the disease.[3] In 2015, diabetes was the seventh leading cause of death in the U.S. It was written as the cause of death for 79,535 individuals and listed as a contributing cause of death for an additional 252,806 deaths.[3] Individuals with diabetes are at an increased risk for a spectrum of complications, including hypertension, stroke, gastroparesis, and neuropathy.[4]

Diabetes mellitus refers to a group of related metabolic disorders revolving around deficiency or inefficiency in the insulin signaling system. Type 1 and type 2 diabetes are the most common forms of diabetes. Type 1 diabetes is an autoimmune disease in which the immune system destroys the insulin-secreting beta cells in the pancreas. The loss of beta cells inhibits the body's ability to secrete insulin; this type of diabetes accounts for 5-10% of diabetes cases.[5] Type 2 diabetes is the more prevalent form and affects the body's ability to metabolize glucose. In the early stages of type 2 diabetes, the pancreas often secretes insulin at levels higher than normal in order to maintain blood glucose levels. However, as the body continues to resist the effects of insulin, the pancreas loses its ability to secrete insulin properly, leading to type 2 diabetes.[6]

In addition to the direct effects of diabetes, there is epidemiological evidence to suggest that individuals with diabetes have a 20-25% increased risk of cancer, depending on the type,

compared to non-diabetic individuals.[7,8] It is also suggested that mortality in cancer patients increases significantly if the patient is also diabetic.[7] Like diabetes, cancer manifests from irregular cell signaling. Genetic and epigenetic alterations allow cells to bypass homeostatic mechanisms that control survival and migration in the body, leading to the overproliferation of cells.[9]

Many risk factors coincide for diabetes and cancer including age, sex, weight, physical activity, nutrition, alcohol, and smoking.[7,10] However, the link between these two diseases is not fully understood. Several hormones that participate in both metabolic and mitogenic pathways have been identified. The metabolic pathways are key in energy regulation and associated with diabetes, while the mitogenic pathways are key in cell growth and associated with cancer. Due to the prevalence of chronic disease worldwide, it is necessary to explore and understand these hormones, their structures, and their pathways. These hormones include but are not limited to insulin, insulin-like growth factors (IGF), and the recently discovered preptin.

B. Preptin

Insulin-like growth factors are peptide hormones that function to stimulate growth but also can decrease the levels of blood glucose.[11] The two types of insulin-like growth factors are IGF-I and IGF-II. The translation product of the insulin-like growth factor II gene is pre-pro-IGF-II, which is 180 amino acids long and organized into five domains (A through E). Pre-pro-IGF-II contains a 24 amino acid long N-terminal signal peptide, a 67 amino acid long mature protein (IGF-II), and an 89 amino acid long C-terminal E domain.[12,13] Proteolytic cleavage, the process of breaking peptide bonds between amino acid residues in protein, of the signal peptide forms pro-IGF-II. Subsequent proteolysis of pro-IGF-II results in the cleavage of mature IGF-II protein and the E-domain.[13] Preptin is the 34 amino acid peptide corresponding to Asp^{69}-Leu^{102} of pro-IGF-II E-peptide.[14]

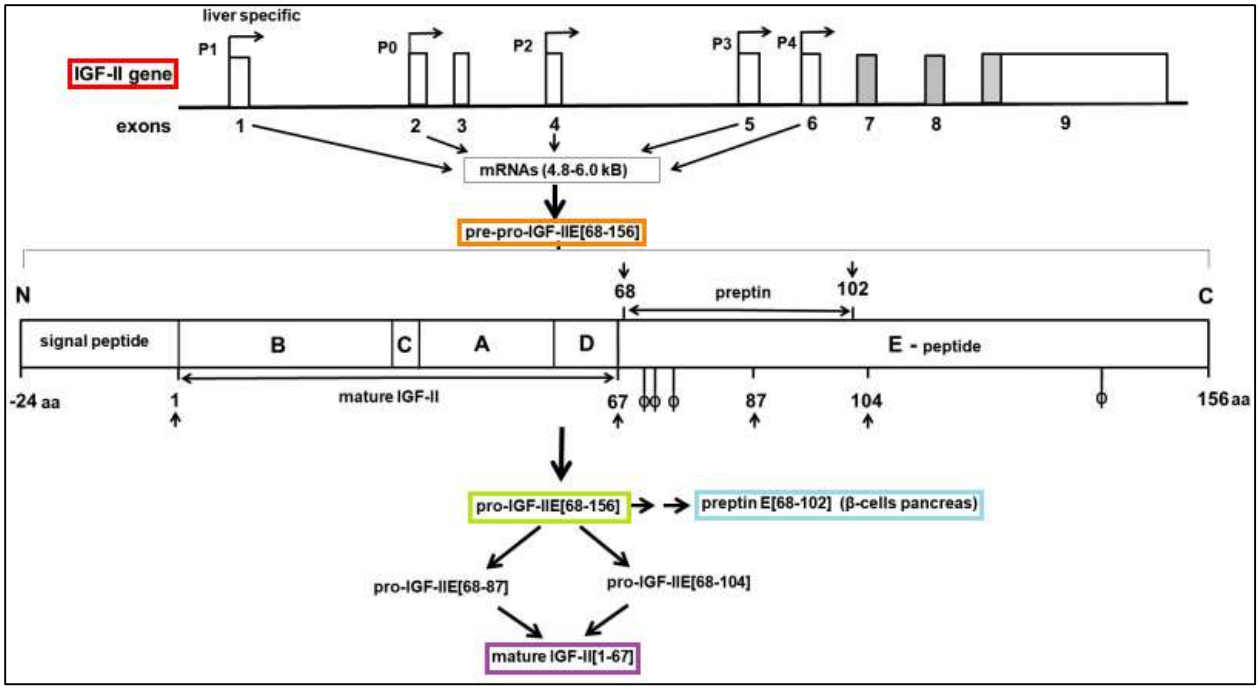


Figure 1. Schematic representation of the IGF-II gene (red), which is translated to pre-pro-IGF-II (orange). Pre-pro-IGF is proteolytically cleaved into pro-IGF-II (green). Subsequent proteolysis yields mature IGF-II (purple) and the E-domain. Preptin (blue) is the 34 amino acid peptide of the E-domain.[13]

In 2001, preptin was isolated from βTC6-F7 cells in rats by Buchanan et al., and its sequence has been greatly conserved throughout evolution in both humans and mice (Figure 3).[14,15] Human and mouse preptin are 79.41% similar, human and rat preptin are 75.53% similar, and rat and mouse preptin are 94.12% similar.[15]

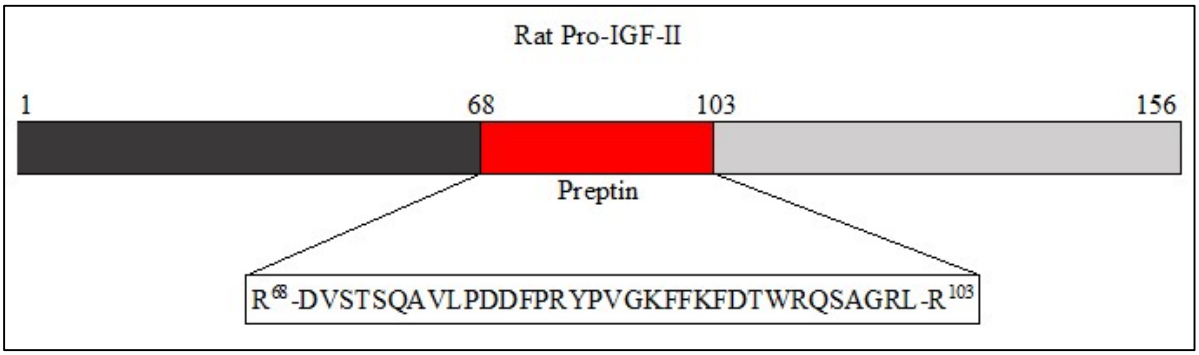


Figure 2. Preptin cleavage product of rat pro-insulin-like growth factor II between residues 68 and 103.

Comparative amino acid sequences among human, rat, and mouse preptin	
Human	R^{68}-DVSTPPTVLPDNFPRYPVGKFFQYDTWKQSTQRL-RR^{104}
Rat	R^{68}-DVST**SQA**VLPD**D**FPRYPVGKFF**KF**DTW**R**QS**AG**RL-RR^{104}
Mouse	R^{68}-DVST**SQA**VLPD**D**FPRYPVGKFFQYDTW**R**QS**AG**RL-RR^{104}

Figure 3. Preptin sequence among humans, rats, and mice with differences indicated in bold.

The main sources of preptin are cells of the kidney, liver, pancreas, salivary gland, and mammary tissue.[15] The physiological and biochemical functions of preptin include cell differentiation, cell proliferation, bone density regulation, metabolic adaptation, and homeostasis regulation.[15] Its ability to affect cell differentiation and proliferation relate to its mitogenic activity in bone. Preptin also affects homeostasis regulation as it pertains to glucose metabolism. Preptin and insulin are co-secreted from pancreatic beta cells in response to glucose, and they help shuttle glucose into the muscles and tissues for energy use. Preptin behaves in a concentration dependent manner, like insulin.[14,16] If more glucose is present, there will be a greater secretion of both preptin and insulin to balance the levels of glucose in the blood within the narrow range required for homeostasis.

Preptin's metabolic properties can be seen with its similar behavior to the FDA-approved drug glibenclamide. Glibenclamide belongs to the widely-used class of drugs known as sulfonylureas, and it is used as a treatment for type 2 diabetes. Individuals with type 2 diabetes have difficulty metabolizing glucose because their bodies begin to resist the effects of insulin, and the pancreas loses its ability to secrete insulin properly. Glibenclamide helps with this by stimulating pancreatic beta cells to secrete insulin, and it can do this in the presence or absence of a glucose challenge by "blocking ATP-sensitive potassium channels in pancreatic beta cells."[15,16,18] These ATP-sensitive potassium channels are also shared by the cardiovascular

system. During ischemia, the restriction of blood supply to the myocardial tissue, these channels open because the levels of ATP in the cytosol decrease.[18] There is a protective mechanism for the cardiovascular system called "ischemic preconditioning," which utilizes a brief period of ischemia, and it aims to protect against prolonged ischemic injury to the heart.[19] Sulfonylureas, such as glibenclamide and tolbutamide, have been found to block the opening of the ATP-sensitive potassium channels in myocardial tissue and vascular smooth muscle cells, which could lead to interference of the cardioprotective mechanism of ischemic preconditioning.[20] Furthermore, because glibenclamide can stimulate insulin secretion without the presence of glucose, it can lead to severe hypoglycemia, which is a state of low blood glucose that can be characterized by seizures and loss of consciousness.[18,21] This hypoglycemic state is the most likely culprit in the adverse consequences of glibenclamide and other sulfonylureas on the cardiovascular system.[18] The risks of glibenclamide, along with other diabetic drugs, drives the search for other potential treatments, one of which is preptin. In rats, preptin stimulates insulin secretion in similar concentrations to glibenclamide *in vitro*, although preptin's pathway is not fully understood.[15] Unlike glibenclamide, however, preptin is self-regulating, meaning that when glucose levels are not elevated, the body's natural regulatory system resists preptin's action and prevents hypoglycemia.

In addition to preptin's potential as a treatment for diabetes, it is also being considered as a treatment for osteoporosis because of its osteogenic nature *in vivo* and *in vitro*.[22] Osteoporosis is characterized by an imbalance in bone metabolism, the cycle of growth and resorption of bone guided by osteoblasts, osteoclasts, hormonal and regulatory signals.[23] If resorption of bone is greater than the growth of bone, osteoporosis can develop.[23] The current therapies for preventing osteoporosis target bone resorption but have limited ability to improve bone mass.[22] This makes preptin a good candidate for treatment of osteoporosis, as it stimulates bone formation and can

positively impact bone mass.[22] Similar to insulin and other products of the pancreatic beta cell, preptin encourages bone anabolism via the proliferation and survival of osteoblasts.[22] However, preptin does not impact osteoclast activity and the removal of bone tissue.[22]. Homeostasis of bone is mediated by both osteoblast and osteoclast activity, so preptin's ability to affect only the bone-forming side of this equation could lead to uncontrolled cell growth, manifesting diagnostically as cancer. Preptin is also expressed alongside insulin-like growth factor II, as they are both cleavage products of the pro-insulin-like growth factor II. Upon binding of IGF-II to its receptor, mitogenic (cell growth) pathways are initiated.[24] For preptin to become a useful drug, its metabolic and osteogenic activities must be separated to avoid a cancerous state.

Kowalczyk et al. previously established that the first 16 amino acid residues of preptin's sequence stimulated bone growth and increased the survival of osteoblasts in rats.[25] This fragment was not active in the pancreas nor did it affect glucose metabolism.[25] However, because this peptide fragment was still relatively large for a therapeutic, this group performed truncations to determine the smallest fragment that performed the same function of bone anabolism.[25] A short eight amino acid fragment of preptin (1-8) was identified as critical to the proliferation of osteoblasts and the enhancement of bone formation in rats.[25] The work by Kowalczyk et al. was done to elucidate preptin's mitogenic activity in bone, but no such work has been done to identify fragments necessary for its metabolic activity.

Because research has not been done to investigate preptin's metabolic activity, this work aimed to address that disparity. A review of the C-terminal end of the rat preptin sequence revealed a three-residue region with interesting properties. Residues Trp[27] (tryptophan), Arg[28] (arginine), and Gln[29] (glutamine) each have long side-chain when compared to other amino acids (Figure 4). The proximity of these long side-chains coupled with their properties suggests that this region of the

peptide may interact with the target receptor, which is currently unknown. One chemical property exhibited tryptophan is aromaticity. Aromatic rings possess stability from the pi-electrons above and below the plane of the ring; these aromatic rings have been demonstrated to participate in pi-stacking with other side chains in protein-protein interactions.[26] In addition, arginine exhibits the physical property of electric charge. Arginine is a positively charged amino acid, and repetitive arginine residues have been shown to be a key recognition site in protein-protein interactions. Taken together, this region presented a good place to start investigating the structure-activity-relationship of these three residues of preptin. The goal of this work was to analyze amino acid residues Trp[27], Arg[28], and Gln[29] of rat preptin using alanine scanning mutagenesis to better understand the structure-activity relationship of these residues to the peptide.

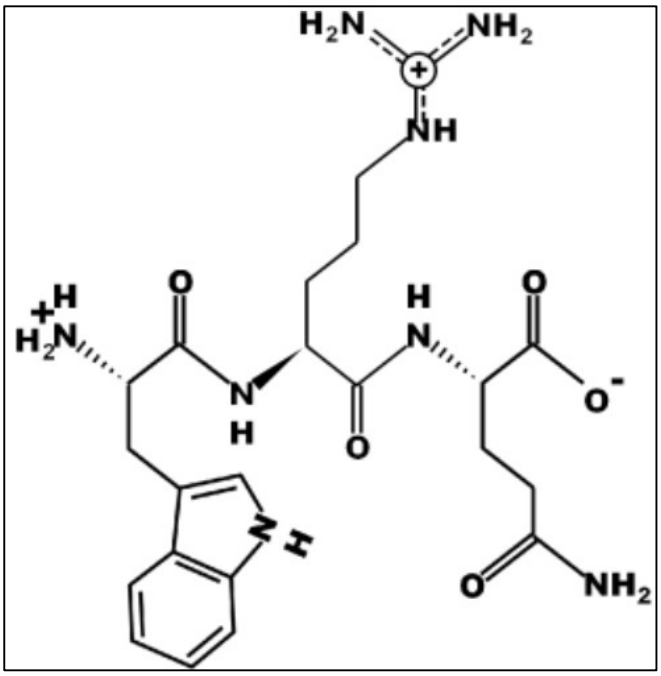


Figure 4. Amino acid residues Trp[27], Arg[28], and Gln[29] of preptin's 34 amino acid long sequence.

C. Alanine Scanning Mutagenesis

Site-directed mutagenesis is used to create specific changes to double-stranded plasmid DNA. This is often done to screen for mutations, introduce or remove endonuclease sites, or change the sequence of a gene product.[27] Alanine-scanning mutagenesis is the commonly used form of site-directed mutagenesis, and it alters the DNA sequence to encode an alanine in place of the naturally occurring amino acid of interest.[28] Substitution of an alanine into the polypeptide gene product does not alter the "main-chain conformation" like glycine or proline and keeps the native protein intact.[28] It also does not inflict electrostatic or steric effects because it is uncharged and its side chain is a simple methyl group.[28] However, alanine substitution does alter the side-chain profile of the polypeptide by essentially removing the native side chain. If an altered polypeptide shows reduced activity when the native side-chain is removed, that is an indication that the native side-chain participates in the active or binding site of the polypeptide.

D. Cloning

Cloning is a technique of making identical copies of a piece of DNA. A cloning vector is a piece of DNA into which a desired DNA fragment can be inserted. In traditional cloning methods, vectors are based on bacterial plasmids; these plasmids have DNA that is circular and double-stranded, and they replicate independently of genomic DNA.[29] Cloning vectors that are based on plasmids have various elements including an origin of replication, restriction enzyme sites or multiple cloning site (MCS), and markers that help determine successful uptake of the desired vector.[29]

There are four steps to the traditional cloning procedure: use restriction enzymes to select a desired gene from an organism, insert the desired gene into a vector, transform the vector into

bacteria, and isolate the vector DNA for purification.[30] Because DNA is selected from one organism and transplanted into a vector, cloning can also be referred to as recombinant DNA technology. It can be used to determine the function of a gene, determine how mutations affect the function of a gene, and make large quantities of protein encoded by the gene.[30]

E. Inverse Polymerase Chain Reaction Mutagenesis

Polymerase chain reaction (PCR) cloning differs from the traditional cloning procedure because the vector and the DNA fragment of interest can be amplified and ligated together without restriction enzymes.[31] The polymerase chain reaction also allows for the exponential amplification of small quantities of DNA. The process involves three steps: initial denaturation, thermocycling, and final extension of DNA. During the initial denaturation, the double-stranded DNA is subjected to a high temperature, typically between 94 and 98°C, and incubated for a longer time to guarantee that the double-stranded DNA is fully separated prior to the thermocycling.[32,33] The time must be long enough to ensure the denaturation of the strands for priming but not long enough to damage the DNA.[30] The high temperature disrupts the hydrogen bonds between nucleotide bases of the double-stranded DNA to separate it into single strands.[33,34] Initial denaturation helps to ensure that optimal amplification occurs during the first amplification cycle. Thermocycling entails the repetition of another denaturation step along with annealing and extension and is typically carried out for 25 to 35 cycles, although this is determined by the amount of DNA to be amplified and the desired yield.[33,35] PCR primers are short segments of single-stranded DNA that bind to their complementary sequences on single-stranded DNA during the annealing step. The primer-template heteroduplex forms, acting as an initiation complex for the DNA polymerase to perform extension.[33,34] Extension is the last step in thermocycling, and this is when DNA *Taq* polymerase,

the most commonly used polymerase in PCR, extends the primer sequences at the 3' end to the end of the amplicon.[33] This step is important to amplification because of its specificity: amplification will only occur to molecules with sequences complementary to the primers.[31] The ideal temperature for extension is 72$^{\circ C}$, as that is the optimal temperature for the DNA Taq polymerase.[34] The final extension step occurs after the last cycle and DNA polymerase extends the primer sequences.[33] Its duration depends on the length of the amplicon and must be long enough to ensure complete polymerization and adequate yield of the DNA.[34]

The traditional PCR technique used for site-directed mutagenesis utilizes primers that flank the target gene or fragment that is to be copied. The primers are designed with the desired mutation, and during the PCR, the mutation is incorporated into the gene or fragment. However, this method is limited in amplification to the sequence between two inward-pointing primers, and extension cannot exceed the segments upstream and downstream from the region of interest.[36,37] This method also requires additional steps to ligate the amplified DNA sequence into a linear plasmid.

On the other hand, inverse PCR mutagenesis is a faster and more efficient method of mutagenesis when compared to the traditional mutagenesis protocol. Inverse PCR mutagenesis differs because it uses back-to-back non-overlapping primers (Figure 5) to amplify the entire plasmid instead of just a target segment. The forward primer has the desired mutation, and the reverse primer allows them to anneal back-to-back. The segment of DNA containing the mutation and the vector itself can be amplified and ligated together without restriction endonucleases.[31]

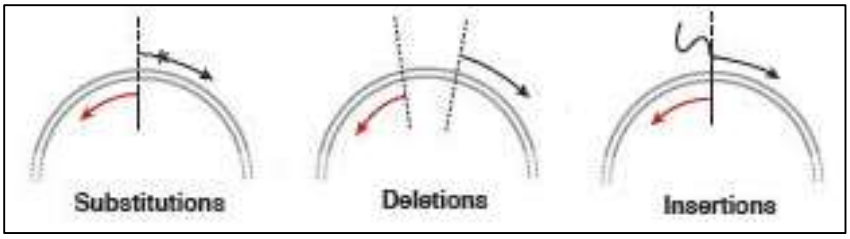


Figure 5. Back-to-back primers are indicated by the arrows, and they are efficient at incorporating substitutions, deletions, and insertions into double-stranded plasmid DNA.[38]

Inverse PCR mutagenesis produces linear plasmid products that require circularization. This is accomplished by incubation with the enzymes kinase and ligase. Kinase phosphorylates the 5' ends of the linear PCR products so they can be circularized by the enzyme ligase. Ligase recognizes the phosphorylated 5' ends of the linear double-stranded PCR products and creates a new phosphodiester bond with the free 3' ends. This reaction completes the plasmid syntheses process. However, to prevent the introduction of the unmutated template plasmid into the cells, it is removed by DpnI, an enzyme that digests methylated DNA.[38] The template plasmid was isolated from a bacterial source, so it is naturally methylated. The methylated nucleotides are recognition sites for DpnI, and DpnI cleaves the double-stranded plasmid DNA. Since the PCR products are not methylated, the nucleotides are not recognized by DpnI, and the DNA ise not cut. Once the template plasmid has been digested, the mutated circular plasmids can be introduced to *E.coli* cells.[38] This is known as transformation, and it is the process by which exogenous DNA is acquired by an organism. The entirety of the cloning procedure can be seen in *Figure 6*.

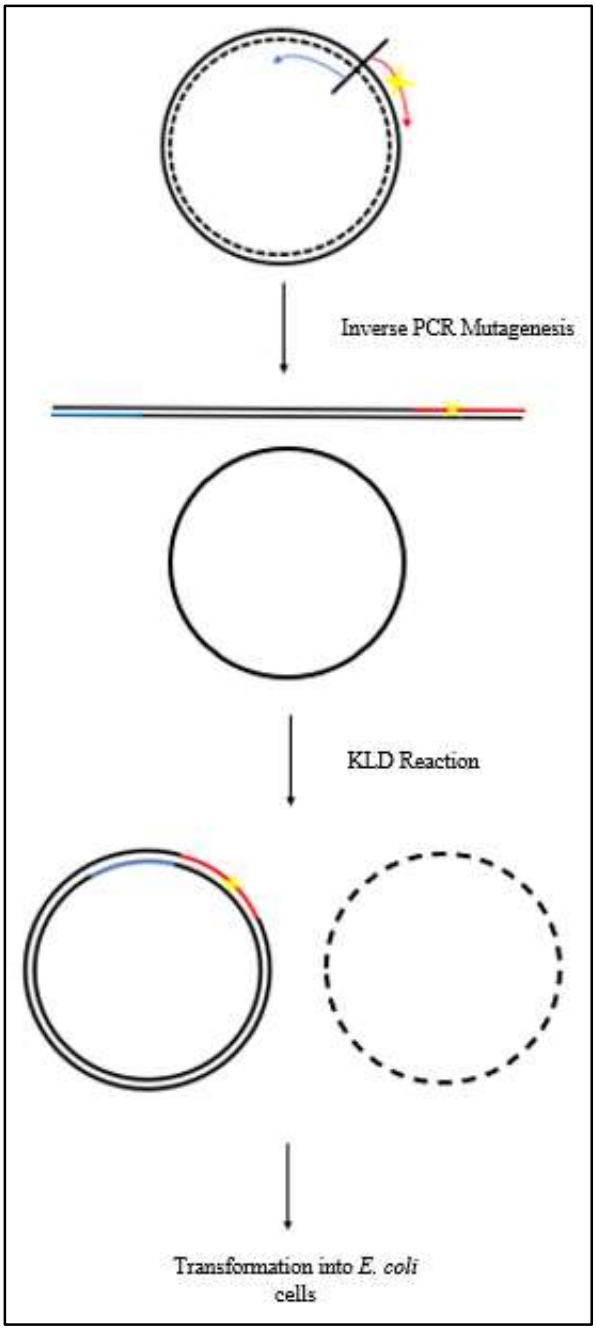


Figure 6. Basic schematic diagram of inverse PCR mutagenesis, KLD reaction, and transformation. The forward primer with desired mutation is indicated by the red arrow while the reverse primer is indicated by the blue arrow. Upon completion of amplification of the linearized plasmid with the desired mutation, the KLD reaction is performed to circularize the DNA and eliminate the template plasmid.

F. Primer Design

Primers are short sequences of nucleic acids, either DNA or RNA, that are complementary to the template DNA sequence. DNA replication cannot occur without primers. DNA polymerase can only add nucleotides to a free 3' hydroxyl end, which is not available on a single-strand of template DNA. Primers base-pair to single-stranded template DNA and offer a free 3' hydroxyl end for the polymerase to add an incoming nucleotide. During the polymerase chain reaction, primers allow for the association of DNA polymerase after denaturation of double-stranded DNA and before elongation of the sequence.

Primer design is crucial to the PCR process, and it helps balance the specificity and efficiency of DNA amplification.[39] The specificity refers to the likelihood that a mispriming event occurs while efficiency refers to the ability of the primer pair to amplify the PCR product to the theoretical maximum, which is a twofold increase of the PCR product per cycle.[39] The length of the primer and the annealing temperature of the PCR reaction are two variables that control the specificity, and the typical length for most PCR amplification is 18 to 30 bases.[39,40] A shorter primer could lead to the formation of nonspecific and undesirable PCR products, and a longer primer could result in a smaller quantity of primed templates during annealing.[39,40] In the latter case, this small inefficiency would augment at each annealing step during thermocycling and there would be a noticeable reduction in the amplified product. Primers of the appropriate length are also sequence-specific if the annealing temperature is within a few degrees of the melting temperature (T_m) of the primer.[39] The primer sequence is another important feature in primer design, especially the 3'-end.[40] The 3'-end is critical because its composition regulates the specificity and sensitivity of the PCR reaction. There should not be 3 or more guanosine or cytosine bases on this end of the primer, as this could stabilize nonspecific annealing onto the template.[40]

A thymidine on the 3'-end would also affect specificity because this nucleotide is more susceptible to mispriming.[40] Design of the 3'-end primer should include a restriction site, stop codon, and overlap with the strand opposite the 3'-end of the specific gene.[40] The 5'-end of the primer is not as important for annealing, but there are several elements that it must contain: a restriction site, 5'-extension to the site, start codon, and overlap with the specific gene target.[40]

For inverse PCR mutagenesis, back-to-back non-overlapping primers are utilized. These primers can be used for substitutions, deletions, small and large insertions. Substitutions are made by incorporating the desired nucleotide substitutions in the forward primer, leaving at least 10 nucleotide bases on the 3' end. The reverse primer must be designed so that the 5' ends of both primers anneal back-to-back (Figure 7).

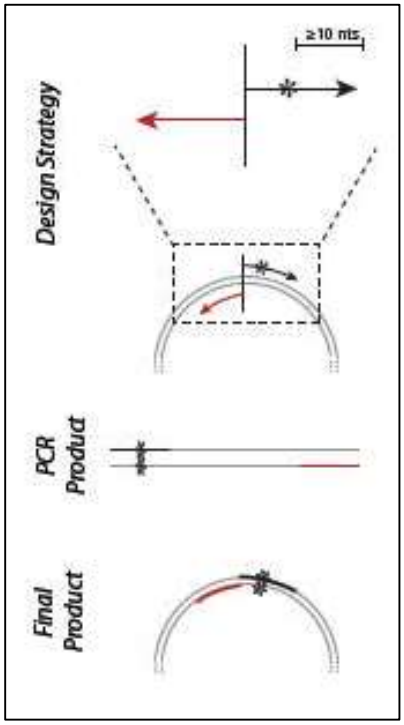


Figure 7. Primer design for substitution with the forward primer (black), reverse primer (red), and desired nucleotide change (*).[38]

G. Gel Electrophoresis

Agarose gel electrophoresis is a method for separating, identifying, and purifying DNA fragments and can also be used to separate RNA and proteins. Gel electrophoresis utilizes an electric field to move DNA, which is negatively charged, through an agarose gel towards a positive electrode.[41] It has three main steps with the first being the preparation of the gel with a concentration of agarose appropriate for the size of the DNA.[42] The second step is loading the DNA samples containing dye into sample wells alongside a molecular weight ladder and running them at a voltage using a voltage source and electrodes.[41,42] A molecular weight ladder is a collection of fragments of DNA of known lengths, and this ladder is necessary to estimate the length of the simultaneously run samples.[41] The last stage involves either the staining of the gel or illumination of the gel containing ethidium bromide with UV light.[42]

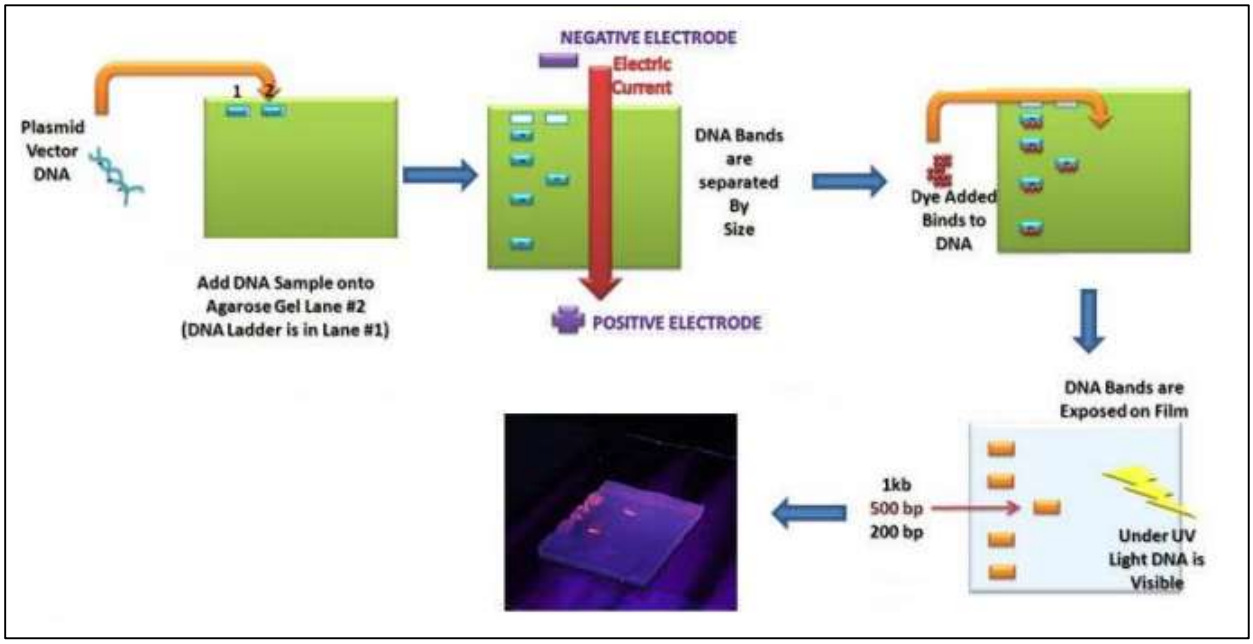


Figure 8. Schematic diagram of gel electrophoresis illustrating the loading of the wells, the movement of DNA and separation into bands, and the UV illumination of the bands in the gel containing ethidium bromide.[43]

II. Methods and Materials

A. Plasmid Design

1. Rat Preptin Ala^{21} Gene
2. His^{6} Affinity Tag
3. GFP fusion protein
4. TEV Protease recognition site
5. Ampicillin Selection Marker
6. LAC Operon Inducible Promoter

A custom synthetic plasmid (pD444-SR:His6-GFP-TEV-Preptin Ala 21, Figure 9) was purchased from ATUM. The plasmid was designed to encode for a rat-preptin analog with a His^{6}-green fluorescent protein (GFP) tag. The preptin analog contained an alanine mutation at residue 21, previously a phenylalanine, to prolong the half-life of the hormone *in vivo*.[44] Our lab lacks access to a preparative HPLC or UV detector capable of purifying our fusion protein. Therefore, the gene for the GFP protein was inserted on the 5' end of the preptin gene. GFP was isolated from the *Aequorea Victoria* jellyfish, and it produces a strong fluorescence emission peak at 509 nm and an excitation peak at 395 nm. This allows the fusion protein to be tracked during purification using an affordable UV penlight. After the fusion protein has been isolated, the His^{6}-GFP tag can be removed by incubation with the tobacco etch virus (TEV) protease. The TEV protease is highly site-specific, and the gene for the His^{6}-GFP fusion tag includes the recognition sequence for cleavage on its C-terminal end.

The plasmid also contains a gene for ampicillin resistance. This was useful during the transformation of the mutated plasmids into *E. coli* cells, as only the cells that were successful at incorporating the plasmid survived and grew while those that did not were killed by the ampicillin on the plates.

In *E. coli*, the *lac* operon, which is the orange portion of the plasmid map, codes for enzymes involved in the metabolism of lactose. The *lac* operon is efficiently expressed when lactose is present and glucose is not, and transcription of the operon is regulated by the *lac* repressor and the catabolite activator protein (CAP). Isopropyl-β-D-thiogalactopyranoside (IPTG) is a molecular mimic of allolactose (lactose metabolite) and can be used to bind and inhibit the *lac* repressor by changing its conformational state. This allows for the expression of the *lac* operon. IPTG is not metabolized by *E. coli* cells, which ensures that the concentration of IPTG remains constant and continues to induce the expression of the *lac* operon. Thus, IPTG allows for the transcription of the DNA by RNA polymerase and subsequent translation into protein. This gives a method for enhancing protein expression.

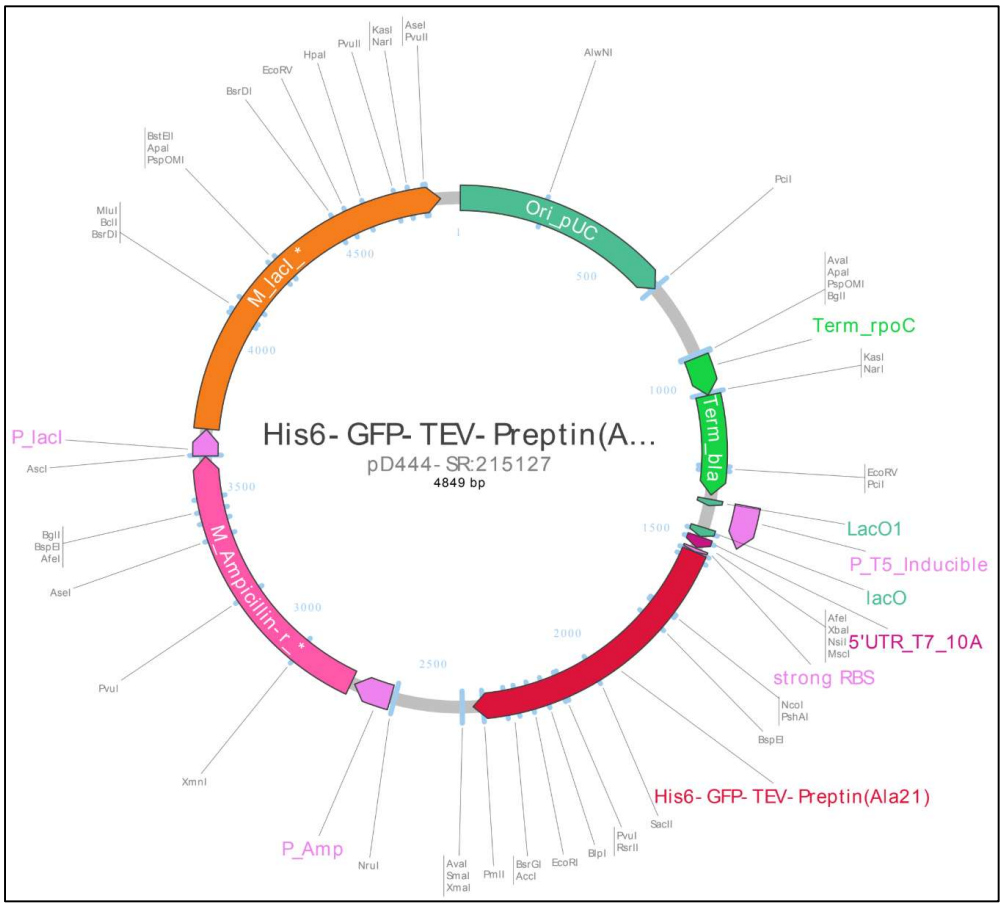


Figure 9. Plasmid map of pD444-SR:His6-GFP-TEV-Preptin Ala 21.

B. Primer Design

NEBasechanger was used to design the forward and reverse primers for the three mutations and calculate the annealing temperatures. Both the forward and reverse primers were necessary to perform the desired alanine substitutions using the NEB Q5 Site-Directed Mutagenesis Kit. The base pairs for the forward primers and the mutations are located in *Table 1* and were obtained from the plasmid sequence in the *Appendix*. *Table 2* features the sequences that were sent to New England BioLabs to create the custom primers along with the melting and annealing temperatures specific to the primers.

Table 1. Base pair sequences for the forward primers and mutations for W27A, R28A, and Q29A.

	Corresponding Base Pair Sequence
W27A Forward Primer	2349-2371
W27A Mutation	2359-2361
R28A Forward Primer	2352-2374
R28A Mutation	2362-2364
Q29A Forward Primer	2355-2377
Q29A Mutation	2365-2367

Table 2. Sequences, melting temperatures (T_m), and annealing temperatures (T_a) of the forward and reverse primers with the mutation to alanine in red.

Primer	Sequence	T_m	T_a
W27A Forward	GTTCGACACGgcaCGTCAGAGCG	61 °C	62 °C
W27A Reverse	TTAAACGCTTTGCCCACC	64 °C	62 °C
R28A Forward	CGACACGTGGgcaCAGAGCGCGG	63 °C	62 °C
R28A Reverse	AACTTAAACGCTTTGCCC	61 °C	62 °C
Q29A Forward	CACGTGGCGTgcaAGCGCGGGTC	62 °C	58 °C
Q29A Reverse	TCGAACTTAAACGCTTTG	57 °C	58 °C

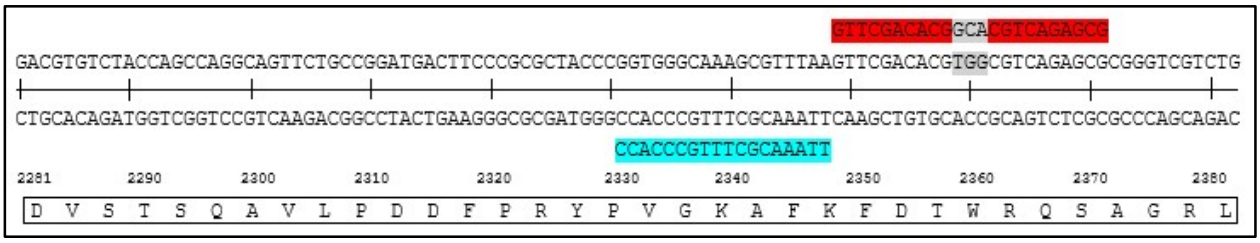


Figure 10. Design of the W27A primer with the forward primer (red), reverse primer (blue), and the alanine mutation (grey).

C. Selection

One of the most common ways to culture bacteria is using Luria broth (LB) agar plates (Figure 11). The LB is a nutrient-rich media, and agar is added to the LB to create a gel that is suitable for bacterial growth; bacteria cannot digest agar but receives nutrients from the LB. During this study, ampicillin was added to the gel to grow up the cells that successfully incorporated the specific, antibiotic-resistant plasmid. Luria broth-ampicillin plates were prepared by spreading 40 µL of 500x ampicillin in 100 µL of distilled water onto 20 mL LB-agar plates.

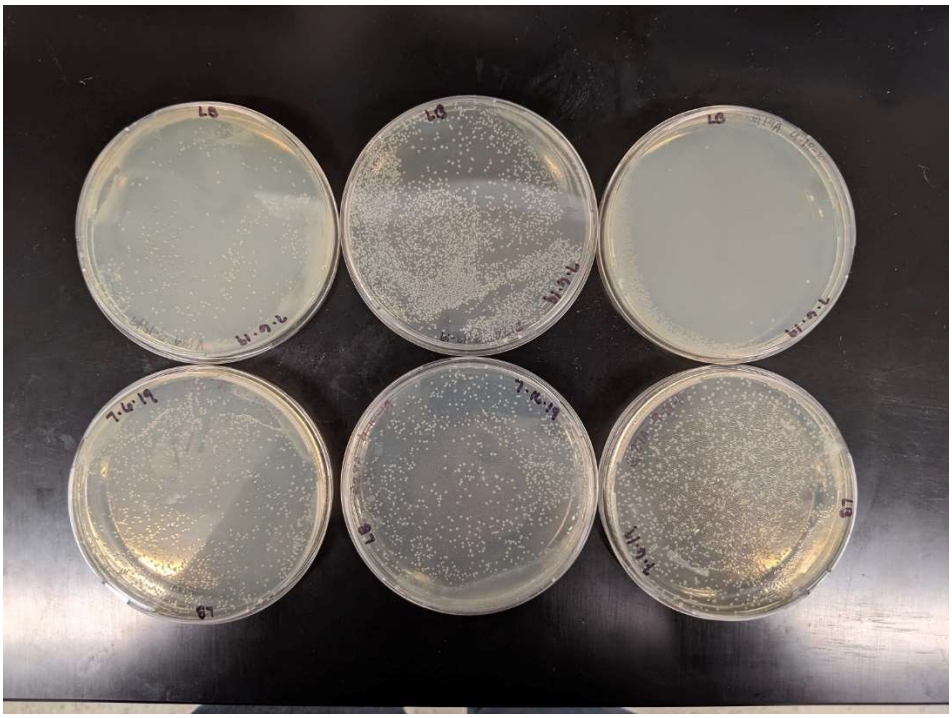


Figure 11. Luria broth-ampicillin plates with visible colonies.

D. Mutagenesis

The commercially available Q5® Site-Directed Mutagenesis Kit (New England BioLabs, Inc) was used to perform the cloning procedure. The overview of this process is seen in *Figure 12*.

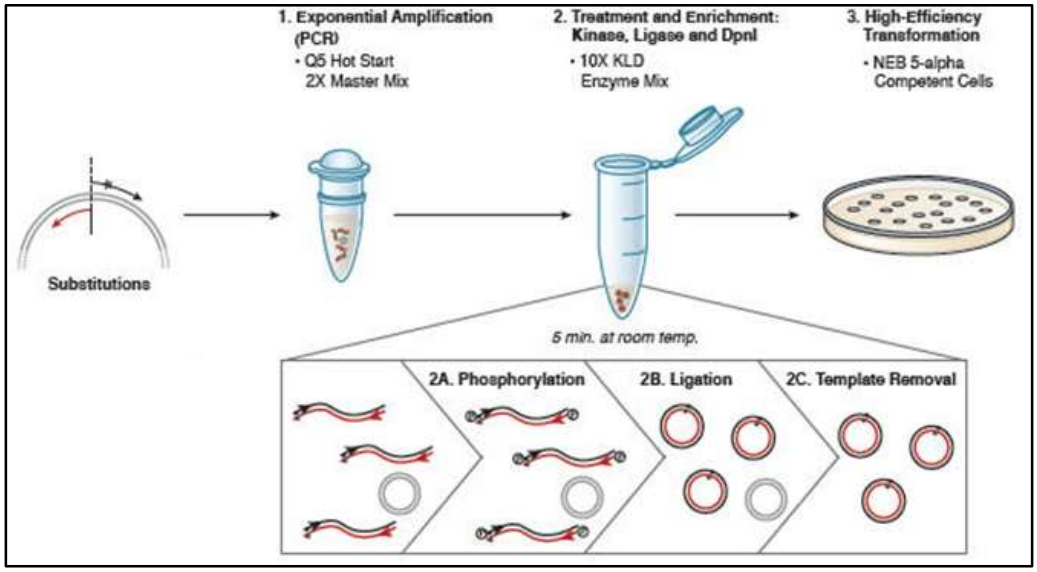


Figure 12. Overview of the NEB Q5 Site-Directed Mutagenesis Kit.[21]

An important note for this kit is that it utilizes a Q5 high-fidelity polymerase instead of *Taq* polymerase. The protocol provided by the manufacturer was followed with minimal changes. The first step of the mutagenesis protocol was exponential amplification. The reagents in *Table 3* were assembled in a PCR tube, mixed completely, and transferred to a Eppendorf Mastercycler Personal thermocyler (PCR reaction conditions are in Table 4).

Table 3. Reagents for exponential amplification.

	25.1 µL RXN	Final Concentration
Q5 Hot Start High-Fidelity 2X Master Mix	12.5 µL	1X
10 µM Forward Primer	1.3 µL	0.5 µM
10 µM Reverse Primer	1.3 µL	0.5 µM
Template DNA (1-25 ng/µL)	1.0 µL	1-25 ng
Nuclease-free water	9.0 µL	

Table 4. Conditions for thermocycling.

Steps	Temperature (°C)	Time
Initial Denaturation	98	30 s
25 Cycles:	98	10 s
	62	30 s
	72	30 s/bp = 80
Final Extension	72	2 mins
Hold	4	

The second step was the simultaneous phosphorylation and ligation of the linear PCR products with concurrent digestion of the template plasmid. The kinase, ligase, and digestion enzymes (Table 5) were added to the PCR tube, mixed thoroughly with a pipette, and incubated at room temperature for five minutes.

Table 5. Reagents for KLD reaction.

	Volume	Final Concentration
PCR Product	1 µL	
2X KLD Reaction Buffer	5 µL	1X
10X KLD Enzyme Mix	1 µL	1X
Nuclease-free water	3 µL	

After the KLD reaction, the mutated circular plasmids were used for transformation of competent *E. coli* cells. NEB 5-alpha competent *E. coli* cells (50 µL, included with the kit) were thawed, and 5 µL of the KLD mixture from the second step was added to the cells. The tube was flicked to mix. Flicking ensured that the cells were not damaged, which could occur if they were vortexed. The mixture was placed on ice for 30 minutes, heat shocked at 42°C for 30 seconds, and placed back on ice for 5 minutes. The heat shock encouraged the bacterial cells to take up the mutated plasmid. Then, 950 µL of room temperature SOC broth (Super Optimal Broth with 20 mM glucose) was pipetted into the mixture and incubated at 37°C for 60 minutes in a Thermo Scientific MaxQ 400 Incubator while shaking at 250 rpm. The cells were then mixed via flicking and inverting the tube, and 75 µL of each transformation reaction was spread onto LB-ampicillin plates and incubated overnight at 37°C. Overnight, the bacterial that did not incorporate the ampicillin-resistant plasmid died off. The bacteria that successfully incorporated the plasmid

survived and reproduced. Each surviving bacterium gave rise to a colony of identical, plasmid-containing bacteria.

E. Overnight Culture for Plasmid Isolation

The morning following the transformation of W27A, R28A, and Q29A analogs, the LB-ampicillin plates were observed for growth. One colony was picked from each plate and used to inoculate 4 mL LB cultures with 8 mL 500X ampicillin. The cultures were then returned to the Thermo Scientific MaxO 400 Incubator for 16 hours at 37$^{\circ C}$ while shaking at 250 rpm. To avoid having to perform the mutagenesis every time plasmid is required, glycerol stock solutions were prepared by mixing 500 µL of the cell culture with 500 µL of 50% glycerol. The glycerol stock was then stored at -80$^{\circ C}$.

F. Plasmid Isolation

An Invitrogen PureLink Quick Plasmid Miniprep Kit was used to isolate the plasmid DNA from *E.coli* cells via cell lysis. The first step in the plasmid isolation protocol was to centrifuge 3.5 mL of the 16-hour overnight LB culture and remove the medium. Resuspension buffer with RNase A (250 µL) was added to the pellet of cells and resuspended until homogenous. Lysis buffer (250 µL) was added and mixed gently by inversion of the tube until homogenous. This was incubated at room temperature for five minutes. Precipitation buffer (350 µL) was added and mixed by inversion until homogenous. The lysate was centrifuged 13,000 x g for ten minutes. The supernatant from the precipitation step was loaded onto a spin column in a 2 mL wash tube, which was then centrifuged at 12,000 x g for one minute. The flow through was discarded, and the column was placed back into the wash tube. An optional wash step was performed in which 500 µL wash buffer (W10) with ethanol was added to the column and incubated for one minute at room

temperature. Wash buffer (W9) with ethanol (700 µL) was added to the column and centrifuged at 12,000 x g for one minute. The flow-through was discarded, and the column was placed back into the wash tube. The column was centrifuged again at 12,000 x g for one minute and the flow-through discarded. The spin column was then placed in a 1.5 mL elution tube, and 75 µL of nuclease free water was added to the center of the column. The column was incubated for one minute at room temperature. The column was centrifuged at 12000 x g for two minutes. This resulted in the purified plasmid DNA being contained in the elution tube.

G. Gel Electrophoresis

A 1% agarose gel was prepared by dissolving 0.50 g agarose in 100 mL of Tris Acetate EDTA (TAE) buffer (pH 8.3), which was prepared by mixing 300 mL of 1X TAE buffer with 7.5 µL of ethidium bromide (visualizing agent). Plasmid samples for gel electrophoresis were prepared by adding 5 µL DNA, 4.2 µL NEB purple loading dye, and 15.8 µL nuclease free water in a microcentrifuge tube. Portions of the plasmid samples (20 µL) were pipetted into separate lanes of the agarose gel and run at 110 V and room temperature.

H. DNA Sequencing

Samples of the mutated plasmids of W27A, R28A, and Q29A were sent to MCLabs in California for analysis. Sanger sequencing was used to confirm the presence or absence of the alanine mutations.

III. Results and Discussion

A. Rationale

Preptin exhibits both metabolic and mitogenic properties, so it has the potentiality of being used as a treatment for types 2 diabetes and osteoporosis. This is especially important due to the adverse effects seen with glibenclamide, one of the most common drugs for type 2 diabetes. Glibenclamide and preptin behave similarly in regard to stimulating insulin secretion. However, the pathway by which preptin accomplishes this is unknown. This work was performed to study the structure-activity relationship of the three chosen amino acid residues in order to ultimately assess the effect of the mutations on the metabolic activity of preptin. Amino acid residues Trp^{27}, Arg^{28}, and Gln^{29} of rat preptin were chosen for their properties, such as aromaticity and charge, and their potential to interact with the unknown preptin receptor.

B. Results

Various factors came into conflict with this research. The first problem occurred during our initial efforts at cloning, which presented itself during transformation of the cells. We were having trouble growing up any successful colonies, so we knew that some step of the cloning procedure was not working properly. To determine which step, the original plasmid, PCR product, and KLD product for one of the analogs were analyzed in an agarose gel. The band for the original plasmid was strong, but the band for the PCR product was barely visible. We hypothesized that this faintness was the presence of the original plasmid that did not mutate, and this meant that the PCR was not working. The KLD did not have any bands, which further supported this hypothesis. The NEB Q5 Mutagenesis kit itself was tested to see if that was the problem, and the control plasmid from the kit, PCR product, and KLD product were analyzed in an agarose gel, and the

lanes displayed that the kit was not the problem. We contacted New England BioLabs and were advised to increase the annealing temperature, as the Q5 high-fidelity polymerase requires higher temperatures. The increase in annealing temperature proved successful.

The second problem we encountered stemmed from the construction being done in LeNoir Hall and our lab. Transformation of the NEB 5-alpha *E. coli* cells with our mutated plasmid and subsequent purification was unsuccessful. Analysis of gel electrophoresis concluded that no plasmid was present because of the absence of bands. From increasing the annealing temperatures and ensuring that cloning was successful, we knew that there had to be another cause and tested the ampicillin. To do this, untransformed cells (not ampicillin-resistant) were streaked onto an LB-ampicillin plate, transformed cells (ampicillin-resistant) were streaked onto an LB-ampicillin plate, and untransformed cells were streaked onto a kanamycin (a different antibiotic) plate. Ideally, the *E. coli* cells that incorporated the ampicillin-resistant plasmid should have grown up on the LB-ampicillin plate in small colonies. However, both the plate with the untransformed cells and transformed cells on LB-ampicillin plates had plaque growth. The kanamycin plate had no growth at all, and this was to be expected since the untransformed cells were not ampicillin- or kanamycin-resistant. We hypothesize that the move to the new lab resulted in our ampicillin being left at improper temperatures for storage, and degradation occurred. We obtained a new ampicillin stock to use during transformation in the plates and the cell cultures.

After resolving the problems with cloning and with degraded ampicillin, purification of the plasmids containing mutations encoding for W27A, R28A, and Q29A from the transformed cells was confirmed, seen in *Figure 13*.

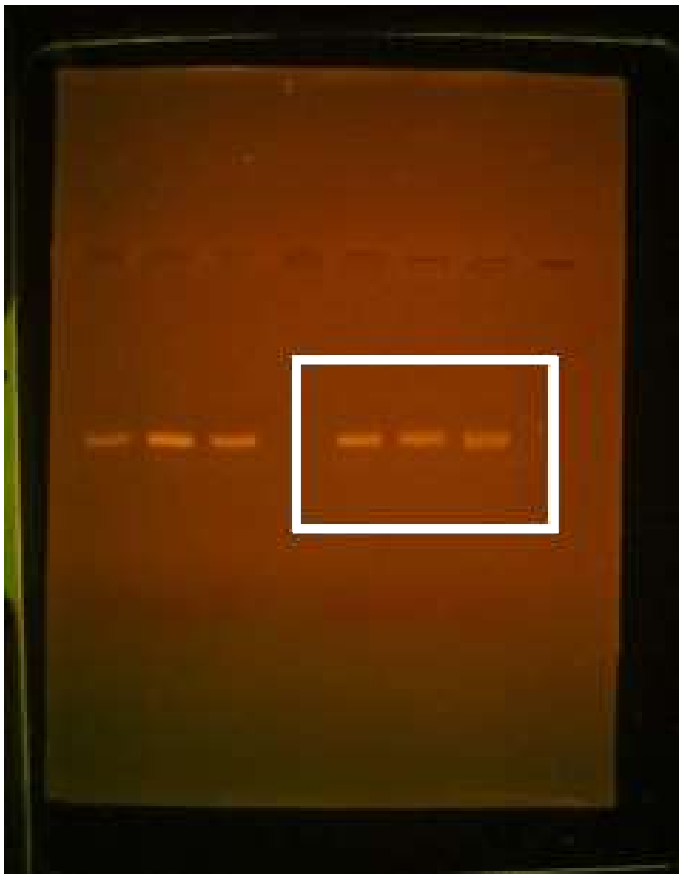


Figure 13. Gel electrophoresis for mutations W27A (Lane 4), R28A (Lane 5), and Q29A (Lane 6).

Figures 14, 15, and *16* display the DNA sequencing results of the mutated plasmids. The mutation at residue 27 was inconclusive, as the sequencing revealed a valine instead of the desired alanine. Further inspection of the residues earlier in the sequence (residues 24-26) revealed disparities from the known sequence for rat preptin. On the other hand, the correct mutations were obtained for residues 28 and 29, with only a couple of residues at the beginning of the preptin sequence different in the latter.

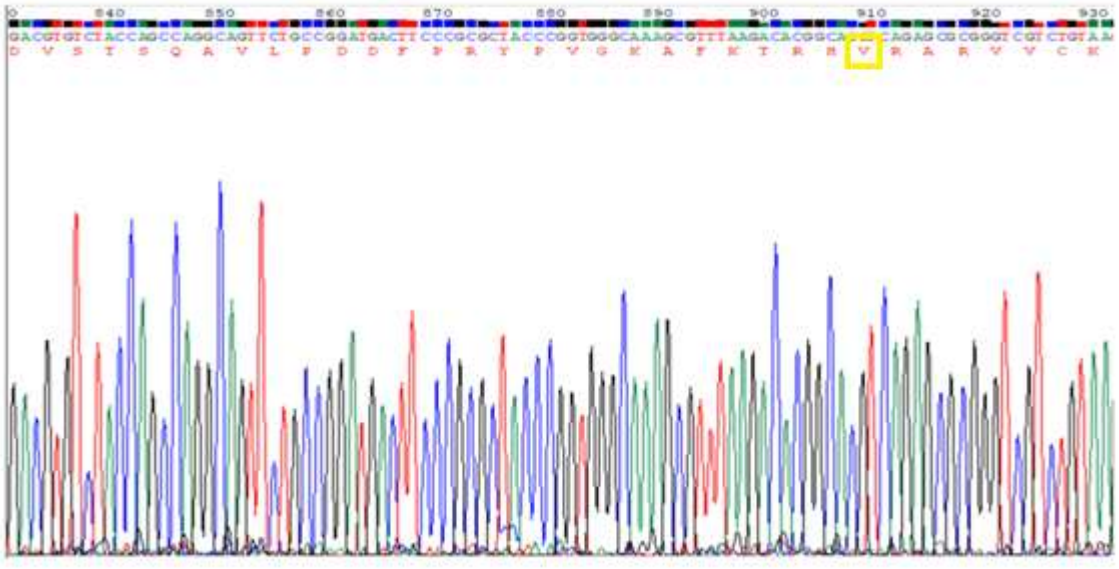


Figure 14. Sequencing results for the unsuccessful mutation of the GFP-Preptin Ala27 gene (highlighted in yellow).

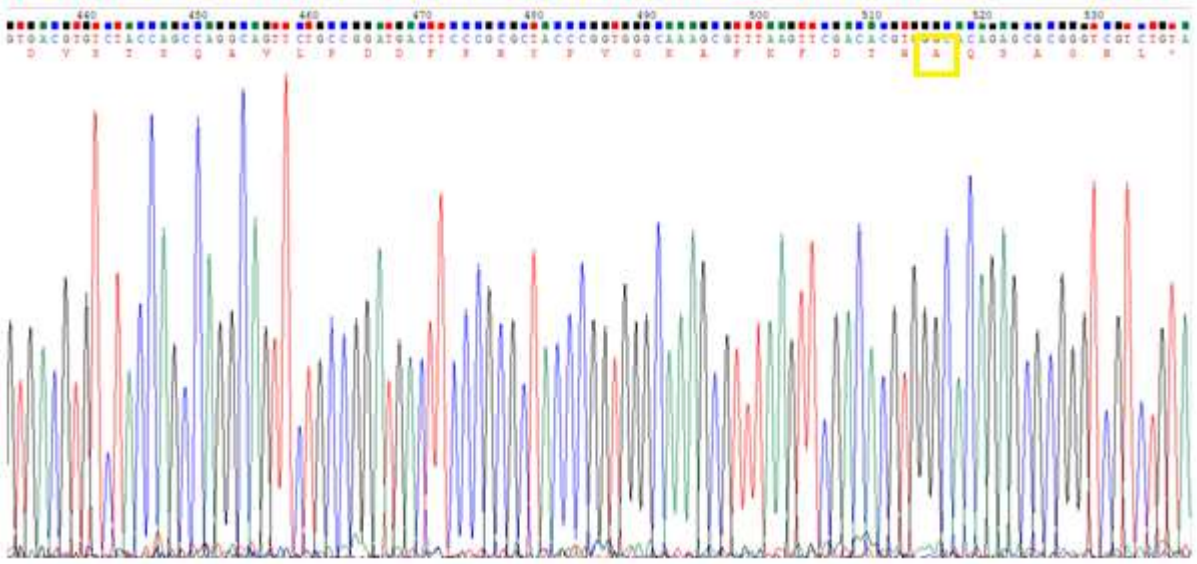


Figure 15. Sequencing results for the successful mutation of the GFP-Preptin Ala28 gene (highlighted in yellow).

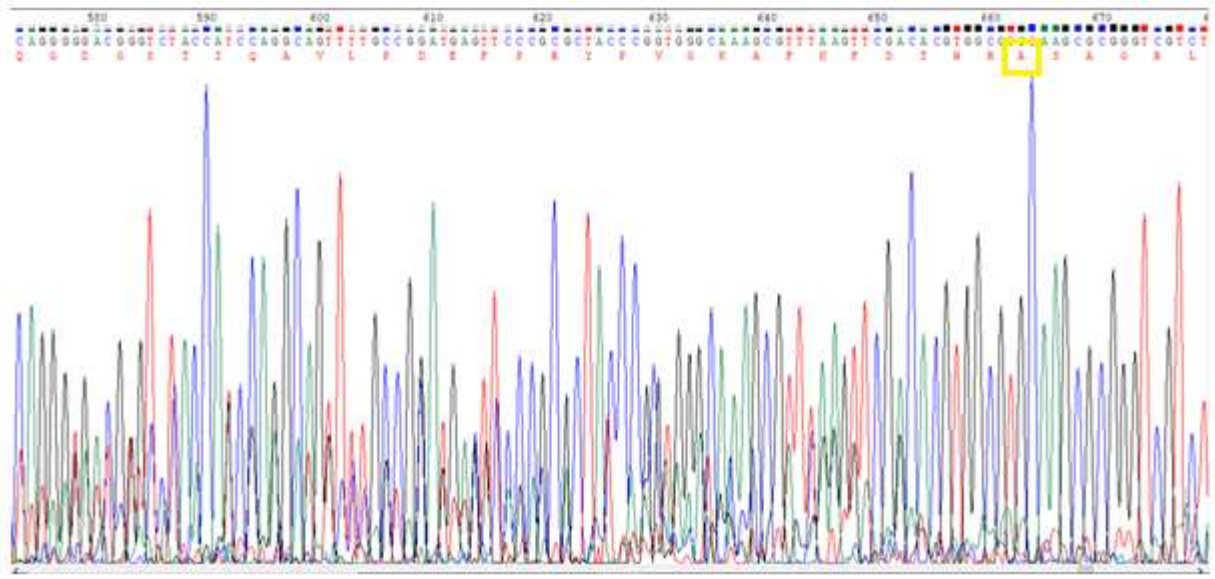


Figure 16. Sequencing results for the successful mutation of the GFP-Preptin Ala29 gene (highlighted in yellow).

There are various factors that affect the quality of sequencing results. For Sanger sequencing, which MCLabs used for our preptin analogs, deoxynucleotides (dNTPs) normally found in DNA are used alongside dideoxynucleotides (ddNTPs) . These ddNTPs have a hydrogen on the 3' carbon instead of the hydroxyl group that is found on dNTPs, and because ddNTPs lack the hydroxyl group, the addition of further dNTPs cannot occur.[45] This results in copious fragments of various lengths, which are reacted with the four ddNTPs. These ddNTPs are labeled with dye, and the wavelengths that cause the fragments to fluorescence are accumulated into a chromatogram.[45] The colored peaks in the chromatogram refer to the nucleotides at those positions, and the codons reflect the amino acids. Errors during this process can be attributed to the absence of a priming site, degraded primer, insufficient amount of DNA template, residual primers, and much more.[45] However, error from the sequencing protocol seems unlikely, as two of the three mutations resulted in the correct alanine mutation. Error most likely occurred during the site-directed mutagenesis, resulting in the plasmid not incorporating the desired alanine mutation.

C. Future Work

The first issue that will need to be addressed with the continuation of this work is obtaining a successful mutation of the 27th residue. After that, the main objective of the future work would be to test the metabolic activity of the preptin analogs using an enzyme-linked immunosorbent assay (ELISA). The ELISA technique would allow us to quantify the ability of the preptin analogs to enhance insulin secretion. Prior to this, however, large-scale production and purification of the analogs must be performed. This will be accomplished by first transforming BL21 cells, which are better than the NEB 5-alpha cells for protein expression, with the mutated plasmids. Isopropyl-β-D-thiogalactopyranoside (IPTG) will be used to induce the cell cultures. IPTG mimics the lactose metabolite, allolactose, and as stated earlier, it allows for the expression of the *lac* operon. Expression of the *lac* operon leads to the transcription of the DNA and the translation to the preptin protein. To optimize the expression of preptin, different concentrations of IPTG will need to be tested to determine the concentration at which the most protein is expressed. After the IPTG treatment, sonication will be used to disrupt the cell membranes of the bacterial cells and extract the DNA, RNA, and protein. Nickel affinity chromatography is used to separate the desired proteins, which have a His^6 affinity tag, from the native bacterial proteins that are present. The nickel columns utilize immobilized metal-affinity chromatography, and the immobilized nickel resin most likely to be used is nickel-nitrilotriacetic acid (Ni-NTA).[46] The NTA is a tetradentate chelating ligand, and it binds the Ni^{2+} ions via four coordination sites, leaving two coordination sites available to interact with the His^6 residues of the preptin analogs.[47] Thus, proteins with the His^6 tag will have a high affinity for the nickel and become immobilized on the resin, and other proteins will bind with low affinity or will not bind at all.[47] Immobilized proteins can be washed through the column using increasing concentrations of imidazole. Size exclusion columns must be

used in addition to Ni-NTA because of the imidazole and other salts that are introduced during the separation of the desired protein from native bacterial protein. To assess the purity and yield of the protein, SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) can be used. SDS-PAGE separates proteins based on charge and molecular weight. SDS is an anionic detergent that breaks hydrogen and hydrophobic bonds in proteins while also binding to proteins and covering their intrinsic charges.[48] The only determining factor for the migration of protein through the gel is the molecular weight, and since the protein is uniformly negatively charged due to SDS, the protein will migrate towards the positive electrode.[49] Proteins with a greater molecular weight take longer to migrate through the gel while proteins with a smaller molecular weight move more quickly through the gel. If the sample being tested is pure and contains one protein, the gel will display one single band representative of the protein; if there are multiple proteins, there will be multiple bands. Once purity and the molecular weight of preptin have been confirmed, the ELISA assay can be done to assess the metabolic activity of the preptin analogs, which will be useful in determining the residues that are vital to the function of the peptide.

IV. References

1. Chronic Disease Data. https://www.cdc.gov/chronicdisease/data/index.htm (accessed March 9, 2019).
2. Chronic Diseases and Their Common Risk Factors. https://www.who.int/chp/chronic_disease_report/media/Factsheet1.pdf (accessed March 9, 2019).
3. Statistics About Diabetes: American Diabetes Association. http://www.diabetes.org/diabetes- basics/statistics/ (accessed March 9, 2019).
4. Diabetes Complications – American Diabetes Association. http://www.diabetes.org/living- with-diabetes/complications/ (accessed March 9, 2019).
5. What is Type 1 Diabetes? https://www.joslin.org/info/what_is_type_1_diabetes.html (accessed March 10, 2019).
6. Felman, A. Insulin Resistance – Medical News Today. https://www.medicalnewstoday.com/articles/305567.php (accessed March 12, 2019).
7. Giovannucci, E.; Harlan, D. M.; Archer, M. C.; Bergenstal, R. M.; Gapstur, S. M.; Habel, L. A.; Pollak, M.; Regensteiner, J. G.; Yee, D. Diabetes and Cancer: A consensus report. *Diabetes Care.* **2010**, *33*, 1674-1685.
8. Scappaticcio, L.; Maiorino, M. I.; Bellastella, G.; Giugliano, D.; Esposito, K. Insights into the relationships between diabetes, prediabetes, and cancer. *Endocrine*. **2017**, *56*, 231-239.
9. Sever, R.; Brugge, J. S. Signal transduction in cancer. *Cold Spring Harb Perspect Med.* **2015**, *5*, 1-21.

10. Scappaticcio, L.; Maiorino, M. I.; Bellastella, G.; Giugliano, D.; Esposito, K. Insights into the relationships between diabetes, prediabetes, and cancer. *Endocrine*. **2017**, *56*, 231-239.

11. Utiger, R. Insulin-like growth factor. https://www.britannica.com/science/insulin-like-growth-factor (accessed April 6, 2020).

12. Holly, J. M. P.; Biernacka, K.; Perks, C. M. The neglected insulin: IGF-II, a metabolic regulator with implications for diabetes, obesity, and cancer. *Cells*. **2019**, *8*, 1-32.

13. Doorn, J. Insulin-like growth factor-II and bioactive proteins containing a part of the E-domain of pro-insulin-like growth factor-II. *Biofactors*. **2020**, 1-20.

14. Buchanan, C. M.; Phillips, A. R.; Cooper, G, J. Preptin derived from proinsulin-like growth factor II (proIGF-II) is secreted from pancreatic islet β-cells and enhances insulin secretion. *Biochem J.* **2001**, *360*, 431-439.

15. Aydin, S. Three new players in energy regulation: Preptin, adropin, and irisin. *Peptides*. **2014**, *56*, 94-110.

16. Cheng, K. C.; Asakawa, A.; Ushikai, M.; Kato, I.; Sato, Y.; Cheng, J. T.; Inui, A. Characterization of preptin-induced insulin secretion in pancreatic beta cells. *J. Endocinol.* **2012**, *215*, 43-49.

17. Hambrock, A.; Loffler-Walz, C.; Quast, U. Glibenclamide binding to sulphonylurea receptor subtypes: dependence on adenine nucleotides. *Br. J. Pharmacol.* **2002**, *136*, 995-1004.

18. Avogaro, A.; De Kreutzenberg, S. V.; Fadini, G. P. The impact of glucose-lowering medications on cardiovascular disease. *Cardiovasc Endocrinol Metab.* **2018**, *7*, 13-17.

19. Iliodromitis, E. K.; Lazou, A.; Kremastinos, D. T. Ischemic preconditioning: Protection against myocardial necrosis and apoptosis. *Vasc Health Risk Manag.* **2007**, *3*, 629-637.

20. Smits, P.; Bijlstra, P. J.; Russel, F. G.; Lutterman, J. A.; Thien, T. Cardiovascular effects of sulphonylurea derivatives. *Diabetes Res Clin Pract.* **1996**, *31*, 55-59.

21. Chahal, H. Comparative Safety and Efficacy of Glibenclamide in the Elderly. https://www.who.int/selection_medicines/committees/expert/19/applications/Sulfonylurea_18_5_A_R.pdf?ua=1 (accessed April 14, 2019).

22. Cornish, J.; Callon, K. E.; Bava, U.; Watson, M.; Xu, X.; Lin, J. M.; Chan, V. A.; Grey, A. B.; Naot, D.; Buchanan, C. M.; Cooper, G. J. S.; Reid, I. R. Preptin, another peptide product of the pancreatic β- cell is osteogenic in vitro and in vivo. *Am. J. Physiol.* **2007**, *292*, 117-122.

23. Bone Metabolism: The Cycle of Bone Growth and Resorption. https://www.alpco.com/cycle-bone-growth-resorption (accessed April 6, 2020).

24. LeRoith, D. Insulin-like Growth Factors and Cancer. Springer Science+Business Media: New York, NY, 2012, pp. 1-290.

25. Kowalczyk, R.; Yang, S. H.; Brimble, M. A.; Callon, K. E.; Watson, M.; Park, Y. E.; Cornish, J. Synthesis of truncated analogues of preptin-(1-16), and investigation of their ability to stimulate osteoblast proliferation. *Bioorg. Med. Chem.* **2014**, *22*, 3565-3572.

26. Anjana, R.; Vaishnavi, M. K.; Sherlin, D.; Kumar, S. P.; Naveen, K.; Kanth, P. S.; Sekar, K. Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: A study based on orientation and distance. *Bioinformation*. **2012**, *8*, 1220-1224.

27. Site Directed Mutagenesis. https://www.neb.com/applications/cloning-and-synthetic-biology/site-directed-mutagenesis (accessed April 22, 2019).

28. Lefevre, F.; Remy, M. H.; Masson, J. M. Alanine-stretch scanning mutagenesis: a simple and efficient method to probe protein structure and function. *Nucleic Acids Res.* **1997**, *25*, 447-448.

29. Traditional Cloning Basics. https://www.thermofisher.com/us/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/molecular-cloning/cloning/traditional-cloning-basics.html (accessed February 23, 2020).

30. DNA cloning. https://www.sciencelearn.org.nz/resources/2031-dna-cloning (accessed April 25, 2020).

31. PCR Cloning Method. https://www.neb.com/applications/cloning-and-synthetic-biology/pcr-cloning (accessed February 23, 2020).

32. Guidelines for PCR Optimization with Thermophilic DNA Polymerases. https://www.neb.com/tools-and-resources/usage-guidelines/guidelines-for-pcr-optimization-with-thermophilic-dna-polymerases (accessed February 23, 2020).

33. Polymerase Chain Reaction. https://www.sigmaaldrich.com/technical-documents/articles/biology/polymerase-chain-reaction.html#ref (accessed February 23, 2020).

34. Stolovitzky, G.; Cecchi, G. Efficiency of DNA replication in the polymerase chain reaction. *Proc Natl Acad Sci USA.* **1996**. *93*, 12947-12952.

35. PCR Cycling Parameters – Six Key Considerations for Success. https://www.thermofisher.com/us/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/pcr-education/pcr-reagents-enzymes/pcr-cycling-considerations.html (accessed February 23, 2020).

36. Ochman, H.; Gerber, A. S.; Hartl, D. L. Genetic applications of an inverse polymerase chain reaction. *Genetics*. **1988**, *120*, 621-623.

37. Green, M. R.; Sambrook, J. Polymerase Chain Reaction. In *Molecular Cloning*. Cold Spring Harbor Protocols; New York, 2019.

38. Q5® Site-Directed Mutagenesis Kit. https://www.neb.com/products/e0554-q5-site-directed-_mutagenesis-kit#Quality,%20Safety%20&%20Legal (accessed April 22, 2019).

39. Diffenbach, C. W.; Lowe, T. M. J.; Dveksler, G. S. General concepts for PCR primer design. *Genome Res*. **1993**, S30-S37.

40. Cloning PCR Strategy Primer Design – EMBL. https://www.embl.de/pepcore/pepcore_services/cloning/pcr_strategy/primer_design/ (accessed February 23, 2020).

41. Agarose Gel Electrophoresis. https://www.addgene.org/protocols/gel-electrophoresis/ (accessed February 25, 2020).

42. Voytas, D. Agarose gel electrophoresis. *Curr Protoc Immunol*, **2001**.

43. Yilmaz, M.; Ozic, C.; Gok, Ilhami. Principles of Nucleic Acid Separation by Agarose Gel Electrophoresis. *In Gel Electrophoresis – Principles and Basics*; Magdeldin, S., Ed.; InTech: Europe, 2012; pp 33-40.

44. Buchanan, C. M.; Peng, Z.; Cefre, A.; Sarojini, V. Preptin analogues: Chemical synthesis, secondary structure and biological studies. *Chem Biol Drug Des*. **2013**, *82*, 429-437.

45. Curci, P. L.; Sonnante, G. How a small double-stranded trick can mislead Sanger sequencing. *J Biomol Tech*. **2015**, *26*, 80-82.

46. Nickel Columns and Nickel Resin. https://www.bio-rad.com/featured/en/nickel-columns-nickel-resin.html (accessed April 13, 2020).

47. Purification of Polyhistidine-Containing Recombinant Proteins with Ni-NTA Purification System. https://www.thermofisher.com/us/en/home/references/protocols/proteins-expression-isolation-and-analysis/protein-purification-protocol/purification-of-polyhistidine-containing-recombinant-proteins-with-ni-nta-purification-system.html (accessed April 13, 2020).

48. Purity analysis SDS-PAGE. http://www.mtoz-biolabs.com/pharm-purity-analysis-sds-page.html (accessed April 13, 2020).

49. Overview of Protein Electrophoresis. https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-electrophoresis.html (accessed April 13, 2020).

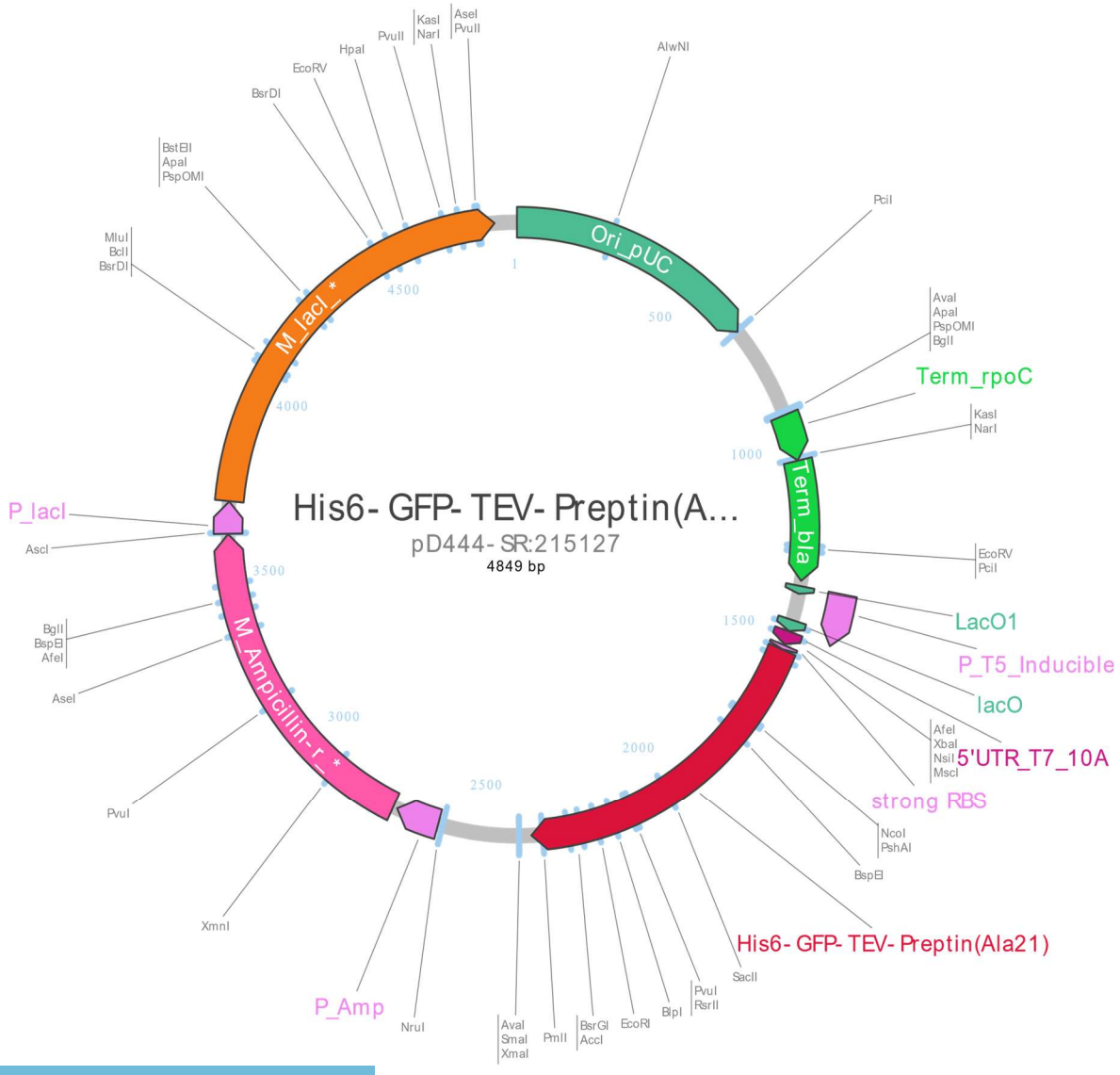
V. Appendix

A. Plasmid Map and Sequence for pD444-SR:His6-GFP-TEV-Preptin Ala21

DNA 2.0

Plasmid Map pD444-SR:215127

Only single and double cutters are shown in the map.



Original Author

DNA2.0, Inc.
1140 OBrien Drive, Suite A
Menlo Park, CA 94025
1-877-DNA-TOGO (Toll free) 1-650-853-8351 (Fax).
info@dna20.com
www.dna20.com

Feature Map

Name	Start	End	Direction
Insert: 215127	1522	2385	Forward
Ori_pUC	1	673	Forward
Term_rpoC	918	1037	Forward
Term_bla	1038	1342	Forward
LacO1	1355	1375	Forward
P_T5_Inducible	1355	1472	Forward
lacO	1443	1472	Forward
5'UTR_T7_10A	1473	1507	Forward
strong RBS	1510	1521	Forward
P_Amp	2637	2754	Forward
Ampicillin-r	2765	3625	Forward
P_lacI	3626	3707	Forward
lacI	3708	4790	Forward

Restriction Map

Name	Sequence	5' Cut Positions
AccI	GTMKAC	2287
AclI	AACGTT	1991, 2950, 3323, 3720
AfeI	AGCGCT	1457, 3501
AlwNI	CAGNNNCTG	262
ApaI	GGGCCC	915, 4269
ApaLI	GTGCAC	357, 2880, 4038
AscI	GGCGCGCC	3629
AseI	ATTAAT	3377, 4743
AvaI	CYCGRG	908, 2418
BclI	TGATCA	4072
BglI	GCCNNNNNGGC	919, 3434
BlpI	GCTNAGC	2164
BspEI	TCCGGA	1767, 3461
BspHI	TCATGA	1398, 2712, 4792
BsrBI	CCGCTC	742, 1363(C), 1424(C), 2710
BsrDI	GCAATG	4105(C), 4471
BsrGI	TGTACA	2252
BssHII	GCGCGC	2159, 3629, 4469, 4822

Name	Sequence	5' Cut Positions
BstEII	GGTNACC	4239
BstXI	CCANNNNNNTGG	3860, 3989, 4112
BtsI	GCAGTG	3133, 3153(C), 4423(C), 4791
EcoRI	GAATTC	2209
EcoRV	GATATC	1255, 4508
HincII	GTYRAC	1273, 2389, 4564
HpaI	GTTAAC	4564
KasI	GGCGCC	1032, 4697
MluI	ACGCGT	4058
MscI	TGGCCA	1546
NarI	GGCGCC	1033, 4698
NcoI	CCATGG	1711
NruI	TCGCGA	2627
NsiI	ATGCAT	1526
PciI	ACATGT	671, 1267
PmlI	CACGTG	2357
PshAI	GACNNNNGTC	1723
PspOMI	GGGCCC	911, 4265
PvuI	CGATCG	2108, 3183
PvuII	CAGCTG	4658, 4751
RsrII	CGGWCCG	2116
SacII	CCGCGG	2006
SmaI	CCCGGG	2420
SspI	AATATT	1024, 1185, 2747
XbaI	TCTAGA	1483
XmaI	CCCGGG	2418
XmnI	GAANNNNTTC	2952
Acc65I	GGTACC	no cuts
AgeI	ACCGGT	no cuts
AsiSI	GCGATCGC	no cuts
AvrII	CCTAGG	no cuts
BamHI	GGATCC	no cuts
BbsI	GAAGAC	no cuts
BglII	AGATCT	no cuts
BsaI	GGTCTC	no cuts
BsiWI	CGTACG	no cuts
BsmBI	CGTCTC	no cuts
BstBI	TTCGAA	no cuts
ClaI	ATCGAT	no cuts
EagI	CGGCCG	no cuts
FseI	GGCCGGCC	no cuts
HindIII	AAGCTT	no cuts

Name	Sequence	5' Cut Positions
KpnI	GGTACC	no cuts
MfeI	CAATTG	no cuts
NdeI	CATATG	no cuts
NheI	GCTAGC	no cuts
NotI	GCGGCCGC	no cuts
PacI	TTAATTAA	no cuts
PmeI	GTTTAAAC	no cuts
PpuMI	RGGWCCY	no cuts
PspXI	VCTCGAGB	no cuts
PstI	CTGCAG	no cuts
SacI	GAGCTC	no cuts
SalI	GTCGAC	no cuts
SanDI	GGGWCCC	no cuts
SapI	GCTCTTC	no cuts
SbfI	CCTGCAGG	no cuts
SfiI	GGCCNNNNNGGCC	no cuts
SnaBI	TACGTA	no cuts
SpeI	ACTAGT	no cuts
SphI	GCATGC	no cuts
SwaI	ATTTAAAT	no cuts
XhoI	CTCGAG	no cuts

Sequence

```
   1 CCCGTAGAAA AGATCAAAGG ATCTTCTTGA GATCCTTTTT TTCTGCGCGT AATCTGCTGC
  61 TTGCAAACAA AAAAACCACC GCTACCAGCG GTGGTTTGTT TGCCGGATCA AGAGCTACCA
 121 ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA TACCAAATAC TGTTCTTCTA
 181 GTGTAGCCGT AGTTAGCCCA CCACTTCAAG AACTCTGTAG CACCGCCTAC ATACCTCGCT
 241 CTGCTAATCC TGTTACCAGT GGCTGCTGCC AGTGGCGATA AGTCGTGTCT TACCGGGTTG
 301 GACTCAAGAC GATAGTTACC GGATAAGGCG CAGCGGTCGG GCTGAACGGG GGGTTCGTGC
 361 ACACAGCCCA GCTTGGAGCG AACGACCTAC ACCGAACTGA GATACCTACA GCGTGAGCTA
 421 TGAGAAAGCG CCACGCTTCC CGAAGGGAGA AAGGCGGACA GGTATCCGGT AAGCGGCAGG
 481 GTCGGAACAG GAGAGCGCAC GAGGGAGCTT CCAGGGGGAA ACGCCTGGTA TCTTTATAGT
 541 CCTGTCGGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT TGTGATGCTC GTCAGGGGGG
 601 CGGAGCCTAT GGAAAAACGC CAGCAACGCG GCCTTTTTAC GGTTCCTGGC CTTTTGCTGG
 661 CCTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT CTGTGGATAA CCGTATTACC
 721 GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA CCGAGCGCAG CGAGTCAGTG
 781 AGCGAGGAAG CGGAAGGCGA GAGTAGGGAA CTGCCAGGCA TCAAACTAAG CAGAAGGCCC
 841 CTGACGGATG GCCTTTTTGC GTTTCTACAA ACTCTTTCTG TGTTGTAAAA CGACGGCCAG
 901 TCTTAAGCTC GGGCCCCCTG GGCGGTTCTG ATAACGAGTA ATCGTTAATC CGCAAATAAC
 961 GTAAAAACCC GCTTCGGCGG GTTTTTTTAT GGGGGGAGTT TAGGGAAAGA GCATTTGTCA
1021 GAATATTTAA GGGCGCCTGT CACTTTGCTT GATATATGAG AATTATTTAA CCTTATAAAT
```

```
1081 GAGAAAAAAG CAACGCACTT TAAATAAGAT ACGTTGCTTT TTCGATTGAT GAACACCTAT
1141 AATTAAACTA TTCATCTATT ATTTATGATT TTTTGTATAT ACAATATTTC TAGTTTGTTA
1201 AAGAGAATTA AGAAAATAAA TCTCGAAAAT AATAAAGGGA AAATCAGTTT TTGATATCAA
1261 AATTATACAT GTCAACGATA ATACAAAATA TAATACAAAC TATAAGATGT TATCAGTATT
1321 TATTATCATT TAGAATAAAT TTTGTGTCGC CCTTAATTGT GAGCGGATAA CAATTACGAG
1381 CTTCATGCAC AGTGAAATCA TGAAAAATTT ATTTGCTTTG TGAGCGGATA ACAATTATAA
1441 TATGTGGAAT TGTGAGCGCT CACAATTCCA CAACGGTTTC CCTCTAGAAA TAATTTTGTT

1501 TAACTTTTAA GGAGGTAAAA AATGCATCAT CATCATCACC ACATGGCCAG CAAGGGCGAA
1561 GAACTGTTTA CGGGCGTCGT TCCGATCCTG GTTGAGCTGG ATGGTGATGT GAACGGCCAC
1621 AAGTTCAGCG TTTCCGGTGA GGGTGAGGGC GATGCTACCT ATGGTAAACT GACGCTGAAA
1681 TTCATTTGCA CGACTGGTAA GCTGCCGGTC CCATGGCCGA CCCTGGTCAC GACCTTTTCC
1741 TATGGTGTTC AGTGTTTCAG CCGTTATCCG GATCACATGA AACAACATGA CTTCTTTAAG
1801 AGCGCGATGC CTGAGGGTTA CGTGCAAGAG CGCACTATTA GCTTCAAAGA TGACGGTAAC
1861 TACAAAACCC GTGCAGAAGT CAAGTTTGAG GGCGACACCT TGGTGAATCG CATCGAGCTG
1921 AAGGGCATTG ACTTCAAGGA AGATGGCAAC ATCCTGGGTC ATAAACTGGA GTACAACTAC
1981 AATTCTCATA ACGTTTACAT TACCGCGGAT AAACAGAAAA ATGGCATTAA GGCCAACTTT
2041 AAGATTCGTC ACAACATCGA AGATGGCTCG GTGCAGTTGG CAGACCACTA TCAACAAAAT
2101 ACCCCGATCG GTGACGGTCC GGTTCTGCTG CCGGATAATC ACTACCTGAG CACGCAAAGC
2161 GCGCTGAGCA AAGACCCGAA TGAAAAGCGT GACCACATGG TCCTGTTGGA ATTCGTGACC
2221 GCTGCGGGTA TCACCCACGG TATGGACGAA CTGTACAAAG AAAACCTGTA TTTTCAGGGT
2281 GACGTGTCTA CCAGCCAGGC AGTTCTGCCG GATGACTTCC CGCGCTACCC GGTGGGCAAA
2341 GCGTTTAAGT TCGACACGTG GCGTCAGAGC GCGGGTCGTC TGTAAGGTTG ACCCCAAGGG
2401 CGACACCCCA TAATTAGCCC GGGCGAAAGG CCCAGTCTTT CGACTGAGCC TTTCGTTTTA
2461 TTTGATGCCT GGCAGTTCCC TACTCTCGCA TGGGGAGTCC CCACACTACC ATCGGCGCTA
2521 CGGCGTTTCA CTTCTGAGTT CGGCATGGGG TCAGGTGGGA CCACCGCGCT ACTGCCGCCA
2581 GGCAAACAAG GGGTGTTATG AGCCATATTC AGGTATAAAT GGGCTCGCGA TAATGTTCAG
2641 AATTGGTTAA TTGGTTGTAA CACTGACCCC TATTTGTTTA TTTTTCTAAA TACATTCAAA
2701 TATGTATCCG CTCATGAGAC AATAACCCTG ATAAATGCTT CAATAATATT GAAAAAGGAA

2761 GAATATGAGT ATTCAACATT TCCGTGTCGC CCTTATTCCC TTTTTTGCGG CATTTTGCCT
2821 TCCTGTTTTT GCTCACCCAG AAACGCTGGT GAAAGTAAAA GATGCTGAAG ATCAGTTGGG
2881 TGCACGAGTG GGTTACATCG AACTGGATCT CAACAGCGGT AAGATCCTTG AGAGTTTTCG
2941 CCCCGAAGAA CGTTTTCCAA TGATGAGCAC TTTTAAAGTT CTGCTATGTG GCGCGGTATT
3001 ATCCCGTATT GACGCCGGGC AAGAGCAACT CGGTCGCCGC ATACACTATT CTCAGAATGA
3061 CTTGGTTGAG TACTCACCAG TCACAGAAAA GCATCTTACG GATGGCATGA CAGTAAGAGA
3121 ATTATGCAGT GCTGCCATAA CCATGAGTGA TAACACTGCG GCCAACTTAC TTCTGACAAC
3181 GATCGGAGGA CCGAAGGAGC TAACCGCTTT TTTGCACAAC ATGGGGGATC ATGTAACTCG
3241 CCTTGATCGT TGGGAACCGG AGCTGAATGA AGCCATACCA AACGACGAGC GTGACACCAC
3301 GATGCCTGTA GCGATGGCAA CAACGTTGCG CAAACTATTA ACTGGCGAAC TACTTACTCT
3361 AGCTTCCCGG CAACAATTAA TAGACTGGAT GGAGGCGGAT AAAGTTGCAG GACCACTTCT
3421 GCGCTCGGCC CTTCCGGCTG GCTGGTTTAT TGCTGATAAA TCCGGAGCCG GTGAGCGTGG
3481 TTCTCGCGGT ATCATCGCAG CGCTGGGGCC AGATGGTAAG CCCTCCCGTA TCGTAGTTAT
3541 CTACACGACG GGGAGTCAGG CAACTATGGA TGAACGAAAT AGACAGATCG CTGAGATAGG
3601 TGCCTCACTG ATTAAGCATT GGTAAGCGGC GCGCCATCGA ATGGCGCAAA ACCTTTCGCG
```

3661 GTATGGCATG ATAGCGCCCG GAAGAGAGTC AATTCAGGGT GGTGAATATG AAACCAGTAA

3721 CGTTATACGA TGTCGCAGAG TATGCCGGTG TCTCTTATCA GACCGTTTCC CGCGTGGTGA

3781 ACCAGGCCAG CCACGTTTCT GCGAAAACGC GGGAAAAAGT GGAAGCGGCG ATGGCGGAGC

3841 TGAATTACAT TCCCAACCGC GTGGCACAAC AACTGGCGGG CAAACAGTCG TTGCTGATTG

3901 GCGTTGCCAC CTCCAGTCTG GCCCTGCACG CGCCGTCGCA AATTGTCGCG GCGATTAAAT

3961 CTCGCGCCGA TCAACTGGGT GCCAGCGTGG TGGTGTCGAT GGTAGAACGA AGCGGCGTCG

4021 AAGCCTGTAA AGCGGCGGTG CACAATCTTC TCGCGCAACG CGTCAGTGGG CTGATCATTA

4081 ACTATCCGCT GGATGACCAG GATGCCATTG CTGTGGAAGC TGCCTGCACT AATGTTCCGG

4141 CGTTATTTCT TGATGTCTCT GACCAGACAC CCATCAACAG TATTATTTTC TCCCATGAGG

4201 ACGGTACGCG ACTGGGCGTG GAGCATCTGG TCGCATTGGG TCACCAGCAA ATCGCGCTGT

4261 TAGCGGGCCC ATTAAGTTCT GTCTCGGCGC GTCTGCGTCT GGCTGGCTGG CATAAATATC

4321 TCACTCGCAA TCAAATTCAG CCGATAGCGG AACGGGAAGG CGACTGGAGT GCCATGTCCG

4381 GTTTTCAACA AACCATGCAA ATGCTGAATG AGGGCATCGT TCCCACTGCG ATGCTGGTTG

4441 CCAACGATCA GATGGCGCTG GGCGCAATGC GCGCCATTAC CGAGTCCGGG CTGCGCGTTG

4501 GTGCGGATAT CTCGGTAGTG GGATACGACG ATACCGAAGA TAGCTCATGT TATATCCCGC

4561 CGTTAACCAC CATCAAACAG GATTTTCGCC TGCTGGGGCA AACCAGCGTG GACCGCTTGC

4621 TGCAACTCTC TCAGGGCCAG GCGGTGAAGG GCAATCAGCT GTTGCCAGTC TCACTGGTGA

4681 AAAGAAAAAC CACCCTGGCG CCCAATACGC AAACCGCCTC TCCCCGCGCG TTGGCCGATT

4741 CATTAATGCA GCTGGCACGA CAGGTTTCCC GACTGGAAAG CGGGCAGTGA CTCATGACCA

4801 AAATCCCTTA ACGTGAGTTA CGCGCGCGTC GTTCCACTGA GCGTCAGAC

Only the synthesized DNA fragment (in red) has been sequence verified. We do not guarantee the vector sequence.