2015

Expression and Characterization of Mouse Notch1 *Abruptex* EGF Repeats in *E. coli*

Amid Paudyal
Louisiana State University and Agricultural and Mechanical College

EXPRESSION AND CHARACTERIZATION OF MOUSE NOTCH1 *ABRUPTEX* EGF REPEATS IN *E. COLI*.

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science

in

The Department of Chemistry

by
Amid Paudyal
B.S., Nicholls State University, 2010
May 2015

This thesis is dedicated to my family and friends for their continued love and support.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Prof. Megan Macnaughtan, for her continued guidance and support throughout this journey. I would also like to thank my committee members, Prof. Kermit Murray and Prof. Jayne Garno, for their advice and support. I also thank my supervisor at the LSU NMR Facility, Dr. Thomas Weldeghiorghis, for all his help and support.

I thank the Macnaughtan Research Group members for their invaluable friendship, help, and support during my stay at LSU.

I would like to thank my family and friends for their continued love and support, without which this would never have been possible. I especially would like to thank my parents and my brother for all that they have done for me, and will always be grateful for that.

Finally, I would like to thank the LSU Department of Chemistry for giving me an opportunity to pursue graduate studies in Chemistry, and for supporting my education through assistantships. I would also like to thank the National Science Foundation (NSF) for funding this project (award number: CHE-1413576).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Notch1 is a transmembrane protein involved in cell signaling, which is found to be linked to various genetic and neurodegenerative diseases.[1-2] The extracellular domain of the Notch1 protein consists of 36 tandem epidermal growth factor-like (EGF) repeats, which are directly involved in ligand binding and are heavily glycosylated.[1] EGF 24-29 repeats lie in the *Abruptex* region of the Notch1 protein, and regulate the ligand binding activity through structural changes.[1] The role of glycosylation in ligand binding events and the overall structure of the *Abruptex* region are not known. Also, the role of calcium binding in Notch activation or inhibition is not clear. The goal of this project is to use NMR spectroscopy to study the structure of EGF domains with and without glycosylation, and a three dimensional structure of the *Abruptex* EGF repeats will be determined. The structures will provide a model to the *Abruptex* region of Notch1 and will further aid in our understanding of ligand binding activity in the extracellular domain of the Notch1.

CHAPTER 1: BACKGROUND AND SIGNIFICANCE

1.1 Proteins and Protein Structure

Proteins are large biomolecules or macromolecules that play an important role in all living things. They are responsible for chemical processes occurring in living organisms, and play a large role in tissue and organ system regulation and function.[3-4] Formation of proteins begins with the transcription process, which transfers genetic information from deoxyribonucleic acid (DNA) to messenger-ribonucleic acid (mRNA). Translation of mRNA produces proteins as shown in *Figure 1*. Ribosomes and translational RNA (tRNA, a non-protein coding gene) bind to the mRNA (protein coding gene) in the cytoplasm of the cell to form an amino acid chain linked by the peptide bond, which then folds to form proteins.[4]

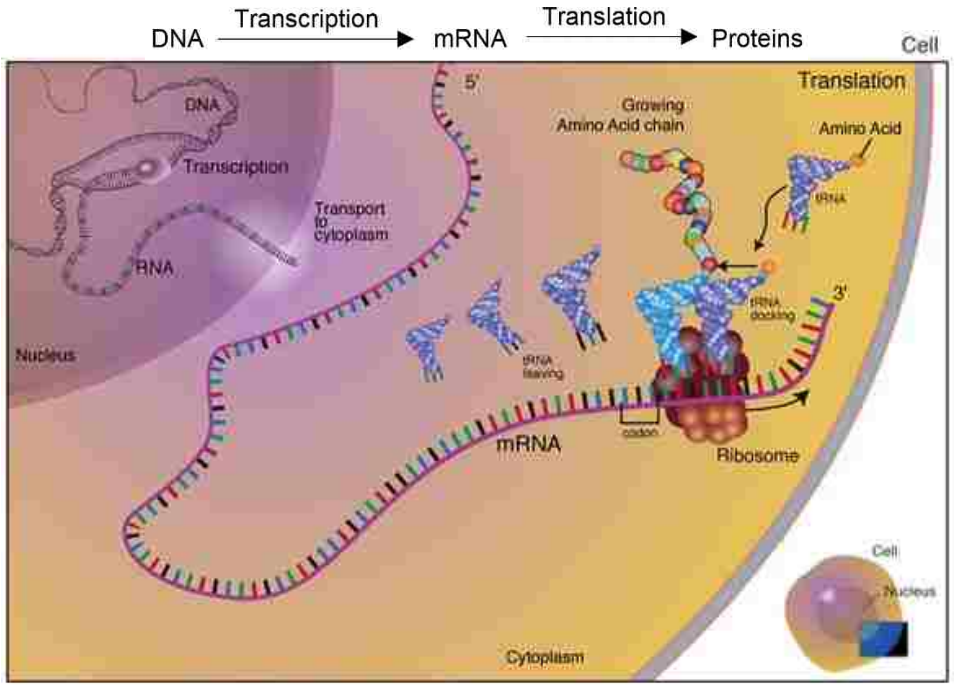

Figure 1. Schematic of protein synthesis in a living cell. DNA transcribes information to mRNA, which is then translated to form proteins. During the transcription process, a double-helix anti-parallel strand of DNA carries genetic information. DNA transcription makes an RNA copy of a gene sequence to form mRNA in the nucleus of a cell. The mRNA is a single-stranded RNA molecule, which is complementary to one of the DNA strands. The mRNA is then transported from the nucleus to the cytoplasm of the cell where proteins are formed by translation. During translation, the ribosome reads the mRNA triplet base codon and translates it into a specific amino acid. During this process, a tRNA molecule confirms the newly formed amino acid in a protein sequence by attaching its three complementary bases with an mRNA codon forming a base pair. http://www.genome.gov/glossary/

Proteins are made of a polypeptide chain comprising 20 different naturally occurring amino acids, which are the building blocks. An amino acid backbone is made of a central carbon atom (C_α) attached to an amino group ($–NH_2$), a carboxyl group (–COOH), and a side chain (–R) that varies from one amino acid to another as shown in *Figure 2 (a)*. Two amino acid residues are joined together by a peptide bond formed by the dehydration reaction that occurs between the amino group of one amino acid residue and the carboxyl group of another amino acid residue to form a polypeptide chain as shown in *Figure 2 (b)*. Each of the 20 different amino acids has a specific side-chain that governs the chemistries of the polypeptide chain to give a shape or conformation to the protein molecule. The different types of side chain that exist in a protein molecule are electrically charged (either positively or negatively charged), polar uncharged, and hydrophobic side chains that usually form non-covalent bonds. Some amino acid side chains have special properties, such as: (1) cysteine, which can form covalent disulfide bonds between two cysteine residues, (2) proline, which has a cyclic bond with the amino group in the backbone, and (3) glycine that is the simplest and one of the most stable naturally occurring amino acids.[4-5] Proteins are further classified into different levels according to their structure.

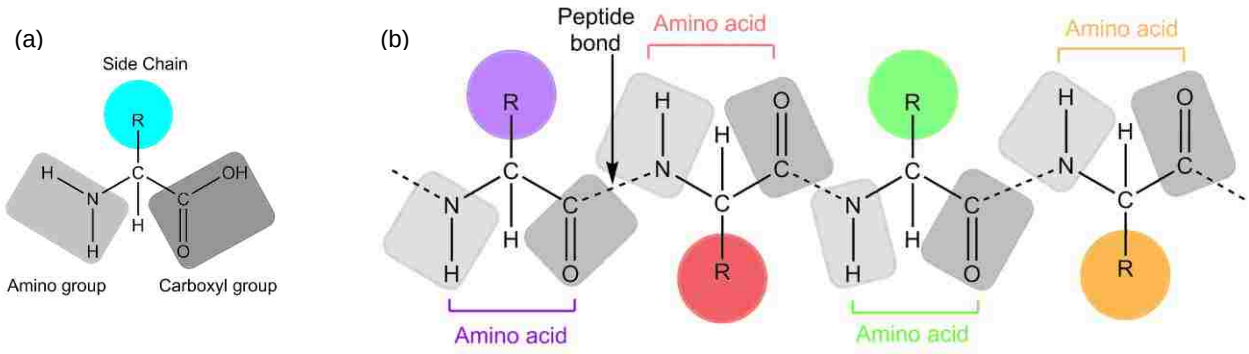

Figure 2. (a) Amino acid backbone consisting of an amino group and a carboxyl group with a variable side chain. (b) Amino acid residues linked together by a peptide bond forming a polypeptide chain.[2]

A linear polypeptide chain of amino acids gives the primary structure of a protein. Hydrogen bonding between the amino acid residues in a protein sequence forms alpha helices and beta (pleated) sheets giving the protein its secondary structure. A tertiary structure is formed

when the alpha helices and beta sheets arrange in a pattern or a fold to give a certain conformation. The structure containing multiple polypeptide chains arranges to form the quaternary structure of a protein.[4-5] *Figure 3* gives a representation of each of these levels of protein structure. Besides these different structures, proteins undergo several modifications once they are formed.

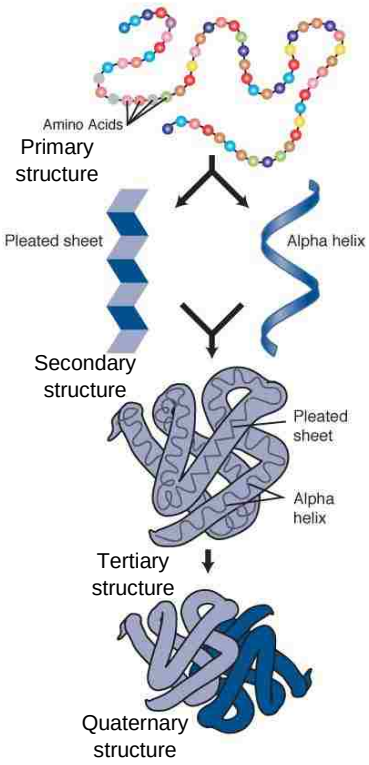

Figure 3. Levels of protein structure showing primary structure, secondary structure, tertiary structure, and quaternary structure. The primary structure is made of a linear polypeptide chain of amino acids. The secondary structure forms when the amino acids are linked together to form alpha helices and beta (pleated) sheets. These secondary structures then attract together to form a tertiary structure. Multiple amino acid chains then form the quaternary structure. http://www.genome.gov/glossary/

1.2 Post-translational Modifications

After translation, proteins further undergo post-translational modification (PTM). PTMs affect several cellular processes in living organisms such as cell cycle, differentiation, proliferation, and apoptosis.[6] PTMs occur either by the covalent addition of a functional group to a protein or a proteolytic cleavage. More than 200 PTMs have been identified. Some of the most

common PTMs that occur are methylation, acetylation, phosphorylation, and glycosylation.[7] Protein phosphorylation is a well-studied modification commonly occurring on serine, threonine and tyrosine when a phosphate group is added to these residues. Protein glycosylation is one of the most common modifications, in which glycans or sugar residues are attached to the functional group of a protein molecule. Protein glycosylation generally occurs as *O*-linked glycosylation and *N*-linked glycosylation. In *O*-linked glycosylation, glycans are covalently bound via the oxygen atom in the hydroxyl group of serine and threonine residues. In *N*-linked glycosylation, glycans are covalently bound to nitrogen of asparagine and arginine residues.[8-9] Protein glycosylation affects protein folding and stability.[7, 10] Protein methylation is the addition of a methyl group to an arginine or lysine amino acid residue.[11] Acetylation is the modification that occurs when an acetyl group is added to the *N*-terminal amino group.[12] One such protein that undergoes various PTMs is Notch1.

1.3 Notch1 Protein

Notch1 is a transmembrane receptor protein involved in cell signaling, which is vital for cell development and activity.[1] Dysregulation in Notch1 signaling has been found to be associated with various developmental disorders in humans including neurodegenerative diseases and cancer.[1-2] For this reason, the Notch1 signaling pathway plays an important role in human genetics regulating cell fate by maintaining a balance between cell differentiation, proliferation, and apoptosis.[13] The Notch1 receptor consists of an extracellular domain (ECD) and an intracellular domain (ICD) as shown in *Figure 4*. The ECD of Notch1 is comprised of tandem epidermal growth factor-like (EGF) repeats. The human Notch1 receptor has two ligands: Delta-like-1 and Jagged-1. Notch1 ligands incorporate a receptor-binding Delta-like-1 or Jagged-1 domain with a variable number of (6-8) tandem EGF-like repeats, and are expressed on the cell surface.[13-15] A cell signal is transmitted when the ECD binds ligands from the signal-sending cells

and passes an intercellular signal to the signal-receiving cells via proteolytic cleavage of the ICD as shown in *Figure 4 (b)*.

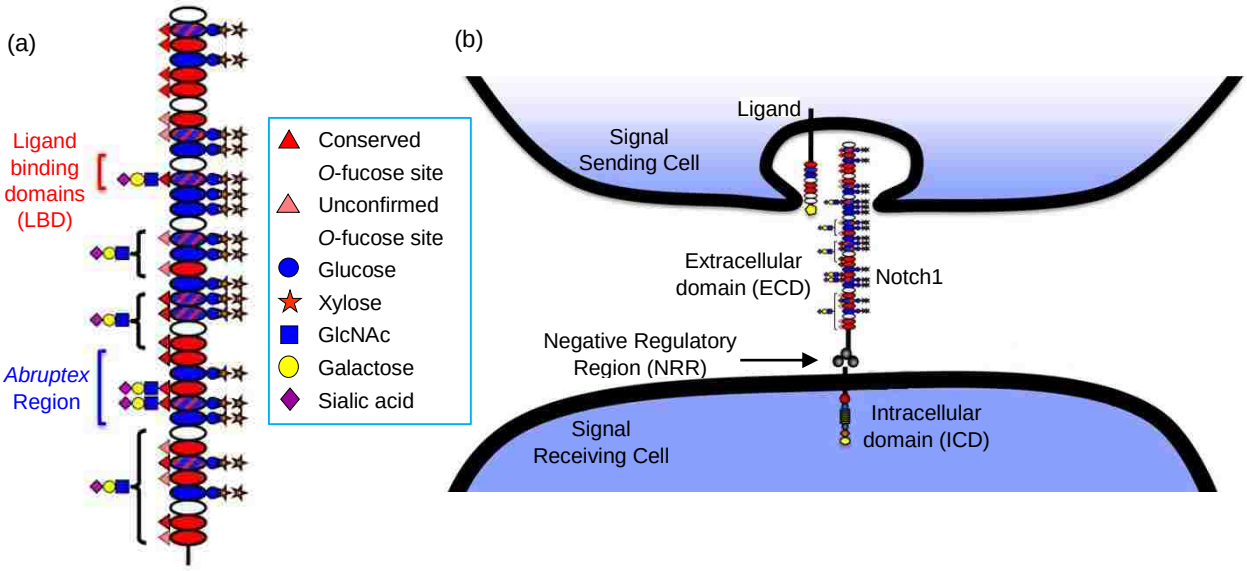

Figure 4. (a) Schematic of the mouse Notch1 extracellular domain (ECD) receptor. The ECD has 36 epidermal growth factor-like (EGF) repeats represented by ovals. EGF 11-12 is the ligand binding domain, and EGF 24-29 is the *Abruptex* region. The red, blue, or red-blue lined ovals represent the consensus sequence for *O*-fucosylation, *O*-glucosylation, or both, respectively. (b) Schematic of the Notch signaling pathway. The signal sending cell transmits genetic information via DSL (Delta, Serrate, Lag-2) ligand interaction with the Notch ECD. The signal further passes down to the NRR and ICD in the signal receiving cell. Reproduced from (Rana, N. A.; Haltiwanger, R. S., Fringe benefits: Functional and structural impacts of *O*-glycosylation on the extracellular domain of Notch receptors. *Current Opinion in Structural Biology* **2011**, *21* (5), 583-589.) with permission from the publisher.

1.4 Notch1 Extracellular Domain and EGF-like Repeats

The Notch1 extracellular domain (ECD) is an important segment of the Notch1 receptor, which is responsible for transmitting signals between cells. The Notch1 ECD is made of 36 tandem EGF-like repeats and a negative regulatory region (*Figure 4*). An EGF-like repeat is a small protein with 40 amino acid residues that consist of six-conserved cysteine (Cys) residues, which forms three disulfide bonds bridged by Cys^1–Cys^3, Cys^2–Cys^4, and Cys^5–Cys^6 as shown in *Figure 5*.[1, 16] Each EGF-like repeat is represented by an oval, as shown in *Figure 4*, and numbered from 1 to 36 (top to bottom). Once the ligand binds, each of these EGF-like repeats then help to transmit the cell signal downstream. EGF 11-12 of the Notch1 ECD is the site where Delta-like-1 and

Jagged-1 ligands interact and is called the ligand binding domain. EGF 24-29 is a flexible region called the *Abruptex* region that regulates ligand binding specificity.[1, 17]

1.4.1 Ligand Binding

Another important domain for cell signaling within the Notch 1 ECD is the *Ligand Binding Domain* that recognizes the ligand and transmits the signal downstream ultimately to the ICD.

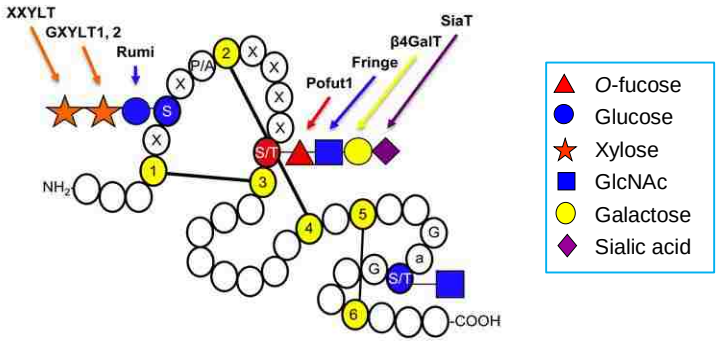

Figure 5. Schematic of Notch1 epidermal growth factor (EGF)-like repeat forming three disulfide bonds between six conserved cysteine residues in a 1-3, 2-4, 5-6 pattern. Each circle represents an amino acid residue: Yellow circles numbered 1-6 are the six conserved cysteine residues. The circle with the letter 'X' may be any amino acid residue and the letter 'a' is an aromatic residue. The blue circle with a letter S is a serine residue and is a site for *O*-glucosylation. The red circle with a letter S/T is a serine or threonine residue and is a site for *O*-fucosylation. The circle with a letter P/A is a proline or alanine residue, and the circle with a letter G is a glycine residue. An *O*-GlcNAc modification can occur between cysteine 5-6 at the serine or threonine residue. A sugar is added to an amino acid residue or glycan by a specific glycosyltransferase enzyme represented by an arrow. Reproduced from (Rana, N. A.; Haltiwanger, R. S., Fringe benefits: Functional and structural impacts of *O*-glycosylation on the extracellular domain of Notch receptors. *Current Opinion in Structural Biology* **2011**, *21* (5), 583-589.) with permission from the publisher.

Ligand binding is regulated by interactions between the *Abruptex* region and the *Ligand Binding Domain* of the Notch ECD.[18] Delta/Serrate/Lag-2 (DSL) domains are the disulfide-rich ligands that are involved in Notch activation by trans-ligand binding or inhibition by cis-ligand binding events.[19-20] *Drosophila* have only one Notch receptor with two ligands, Delta and Serrate, whereas mammalian cells have four Notch receptors and five ligands, Delta-like-1, Delta-like-3, Delta-like-4, Jagged-1, and Jagged-2.[15] EGF25-29 repeats lie in the *Abruptex* region of the Notch1 protein and are believed to regulate ligand binding through structural changes.[1] The activation or inhibition of Notch signaling is also regulated by calcium binding and glycosylation of the EGF

repeats.[15, 19, 21] Several studies have explained the Notch1 activation, which starts with ligand binding to the Notch1 ECD that eventually triggers the proteolytic cleavage of the Notch ICD.[22] Therefore, Notch1 activation is dependent on the binding of the ligands with the Notch1 ECD.

1.4.2 Role of Calcium Binding

Many EGF-like repeats of the Notch1 ECD are capable of binding calcium ions, which help to maintain heterodimeric stability between the ligand and the Notch1 ECD during the cell signaling process.[23] A heterodimer consist of two different polypeptide chains that are linked by non-covalent interactions. This heterodimer formation helps to stabilize the large ECD with the membrane bound ICD.[24] Calcium regulates Notch protein-protein interactions as well as its activation or inhibition by interacting with the DSL ligands.[15, 23] The *Ligand Binding Domain* (EGF 11-12) of Notch1 has a conserved calcium binding site between the domains. Calcium binding provides rigidity and helps to align the tandem EGF-like repeats in a rod-like shape.[25-26] EGF 24-29 is the *Abruptex* region, which is a highly glycosylated and flexible region of the Notch1 ECD as it lacks calcium binding sites between EGF 27-29.[1, 17] Furthermore, attachment of calcium between the tandem EGF domains helps to stabilize the Notch structure.[27]

1.4.3 Roles of Glycosylation

Glycosylation occurs when a sugar molecule attaches to the EGF domain of the Notch1 protein and subsequently regulates cell signaling. A properly folded EGF repeat can be modified by two types of *O*-glycosylation: *O*-fucosylation and *O*-glucosylation.[16] *O*-fucosylation involves the protein *O*-fucosyltransferase (Pofut1) enzyme that catalyzes the transfer of *O*-linked fucose from guanosine 5'-diphosphate (GDP) -fucose to EGF-like repeats. *O*-glucosylation involves the protein *O*-glucosyltransferase enzyme, Rumi/Poglut1, which catalyzes the transfer of *O*-linked glucose from uridine diphosphate (UDP) -glucose to EGF-like repeats.[16] Several studies in the past decade, primarily focusing on *O*-fucosylation of EGF repeats, have revealed its effect on Notch signaling activity.[1, 20, 28] Pofut1 is necessary for the proper folding and ligand interaction of

Notch.[29] A glycosyltransferase enzyme, Fringe, elongates the *O*-fucose residue by adding *N*-acetylglucosamine (GlcNAc), which directly affects the Notch-ligand interactions.[29-30] A study in *Drosophila* has shown that the addition of *O*-fucose to Notch EGF12 suppressed the Notch signaling activity; however, several other studies have shown that *O*-fucosylation helps in ligand mediated Notch activation.[28, 31] In addition, a recent study in *Drosophila* shows that *O*-glucose alters Notch ECD conformation and regulates the activation of Notch.[29]

1.5 EGF-like Repeat Structures, Calcium Binding, and Glycosylation

EGF-like repeat structures form the Notch1 ECD, which regulates cell signaling. An EGF-like repeat of mouse Notch1 consist of amino acid residues that are directly involved in ligand binding and are heavily glycosylated.[1] The six cysteine residues in an EGF domain arrange in a 1–3, 2–4, 5–6 pattern to form three disulfide bonds, which predominantly define its tertiary structure.[23, 25] Calcium binding and glycosylation, which regulate Notch signaling, occur at EGF domains in a conserved manner.

The EGF-like repeats form rod and U-shaped structures in the presence or absence of calcium, respectively.[19, 26, 32-34] The tandem EGF repeats are aligned and are made rigid by calcium binding, and are more flexible in the absence of binding as shown in *Figure 6 (a)* and *(b)*.[25-26] Calcium binding occurs at the linker region of the two tandem EGF repeats that follow a

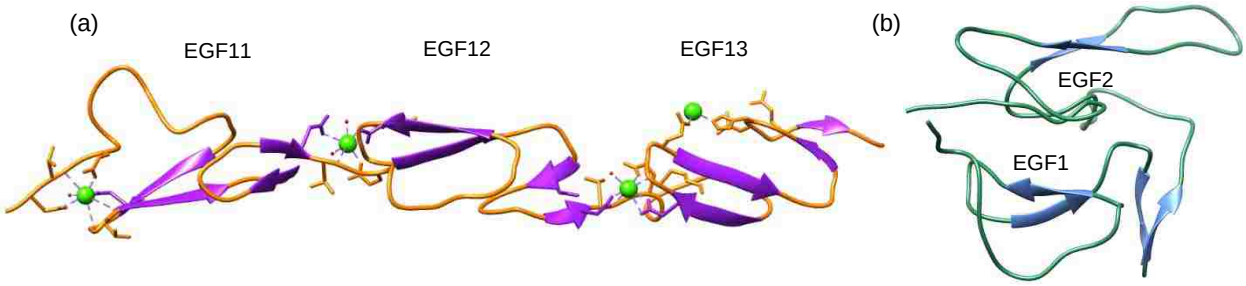

Figure 6. (a) The structure of human Notch1 EGF11-13 (PDB: 2VJ3)[17] showing coils (orange) and strands (purple). Green spheres represent calcium residues bound at the linker region between two tandem EGF-like repeats. Calcium binding gives a distinct rod-shaped structure. (b) The structure of *Plasmodium falciparum* merozoite surface protein 1 (PDB: 1CEJ)[29] showing coils (green) and strands (blue). This protein lacks calcium and is more flexible giving a “U-shaped” structure. The structures were obtained from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB).

consensus sequence for calcium binding. The consensus calcium binding site for the Notch EGF-like repeat follows the sequence (D/N/E)*X*(D/N/E)(E/D/Q)C^1. . .C^3*X*(D/N)*XXXX*(Y/F)*X*C^4, where "/" represents "or", X can be any amino acid residue, and C^1 to C^4 are the first through fourth conserved cysteine residues.[17]

Glycosylation helps to regulate cell signaling by providing the modifications needed for the ligand-binding domain on the properly folded Notch 1 protein to recognize the ligand. Protein folding occurs in the endoplasmic reticulum (ER) of the cell during protein synthesis.[35] A properly folded protein in its native state has its disulfide bonds correctly formed in the ER during protein biosynthesis.[16] An EGF repeat in its native state can be modified by *O*-glucosylation and *O*-fucosylation in the presence of glycosyltransferase enzymes. Several studies done on Notch glycosylation have developed a consensus sequence for the addition of *O*-glucose and *O*-fucose on the EGF domains. The consensus *O*-glucose site on Notch EGF domain is C^1*X*(S)*X*(P/A)C^2 and *O*-fucose site is C^2*XXXX*(S/T)C^3, where "/" represents "or", X can be any amino acid, and C^1 to C^3 are the first through third conserved cysteine residues.[16-17]

1.6 Unknowns

While calcium binding and glycosylation both affect the Notch signaling activity, the molecular mechanism by which Notch is activated or inhibited is still not clear. Also, the effect of glycosylation on inter-domain structure is not known. Since Notch regulation is directly associated with cell signaling for cell growth and development, it is important to understand the molecular mechanism governing the process. Furthermore, Notch ligands are promising therapeutic targets due to their importance in neuroscience, cancer, angiogenesis, and stem cell biology is prominent.[36]

CHAPTER 2. EXPRESSION AND CHARACTERIZATION OF MOUSE NOTCH1 *ABRUPTEX* EGF-LIKE REPEATS IN *E. COLI*

2.1 Introduction

2.1.1 Notch1 *Abruptex* EGF-like Repeats

Notch1 is a transmembrane receptor protein involved in cell signaling. Signal transduction between adjacent cells occur via Notch extracellular domain (ECD) after binding with Delta-like-1 or Jagged-1 ligands. The ECD of Notch1 consists of 36 tandem EGF-like repeats, where glycosylation and calcium binding occurs. The *Abruptex* region of the Notch1 ECD is one of the sites for glycosylation, which plays an important role in Notch regulation. Notch1 EGF repeats also have consensus sites for calcium binding. Calcium coordination occurs at the linker region between two adjacent repeats and provides rigidity to the inter-repeat structures. The rigid rod-shaped inter-repeat structures preclude flexibility, which is important for protein stability and functioning. Some studies in the past have suggested that the *Abruptex* region of Notch1 competes with the *Ligand binding domain* for the receptor activation.[37] However, the molecular mechanism by which the *Abruptex* region affects Notch regulation is not fully understood.

2.1.2 Research Goal

The immediate goal of this research is to express and purify properly folded *Abruptex* EGF repeats in milligram quantities for NMR studies. The long term goals of this project are to determine the three dimensional structures of the EGF repeats in the *Abruptex* region of mouse Notch1 with and without glycosylation to understand the role of the glycans in EGF structure. The results will help to further understand how Notch1 signaling is regulated and will aid in the discovery of new drug targets in the future.

2.2 Materials and Methods

The preliminary step to attain the long-term goals of this project is described in this thesis, which is to express, purify, and characterize properly folded EGF repeats in milligram quantities.

2.2.1 Materials

The pET-20b-pelB-EGF27 plasmid was obtained from Dr. Robert Haltiwanger (Stony Brooke University, NY). pET-28a plasmid, Novagen Bugbuster 10X protein extraction reagent, and MF – membrane filters 0.025µm VSWP were purchased from EMD Millipore. NcoI, XhoI, and NdeI restriction enzymes, T4 DNA ligase, Taq 5x Master Mix, 2-Log DNA Ladder, NEB 10-beta electrocompetent *E. coli* cells, NEB 5-alpha competent *E. coli* cells, SOC outgrowth medium, and Shuffle T7 Express competent *E. coli* cells were purchased from New England BioLabs Inc. Sodium chloride (NaCl), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), urea, imidazole, tris(hydroxymethyl)aminomethane (Tris), sodium azide, 2-(*N*-morpholino)ethanesulfonic acid hydrate (MES), calcium chloride, tricine, acetonitrile, trifluoroacetic acid (TFA), potassium phosphate monobasic (KH_2PO_4), magnesium sulfate ($MgSO_4$), D-glucose monohydrate, ammonium chloride, thiamin hydrochloride, biotin, choline chloride, folic acid, nicotinamide, D-pantothenic acid hemicalcium salt, pyridoxal hydrochloride, riboflavin, tetracycline hydrochloride, sodium dodecyl sulfate (SDS), 2-mercaptoethanol (βME), acetic acid, Coomassie brilliant blue R 250, and 3-(trimethylsilyl)-1-propanesulfonic acid sodium salt (DSS) were purchased from Sigma-Aldrich/Fluka. Profinity immobilized metal affinity chromatography (IMAC) resin, Bio-Gel P-4 gel, an ENrich SEC 70 10x300 mm column, Precision Plus Protein Dual Xtra Standard, 4–20% Tris-HCl Mini-PROTEAN TGX precast gels, 10–20% Mini-PROTEAN Tris-Tricine precast gels, 2x Laemmli sample buffer, and tricine sample buffer were purchased from Bio-Rad. Agarose, sodium phosphate dibasic heptahydrate ($Na_2HPO_4.7H_2O$), Pierce Protease inhibitor tablets – EDTA free, EZ-Run *Rec* Protein Ladder, Coomassie Plus (Bradford) protein assay reagent, and bovine serum albumin (BSA) standard were purchased from Fisher/Thermo Scientific. Difco Luria-Bertani (LB) broth – Miller (developed by Miller) and LB Agar – Lennox (developed by Lennox) were purchased from BD Biosciences. Shigemi NMR tubes, BDH methanol, and centrifuge tubes were purchased from VWR International. ^{15}N labeled – ammonium chloride and deuterated water (D_2O) were purchased from

Cambridge Isotope Laboratories. Other chemicals used were isopropyl β-D-1-thiogalactopyranoside (IPTG) from Anatrace, phenylmethylsulfonyl fluoride (PMSF) from Calbiochem, QIAprep spin miniprep kit from Qiagen, Costar 96 well UV transparent flat bottom plates from Corning, kanamycin sulfate from Genlantis, glycine from Amresco, BL21-Gold(DE3) competent cells from Agilent Technologies, and primers from Integrated DNA Technologies (IDT). A Millipore Direct-Q 3 ultrapure water system was used as a source for all experiments.

2.2.2 Cloning

A pSec Tag plasmid containing the mouse Notch1 (mN1) EGF1-36 nucleotide sequence was obtained from Dr. Robert Haltiwanger at Stony Brook University, NY. The mN1 EGF1-36 protein has 1408 amino acid residues. To produce small protein constructs of the *Abruptex* region (EGF24-29), nucleotide sequences were cloned from the pSec Tag plasmid into other expression plasmids. The sequences for all plasmids made and used in this study are included in the *Appendix* section. The amino acid residues for each plasmid are numbered based on the mN1 EGF1-36 sequence. The EGF26-27-28 sequence (residues E961-S1078 based on mN1 EGF1-36 amino acid sequence) was cloned into the pET-28b expression plasmid with a *C*-terminal 6XHis tag (pET-28b-EGF26-27-28-6XHis). The EGF25-26-27 sequence (residues E923-V1040 based on mN1 EGF1-36 amino acid sequence) was cloned into the pET-28b expression plasmid with a *C*-terminal 6XHis tag (pET-28b-EGF25-26-27-6XHis). The EGF25 sequence has an extra cysteine residue, so the residue count at the 21st position in EGF25-26 sequence (or residue number 943 in the mN1 EGF1-36 sequence) is a cysteine, which was mutated to glycine. The EGF25-26-C21G sequence (residues E924-Q999) was cloned into the pET-28a expression plasmid with a *C*-terminal 6XHis tag (pET-28a-EGF-25-26-C21G-6XHis). The EGF25-26-C21G sequence (residues E924-Q999) was cloned with a TEV cleavage site into the pLIC-EGFP expression plasmid with a *C*-terminal 6XHis tag (pLIC-EGF25-26-C21G-TEV-EGFP-6XHis). The EGF27 (residues D1001-W1042) with a pelB leader sequence (pelB-EGF27) was cloned with a

TEV cleavage site into the pLIC-EGFP expression plasmid with a *C*-terminal 6XHis tag (pLIC-pelB-EGF27-TEV-EGFP-6XHis). The EGF27 (residues D1001-W1042) without pelB was cloned with a TEV cleavage site into the pLIC-EGFP expression plasmid with a *C*-terminal 6XHis tag (pLIC-EGF27-TEV-EGFP-6XHis). All of the cloning work mentioned above was performed by the LSU AgCenter Biotechnology Laboratory (ABL) Protein Facility.

Table 1. EGF constructs with their residue count based on EGF1-36 amino acid sequence.

Plasmid	**Residue count**
EGF1-36	G1-L1408
EGF26-27-28	E961-S1078
EGF25-26-27	E923-V1040
EGF25-26-C21G	E924-Q999
EGF27	D1001-W1042

A sub-cloning experiment was performed to remove EGF27 (residues D1001-W1042) without the pelB leader sequence from pET-20b-pelB-EGF27-6XHis. The EGF27 insert was sub-cloned into the pET-28a expression plasmid to construct the pET-28a-EGF27-6XHis plasmid. The pET20b-pelB-EGF27-6XHis containing the insert with a *C*-terminal 6XHis tag and the pET-28a expression plasmid were double digested separately using the restriction enzymes NcoI and XhoI. A 50 µL double digestion reaction was prepared by mixing 2 µL (1 µg) of plasmid (vector or insert), 5 µL (1X) of 10X NEBuffer 3.1, 41 µL of autoclaved Millipore water, and 1 µL (10 units) each of restriction enzymes NcoI and XhoI. Restriction enzymes were the last thing added in the mixture (*Table 2*).

The two 50 µL double digestion reactions (vector and insert) were incubated at 37°C for one hour. The restriction enzymes were heat-inactivated immediately by incubating the reaction mixture at 80°C for 20 minutes. The EGF27 insert was ligated into the digested pET-28a vector by mixing 2 µL (1X) of 10X T4 DNA ligase buffer, 2 µL each of the double digested vector and

insert, 13 µL of autoclaved Millipore water, and 1 µL of T4 DNA ligase enzyme. T4 DNA ligase enzyme was the last thing added in the mixture (*Table 3*).

Table 2. Protocol for double digestion reaction (50 µL reaction).

Plasmid (vector or insert)	2 µL (1 µg)
10X NEBuffer 3.1	5 µL
Autoclaved Millipore water	41 µL
Restriction Enzyme – NcoI	1 µL (10 units)
Restriction Enzyme – XhoI	1 µL (10 units)

Table 3. Protocol for ligation reaction (20 µL reaction).

10X T4 DNA Ligase Buffer	2 µL
Vector plasmid (double-digested)	2 µL
Insert plasmid (double-digested)	2 µL
Autoclaved Millipore water	13 µL
T4 DNA Ligase	1 µL

The ligation reaction was incubated at 16°C for two hours. The ligase enzyme was heat-inactivated immediately by incubating the reaction mixture at 65°C for 10 minutes and chilled on ice immediately. A post-ligation cut was performed by using the restriction enzyme NdeI to remove incorrectly cloned plasmid. 2 µL of 10X NEBuffer 3.1 and 1 µL (10 units) of NdeI restriction enzyme were added to the completed ligation reaction from above and incubated at 37°C for one hour. The restriction enzyme was heat-inactivated immediately by incubating the reaction mixture at 65°C for 20 minutes and immediately chilled on ice. The pET28a-EGF27 plasmid was desalted into autoclaved Millipore water using a 0.025 µm VSWP membrane filter.

The sub-cloned pET28a-EGF27 plasmid (2.5 µL) was transformed into NEB 5-alpha competent *E. coli* cloning strain cells following NEB heat shock transformation protocol. 2.5 µL of plasmid was mixed with 50 µL of competent cells and was incubated on ice for 2 minutes. Without

mixing or disturbing, the mixture was heat shocked at exactly 42°C for 30 seconds, and was immediately placed on ice for another 2 minutes. 950 µL of room temperature SOC medium was added into the mixture, and the mixture was incubated at 37°C with 250 rpm shaking for 1 hour. The transformation mixture was plated on LB agar containing 50 µg/mL kanamycin and incubated overnight (16 hours) at 37°C.

Table 4. Reaction setup protocol for the polymerase chain reaction (PCR) (25 µL reaction).

10 µM Forward Primer	0.5 µL
10 µM Reverse Primer	0.5 µL
Autoclaved Millipore water	19 µL
Taq 5X Master Mix	5 µL

The presence of pET28a-EGF27 in the colonies was verified by polymerase chain reaction (PCR). Colony PCR was performed by selecting individual colonies from the freshly grown transformation plate. The forward primer used was IDT Readymade 51-01-16-07 (pET 3') 5'-CTAGTTATTGCTCAGCGG-3' and the reverse primer was IDT Readymade 51-01-16-08 (pET 5' or T7) 5'-TAATACGACTCACTATAGG-3'. The PCR reaction mixture consisting of 0.5 µL of 10 µM each of the forward and reverse primers, 19 µL of autoclaved Millipore water, and 5 µL of Taq 5X Master Mix was added together and chilled on ice. Taq 5X Master Mix was the last thing added in the mixture (*Table 4*). A small amount of a single bacterial colony from the transformation plate was picked with a sterile toothpick and dissolved in the PCR reaction mixture. The reaction mixture was transferred to a PCR cycler (Bio-Rad S1000 thermal cycler), and the PCR was performed. The cycling conditions include an initial 1 cycle denaturation at 95°C for 45 seconds, 30 cycles consisting of denaturation at 95°C for 45 seconds, annealing at 40°C for 30 seconds, extension at 72°C for 90 seconds, 1 cycle final extension at 72°C for 5 minutes, and 4°C hold until the instrument is turned off (*Table 5*).

The PCR reaction product was mixed with the DNA loading dye and the mixture was directly loaded onto a 1% agarose gel containing 0.2 µg/mL ethidium bromide for the DNA agarose gel electrophoresis analysis. After confirmation of the pET28a-EGF27 from the gel, a 10 mL starter culture of the selected colony was grown in LB broth media containing 50 µg/mL kanamycin. The starter culture was grown overnight at 37°C on an orbital shaker at 250 rpm. The plasmid was purified from the starter culture using the QIAGEN miniprep plasmid purification kit and was sequenced by the GeneLab in the LSU School of Veterinary Medicine. A 1 mL freezer stock in 15% glycerol was prepared by mixing 500 µL of the overnight starter culture of colony with the correct plasmid sequence with 500 µL of 30% sterile (autoclaved) glycerol stock and stored at -80°C immediately.

Table 5. PCR cycling conditions protocol.

Step	**Temperature**	**Time**
Initial Denaturation	95°C	45 seconds
30 Cycles	95°C	45 seconds
	40°C	30 seconds
	72°C	90 seconds
Final Extension	72°C	5 minutes
Hold	4°C	

2.2.3 Recombinant Protein Expression in *E. coli*

The EGF protein constructs were expressed by transforming the appropriate plasmid into BL21-Gold (DE3), Shuffle T7 Express, or Shuffle T7 Express *lysY* competent *E. coli* cells using a heat shock transformation protocol. Transformed colonies were grown on LB agar plates supplemented with the appropriate antibiotics. Unless otherwise stated, the working concentration of kanamycin (KAN) used is 50 µg/mL and tetracycline (TET) is 12.5 µg/mL.

For BL21-Gold (DE3) cultures, 10 mL starter cultures innoculated with single colonies were grown in LB broth media supplemented with appropriate antibiotics. The starter culture was grown at 37°C with 250 rpm shaking overnight. A 1 mL freezer stock in 15% glycerol was prepared as described previously and immediately stored at -80°C. A 1 L volume culture was grown by innoculating 1 L of LB broth with 10 mL of the starter culture and adding the appropriate antibiotics. The culture was grown at 37°C on an orbital shaker at 250 rpm until the optical density of the cells at 600 nm absorbance (OD_{600}) reached 0.5 – 0.6. The culture was induced with 0.4 mM IPTG and incubated overnight at 16°C or for 3 hours at 37°C.

For Shuffle T7 Express and Shuffle T7 Express *lysY* cultures, 10 mL starter cultures were grown in LB broth media supplemented with appropriate antibiotics by selecting a single colony from a freshly grown plate (either a new transformation or a freezer stock). The starter cultures were grown at 30°C with 250 rpm shaking overnight. Freezer stocks in 15% glycerol were prepared as described previously and immediately stored at -80°C. A 1 L volume culture was grown by innoculating 1 L of LB broth with 10 mL of the starter culture and adding the appropriate antibiotics. The cultures were grown at 30°C with 250 rpm shaking until the OD_{600} reached 0.4 – 0.5. The cultures were then induced with 0.4 mM IPTG and incubated overnight at 16°C or for 4 to 6 hours at 30°C.

For each expression, 1 mL samples of the culture were collected before and after induction for SDS-PAGE analysis. The cells were spun down at 4000 x g for 5 min on a benchtop centrifuge and the supernatant was discarded. The cell pellets were frozen at -20°C until sample preparation for analysis.

The 1 L cultures were harvested by centrifugation at 4000 x g for 10 min at 4°C using a Thermo Scientific Sorvall RC 6+ centrifuge with a FIBERLite F9-4x1000y rotor. The supernatant was discarded and the pellets were stored at -80°C until lysis.

Table 6. Recipe for 1 liter 10x M9 Salts.

$Na_2HPO_4•7H_2O$	128 g
KH_2PO_4	30 g
NaCl	5 g

Table 7. Recipe for 100 mL 1000x Trace Minerals (10 mM each).

$FeCl_3•6H_2O$	0.2703 g
$CuSO_4•5H_2O$	0.2497 g
$MnSO_4•H_2O$	0.1690 g
$ZnSO_4•7H_2O$	0.2875 g

Table 8. Recipe for 1 liter 200x Vitamins.

0.1% Thiamin hydrochloride	1 g
0.02% Biotin	200 mg
0.02% Choline chloride	200 mg
0.02% Folic acid	200 mg
0.02% Nicotinamide	200 mg
0.02% D-Pantothenic acid hemicalcium salt	200 mg
0.02% Pyridoxal hydrochloride	200 mg
0.002% Riboflavin	20 mg

Table 9. Recipe for 1 liter Minimal media.

10x M9 Salts	100 mL
1 M $MgSO_4$	2 mL
200x Vitamins	5 mL
1000x Trace Minerals	1 mL
D-Glucose monohydrate	4 g

Table 9 continued

^{15}N labeled NH_4Cl	1 g
Kanamycin (50 mg/mL)	1 mL
0.1 M $CaCl_2$	1 mL
Innoculum	10 mL

2.2.4 Cell Lysis and Protein Purification

Cells were lysed with 20 mL of lysis buffer (50 mM Tris, pH 6.9, 1 mM PMSF, 2 mL of 10x Bugbuster, and 20 µL Benzonase nuclease) followed by four cycles of mechanical lysis in a French press at 1,000 psi. The cell lysate was centrifuged at 30,000 x g for 30 minutes at 4°C using a Thermo Scientific Sorvall RC 6+ centrifuge with a FIBERLite F21-8x50y rotor. The 6X-His-tagged protein was purified from the cleared lysate with a Profinity IMAC Ni affinity chromatography column (4 mL) followed by desalting using a P-4 gel column or gel filtration/size-exclusion chromatography column (Hi-Prep 26/60 Sephacryl S-100 HR or ENrich SEC 70). The equilibration/wash buffer used for the IMAC column was 25 mM Tris, 0.5 M NaCl, 10 mM imidazole, pH 6.9 and the elution buffer was 25 mM Tris, 0.5 M NaCl, 500 mM imidazole, pH 6.9. The buffer used for the P-4 gel or gel filtration chromatography was 25 mM Tris, 150 mM NaCl, pH 6.9.

2.2.5 SDS-PAGE Gel Electrophoresis

The cell pellet samples from a culture and protein obtained after chromatographic purification were analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The cell pellet sample was resuspended completely in 50 µL of Millipore water, and about 1-15 µL of the mixture was used as a sample for SDS-PAGE analysis. The samples for SDS-PAGE using a Bio-Rad 4-20% TGX precast gels were prepared by mixing the sample and sample loading buffer in a 1:1 ratio. The sample loading buffer was prepared fresh each time by mixing 5% v/v βME to 95% v/v 2x Laemmli sample buffer. The samples for SDS-PAGE using a

Bio-Rad 10-20% Tris-Tricine precast gels were prepared by mixing the sample and sample loading buffer in a 1:2 ratio. The sample loading buffer was prepared fresh each time by mixing 2% v/v βME to 98% v/v Tricine sample buffer. The samples were heated at 95°C for 15 minutes and centrifuged briefly before loading into the gel well. The SDS-PAGE was run on a Bio-Rad gel electrophoresis kit at a constant voltage setting of 125V for 60 minutes with 25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3 for the 4-20% TGX gels and 100 mM tris, 100 mM tricine, 0.1% SDS, pH 8.3 for the tris-tricine gels. After the electrophoresis is complete, the gels were removed from the plastic cassette housing and stained with 10 mL of Coomassie stain solution (0.1% w/v Coomassie brilliant blue R 250 in 50% methanol, 40% water and 10% acetic acid) with a constant shaking for 1 hour at room temperature followed by rinsing in a destain solution (50% methanol, 40% water and 10% acetic acid) for 2 hours.

2.2.6 Reversed-phase HPLC

The properly folded and mis-folded protein was separated by purification using a reverse-phase high performance liquid chromatography (RP-HPLC) column (Purity Semi-Prep HPLC C18 column). A linear gradient from 0 to 75% solvent B (100% acetonitrile containing 0.1% TFA) in solvent A (water with 0.1% TFA) for 45 minutes at 2 mL/min flow rate was applied for eluting the protein from the column. The eluted protein was monitored at 214 and 280 nm absorbance. Fractions of eluents from individual peaks were collected and dried on a Speed-Vac (Thermo Savant SPD121 with UVS 400). All the significant peaks observed on RP-HPLC were analyzed by Bruker UltrafleXtreme MALDI-TOF/TOF mass spectrometry to confirm the molecular weight of the purified protein.

2.2.7 MALDI Mass Spectrometry

The protein obtained from RP-HPLC was analyzed using matrix-assisted laser desorption/ionization (MALDI) time of flight (TOF) mass spectrometry (MS). The MALDI matrix was prepared by mixing 30 mg of sinapinic acid in 1 mL of 30% acetonitrile, 70% water, and 0.1%

v/v TFA. The matrix solution was mixed using vortex for 10 minutes and centrifuged to separate the supernatant. MALDI samples were prepared by mixing 1 µL of purified protein sample with 1 µL of sinapinic acid matrix. In general, 10-50 pmol/µL protein sample is sufficient for MALDI analysis of proteins with molecular weights greater than 1 kDa and less than 20 kDa.[38] 1 µL of the sample mixture was pipetted on a MALDI target plate (Bruker Daltonics MTP 384 massive target T) and allowed to air dry. The sample spot on the target plate was analyzed with a Bruker UltrafleXtreme MALDI-TOF/TOF MS in the LSU Mass Spectrometry Facility. Spectra were recorded in positive ion linear mode with an ion acceleration potential of 25 kV. The ionization was performed using a Nd:YAG laser with 355 nm wavelength and 3 ns pulse. Laser shots (6500 shots) were summed together to generate each spectrum. The spectra were recorded in the mass range of 4000–25000 Da.

2.2.8 Coomassie (Bradford) Protein Assay

The protein concentration after chromatographic purification was measured using Coomassie Plus (Bradford) protein assay reagent. The samples were prepared in a UV transparent 96-well plate. A standard curve plot was drawn using a series of bovine serum albumin (BSA) standards at varying concentrations. The absorbance of each BSA samples was measured at 595 nm using a UV/vis spectrophotometer (BioTek PowerWave XS2) and plotted against the known concentration of BSA. The resulting line was fitted using the least squares method, and the final concentration of the unknown protein sample was calculated. The volume of each reagent used to make a standard curve plot is listed in *Table 10*.

Table 10. Volume of reagents needed to plot a standard curve for Coomassie protein assay.

	Volume in µL					
Coomassie Reagent	200	200	200	200	200	200
Protein Elution Buffer	50	40	30	20	10	40
BSA Standard (40 µg/mL)	0	10	20	30	40	0
Protein Sample	0	0	0	0	0	10

2.2.9 ^{1}H-^{15}N HSQC NMR to Detect Protein Folding and Assignments

^{1}H-^{15}N Heteronuclear single quantum coherence (HSQC) NMR spectroscopy was used to determine the foldedness of the protein. In order to obtain a good HSQC NMR signal, about 1 mM of the protein is needed. For example, for the EGF27 protein with a molecular weight of 7.1 kDa, nearly 2 mg of pure protein is needed to make 1 mM concentration in a 250 µL sample volume. The NMR sample for EGF27 protein was prepared in 10 % D_2O and 90% H_2O. 0.5 µL of 1% (w/v) DSS in D_2O standard was added for chemical shift referencing. The sample was transferred into a 5 mm Shigemi tube for running the NMR experiments. The NMR spectra were collected on a Varian 700 MHz spectrometer with a 5-mm-HCN-5922 probe at the LSU NMR Facility. A gNhsqc (sensitivity enhanced gradient ^{1}H-^{15}N HSQC) experiment was run with 16 scans and 256 increments. The 2D HSQC spectrum was processed on a Linux operating system using NMRPipe[39] software developed by the National Institute of Health. NMRView[40] software was used to visualize the spectra.

2.3 Results and Discussion

EGF constructs used in our study are listed in *Table 11* with their antibiotic resistance, properties, and the *E. coli* strains used for protein expression. The amino acid sequences of the expressed proteins and their molecular weights are listed in the appendix.

Table 11. Various EGF constructs with their properties.

Protein expressed	Plasmid name	Antibiotic resistance	Properties	Expression *E. coli* strain
EGF26-27-28-6XHis	pET28b-EGF26-27-28	Kan	With a C-terminal His-tag	BL21-Gold(DE3)
EGF25-26-27-6XHis	pET28b-EGF25-26-27	Kan	With a C-terminal His-tag	BL21-Gold(DE3)
EGF25-26-C21G-6XHis	pET28a-EGF25-26-C21G	Kan	EGF25-26 construct with an extra cysteine mutated to glycine; consists a C-terminal His-tag	BL21-Gold(DE3)

Table 11 continued

Protein expressed	Plasmid name	Antibiotic resistance	Properties	Expression *E. coli* strain
EGF25-26-C21G-TEV-EGFP-6XHis	pLIC-EGF25-26-C21G-TEV-EGFP-His	Kan	Green fluorescent protein fusion with a TEV cleavage site with C-terminal His-tag; an extra cysteine mutated to glycine	BL21-Gold(DE3), Shuffle T7 Express
pelB-EGF27-TEV-EGFP-6XHis	pLIC-pelB-EGF27-TEV-EGFP-His	Kan	Green fluorescent protein fusion with a TEV cleavage site with C-terminal His-tag	BL21-Gold(DE3), Shuffle T7 Express
EGF27-TEV-EGFP-6XHis	pLIC-EGF27-TEV-EGFP-His	Kan	Green fluorescent protein fusion with a TEV cleavage site with C-terminal His-tag	BL21-Gold(DE3), Shuffle T7 Express
EGF27-6XHis	pET28a-EGF27 (without pelB)	Kan	With a C-terminal His-tag	BL21-Gold(DE3), Shuffle T7 Express, Shuffle T7 Express *lysY*

2.3.1 Expression and Purification

EGF26-27-28:

EGF26-27-28-6XHis (EGF26-28) was expressed at 16°C overnight after induction with 0.5 mM IPTG in BL21-Gold(DE3) competent *E. coli* cells using the pET28b-EGF26-27-28 vector. The expression culture was analyzed by SDS-PAGE gel electrophoresis (*Figure 7*) to monitor expression levels. The molecular weight of EGF26-28 is 14.1 kDa. The first lane in *Figure 7* is a protein molecular weight marker, or the ladder, which is used for referencing the bands on the gel. The second lane is the whole cell sample before inducing the culture with IPTG. This lane serves as a background when analyzing protein being produced over time after induction. Lanes 3-6 are the whole cell samples after the culture was induced with IPTG. Lane 6 is the sample before harvesting the culture, which shows a significant dark band just above 15 kDa (marked by

an asterisk). This band suggests that EGF26-28 is being abundantly produced, or overexpressed, compared to the background *E. coli* proteins.

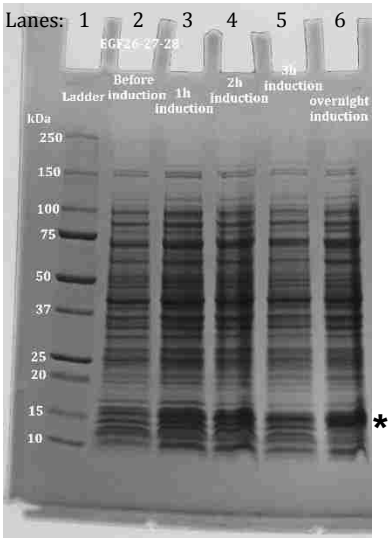

Figure 7. SDS-PAGE gel image of EGF26-28 expression culture. Samples include time point before induction (lane 2) and after induction (lane 3 = 1h, lane 4 = 2h, lane 5 = 3h, and lane 6 = overnight or approximately 16h).

EGF26-28 was purified from the cleared lysate using Profinity IMAC nickel affinity chromatography (*Figure 8a*) followed by gel filtration chromatography using a Hi-Prep 26/60 Sephacryl S-100 HR column (*Figure 8b*). The large elution profile with a sharp peak in *Figure 8 (a)* between 40 and 60 mL shows that the eluted protein from the Ni-affinity column is abundant. *Figure 8 (b)* shows the gel filtration chromatogram, which consists of two eluted peaks. Based on separations performed with molecular weight standards, the first eluting large peak at 140 mL (peak 1) is protein with relatively high molecular weight (>100 kDa). The second peak at 220 mL (peak 2) is consistent with the elution of monomeric EGF26-28.

The elution fractions of EGF26-28 from Ni-affinity and gel filtration chromatography were analyzed with SDS-PAGE. Samples were prepared with and without βME to evaluate the extent of intermolecular cross-linking. If the disulfide bonded protein gives a clear band at approximately the correct molecular weight in the absence of βME, then we can deduce that the protein is not cross-linked by erroneous intermolecular disulfide bonds. If a smear is observed throughout the lane instead of a clear band, then the protein is most likely cross-linked and in multimeric form. *Figure 8 (c)* shows the SDS-PAGE gel image of the Ni-elution with and without βME (lanes 2 and 3), peak 1 with and without βME (lanes 4 and 5), and peak 2 with and without βME (lanes 6 and 7). From the gel image in *Figure 8 (c)*, it is clear that the IMAC elution sample is a mixture of monomer EGF26-28 and cross-linked protein (lanes 2 and 3). The monomeric protein and cross-linked protein can be separated by gel filtration. Peak 1 is mostly cross-linked (lane 5). We can further deduce from the gel image that peak 2 is the monomeric form of the protein, which is

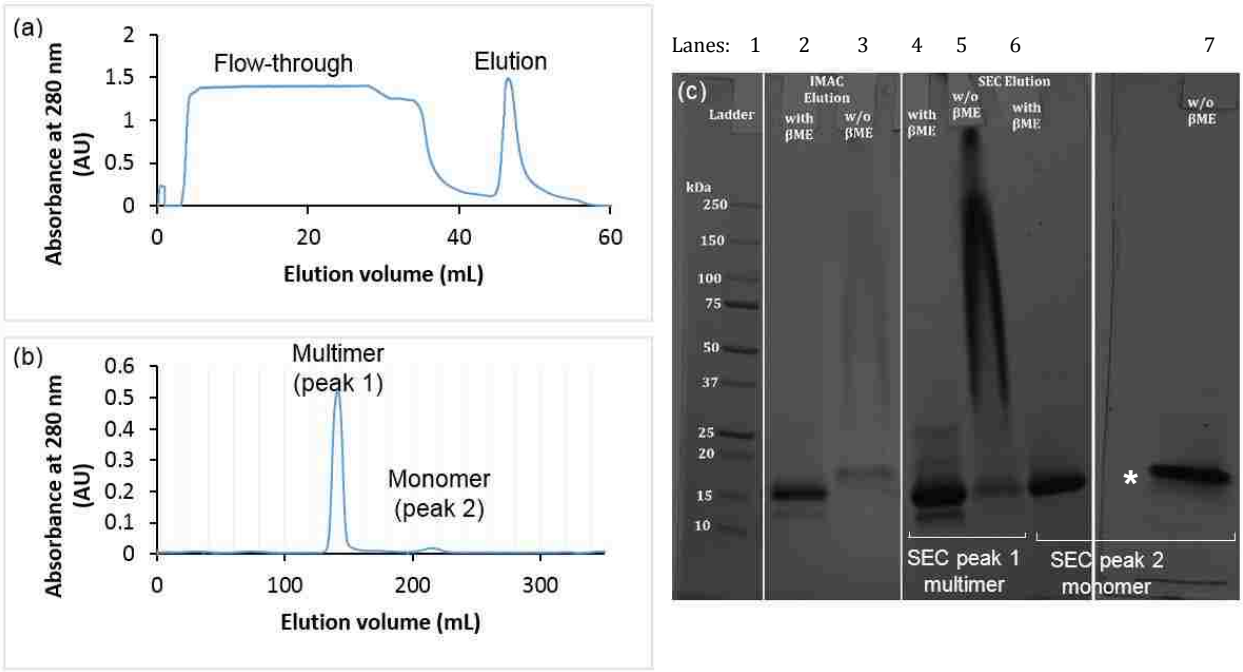

Figure 8. (a) Profinity IMAC nickel affinity chromatography purification of EGF26-28. (b) Size-exclusion chromatography (SEC) purification of the IMAC eluted EGF26-28. Two peaks elute at 140 mL (peak 1) and 220 mL (peak 2). (c) SDS-PAGE gel image of the IMAC and SEC purified EGF26-28 with and without βME: lane 1 (protein ladder), lane 2-3 (IMAC elution), lane 4-5 (SEC peak 1), and lane 6-7 (SEC peak 2).

confirmed by a single band (lane 7). Although we were able to produce the monomeric form of the protein, the quantity that was produced was too low for NMR studies.

EGF25-26-27:

EGF25-26-27-6XHis (EGF25-27) was expressed at 16°C overnight after induction with 0.5 mM IPTG in BL21-Gold(DE3) *E. coli* cells using the pET28b-EGF25-26-27 vector. *Figure 9* shows the SDS-PAGE gel image of the expression culture to monitor expression levels. The molecular

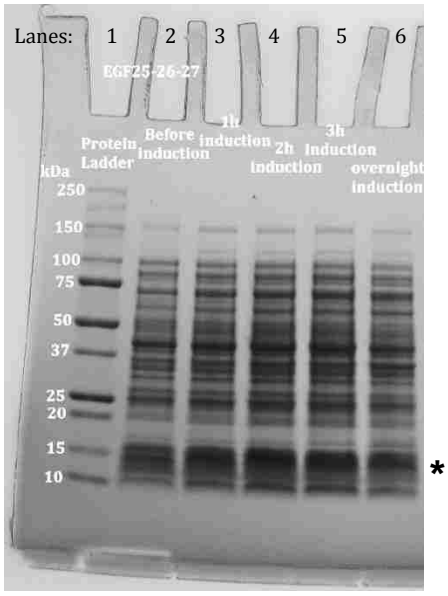

Figure 9. SDS-PAGE gel image of EGF25-27 expression culture. Samples include time point before induction (lane 2) and after induction (lane 3 = 1h, lane 4 = 2h, lane 5 = 3h, and lane 6 = overnight or approximately 16h).

weight of EGF25-27 is 13.8 kDa. As compared to the background *E. coli* proteins on the gel image, it appears that EGF25-27 did not express in overabundance. *Figure 10 (a)* and *(b)* shows the IMAC and gel filtration chromatograms, respectively. *Figure 10 (c)* and *(d)* shows the SDS-PAGE gel images of the chromatographic samples. In *Figure 10 (a)*, the elution profile indicates a high yield, but the SDS-PAGE in *Figure 10 (c)* shows a mixture of many different proteins indicating that EGF25-27 was produced in low yield, if at all. The gel filtration absorbance maximum in *Figure 10 (b)* confirms the low protein yield, and the multiple peaks confirm a mixture of proteins. Peak 4, which elutes where the monomeric EGF25-27 is expected, runs at a higher

MW on SDS-PAGE (*Figure 10d*), even with βME (lane 2). These results indicate that EGF25-27 does not express well in BL21-Gold(DE3) cells.

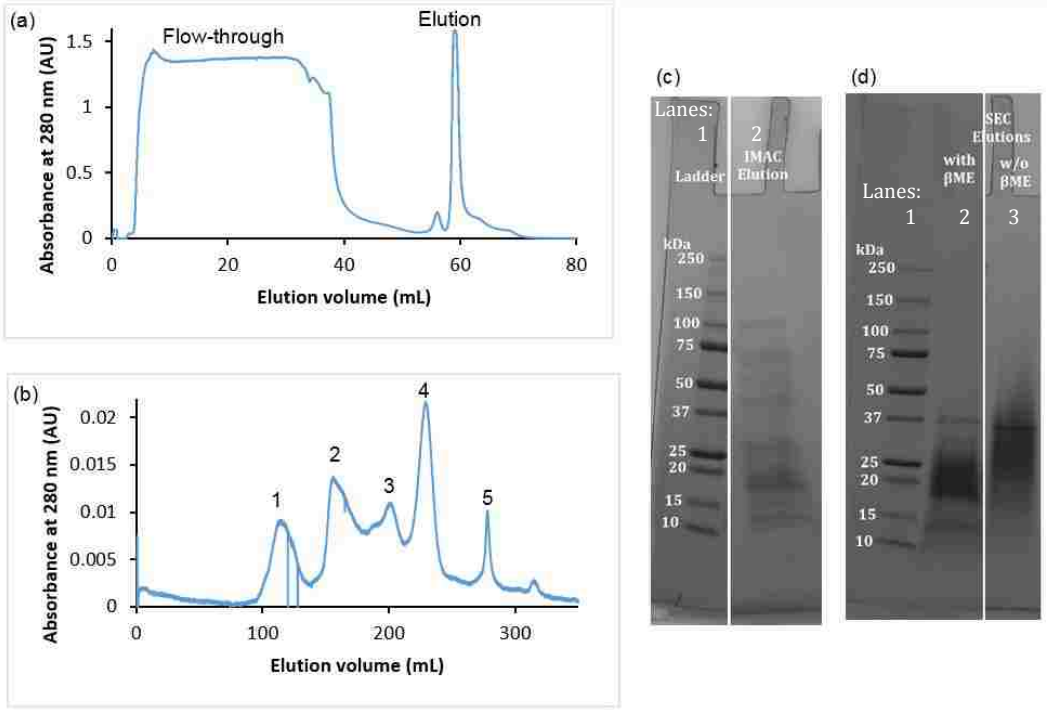

Figure 10. (a) Profinity IMAC nickel affinity chromatography purification of EGF25-27. (b) Size-exclusion chromatography (SEC) purification of the IMAC eluted EGF25-27. Multiple peaks elute between 100-300 mL. (c) SDS-PAGE gel image of the IMAC purified EGF25-27 with βME: lane 1 (protein ladder) and lane 2 (IMAC elution). (d) SDS-PAGE gel image of the SEC purified EGF25-27 with and without βME: lane 1 (protein ladder), lane 2-3 (SEC peak 4 at 230 mL).

EGF25-26-C21G:

EGF25-26-C21G-6XHis (EGF25-26) was expressed at 20°C overnight after induction with 0.5 mM IPTG in BL21-Gold(DE3) *E. coli* cells using the pET28a-EGF25-26-C21G vector. The molecular weight of EGF25-26-C21G is 8.99 kDa. *Figure 11* shows the SDS-PAGE gel image of the expression culture, which indicates that EGF25-26 did not express in abundance. *Figure 12 (a)* and *(b)* shows the IMAC and gel filtration chromatograms, respectively. The elution profile from the IMAC chromatogram suggests that the protein is in high yield. *Figure 12 (b)* shows the gel filtration elution profiles with two peaks. Peak 2, eluting at 225 mL, appears to be a monomer based on the elution profile of the standards. *Figure 12* (*c)* and *(d)* shows the SDS-PAGE gel

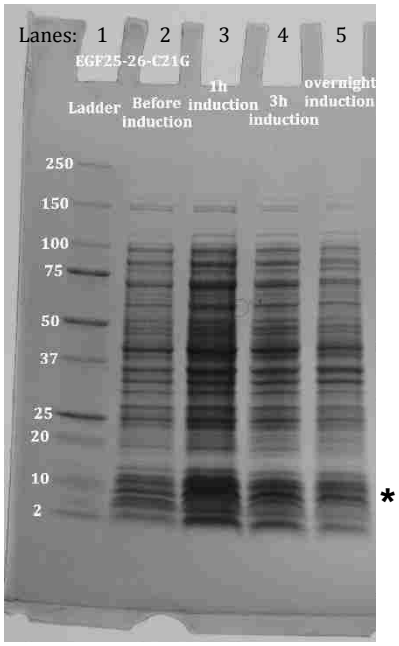

Figure 11. SDS-PAGE gel image of EGF25-26-C21G expression culture. Samples include time point before induction (lane 2) and after induction (lane 3 = 1h, lane 4 = 3h, and lane 5 = overnight or approximately 16h).

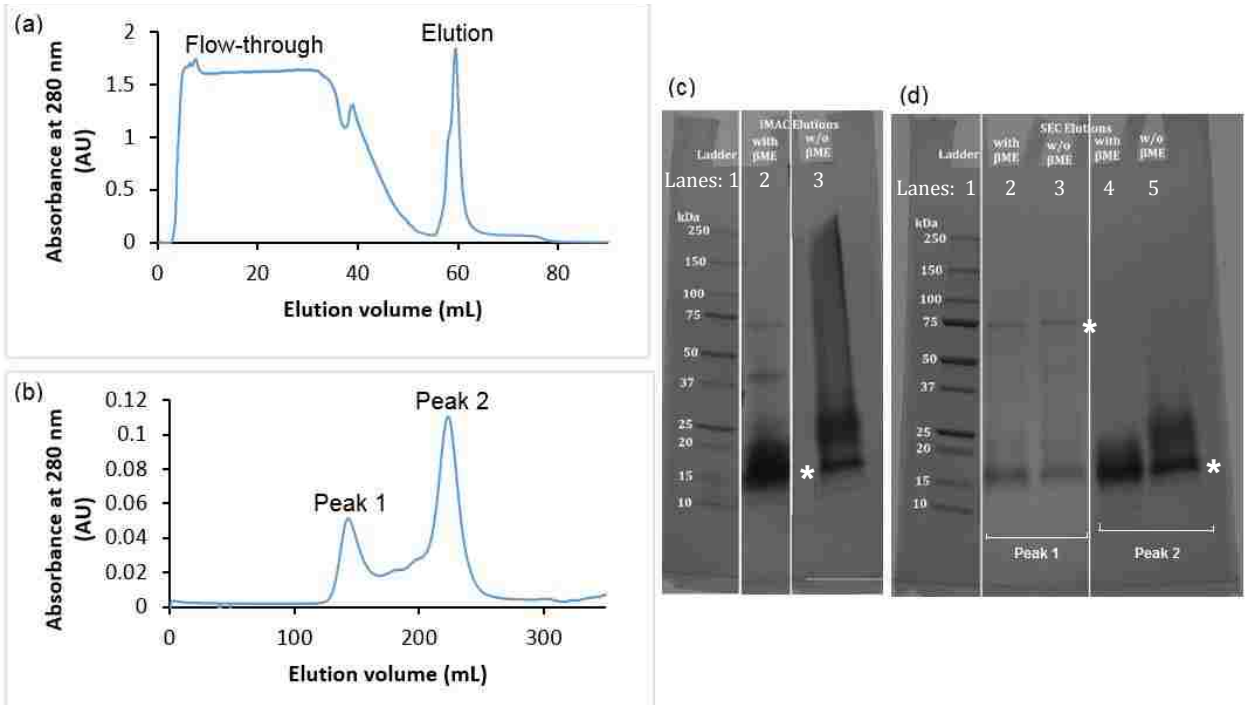

Figure 12. (a) Profinity IMAC nickel affinity chromatography purification of EGF25-26-C21G. (b) Size-exclusion chromatography (SEC) purification of the IMAC eluted EGF25-26-C21G. Two peaks elute at 145 mL (peak 1) and 225 mL (peak 2). (c) SDS-PAGE gel image of IMAC purified EGF25-26-C21G with and without βME: lane 1 (protein ladder) and lane 2-3 (IMAC elution). (d) SDS-PAGE gel image of SEC purified EGF25-26-C21G with and without βME: lane 1 (protein ladder), lane 2-3 (SEC peak 1), and lane 4-5 (SEC peak 2).

images of the chromatographic samples. The IMAC elution sample in *Figure 12 (c)* in the absence

of βME (lane 3) appears as a smear with a band at about 15 kDa (marked by an asterisk). After gel filtration purification, peak 1 samples (*Figure 12b*) on SDS-PAGE (*Figure 12d*) look like a mixture of cross-linked protein (MW ~75 kDa) and monomeric EGF25-26. The band observed near 15 kDa is not a dimer because the MW is below 18 kDa, so it is most likely monomeric EGF25-26. Therefore, peak 2 sample is most likely the monomeric EGF25-26.

EGF25-26-C21G-TEV-EGFP:

EGF25-26-C21G-TEV-EGFP-6XHis (EGF25-26-EGFP) was expressed at 16°C overnight after induction with 0.5 mM IPTG in BL21-Gold(DE3) and Shuffle T7 Express *E. coli* cells using the pLIC-EGF25-26-C21G-TEV-EGFP-His vector. *Figure 13 (a)* and *(b)* are the SDS-PAGE gel

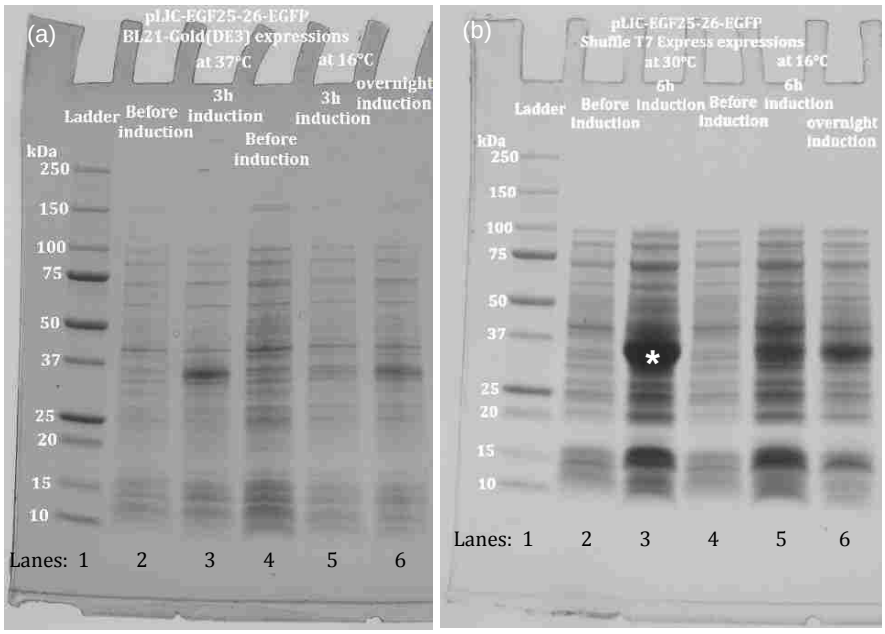

Figure 13. SDS-PAGE gel images of EGF25-26-EGFP expression cultures. (a) Expressions in BL21-Gold(DE3) cells. Samples include time points before induction (lane 2 and 4) and after induction (lane 3 = 3h at 37°C, lane 5 = 3h at 16°C, and lane 6 = overnight or approximately 16h at 16°C). (b) Expressions in Shuffle T7 Express cells. Samples include time points before induction (lane 2 and 4) and after induction (lane 3 = 6h at 30°C, lane 5 = 6h at 16°C, and lane 6 = overnight or approximately 16h at 16°C). The asterisk indicates overexpression of the protein.

images of the cultures in BL21-Gold(DE3) and Shuffle T7 Express cells, respectively, showing the expression levels in the cultures. The molecular weight (MW) of the EGF25-26-EGFP fusion

protein is 38.5 kDa. After cleavage by TEV protease, the MW of EGF25-26-C21G is 10.6 kDa and the EGFP-His is 27.9 kDa. As compared to the background *E. coli* proteins on the gel image in *Figure 13 (b)*, the fusion protein is seen to be overexpressed (marked by an asterisk). Shuffle cell expressions at 30°C for 6 hours after induction with IPTG had the highest expression level of EGF25-26-EGFP and was selected for further purification and analysis. Although the MW of the fusion protein is 38.5 kDa, the dark band of the expressed protein on the gel is seen a little below the 37 kDa marker. This difference in migration is not unusual because the way a protein migrates on a gel depends on its shape and net charge.

Figure 14 (a) shows the IMAC chromatogram of EGF25-26-EGFP purification from Shuffle T7 Express *E. coli* cells. *Figure 14 (b)* shows the SDS-PAGE gel image of the IMAC eluted samples in lane 2 and the reaction with TEV protease (AcTEV) in lane 3. 1 µL of AcTEV protease (10 units) was added to 30 µg of IMAC purified fusion protein and incubated for 16 hours at 30°C. There were two resulting bands observed after TEV cleavage, a darker band between 25-30 kDa

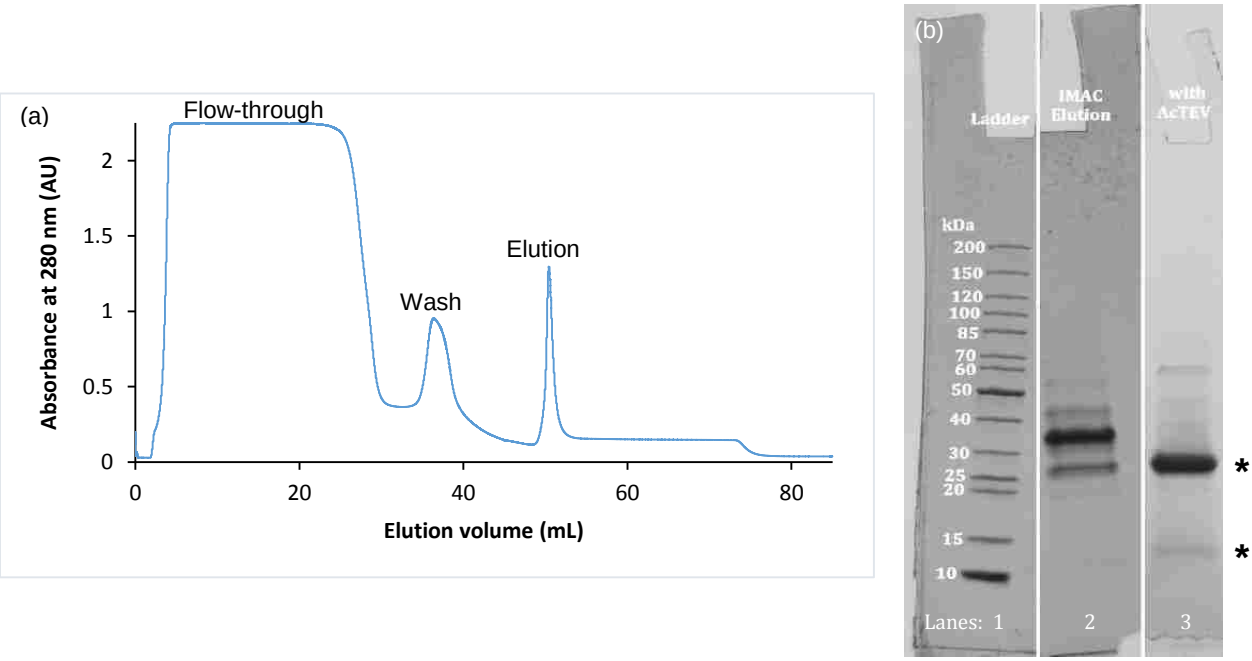

Figure 14. (a) Profinity IMAC nickel affinity chromatography purification of EGF25-26-EGFP fusion protein in Shuffle T7 Express cells. (b) SDS-PAGE gel image of the IMAC purification: lane 1 (protein ladder), lane 2 (IMAC elution of the fusion protein sample), and lane 3 (IMAC elution sample reaction with AcTEV protease).

and a faint band between 10-15 kDa (marked by asterisks) with respect to the protein ladder on the gel. The band between 25-30 kDa is most likely EGFP-His after TEV cleavage from the fusion protein, and the band between 10-15 kDa is probably our target protein EGF25-26-C21G. These observations confirm that AcTEV is cleaving the fusion protein at the TEV cleavage site, however the target protein (EGF25-26-C21G) is not being produced as much. The protein did not noticeably precipitate during the purification steps. Thus, the reason for not being able to get enough yield of the target protein after TEV cleavage, despite having success in producing the fusion protein in high yield, is not known.

EGF27-TEV-EGFP:

EGF27-TEV-EGFP-6XHis (EGF27-EGFP) was expressed at 16°C overnight after induction with 0.5 mM IPTG in BL21-Gold(DE3) and Shuffle T7 Express *E. coli* cells using the pLIC-EGF27-TEV-EGFP-His vector. *Figure 15 (a)* and *(b)* are the SDS-PAGE gel images of the cultures in BL21-Gold(DE3) and Shuffle T7 Express cells, respectively, showing the expression levels in the cultures. The MW of the EGF27-EGFP fusion protein is 35.5 kDa. After TEV cleavage, the MW of EGF27 is 7.5 kDa and the EGFP-His is 27.9 kDa. As compared to the background *E. coli* proteins on the gel image in *Figure 15*, the fusion protein is seen to be overexpressed (marked by an asterisk). BL21-Gold(DE3) cell expressions overnight at 16°C after induction with IPTG had the highest expression of pLIC-EGF27-TEV-EGFP-6XHis protein and was selected for further purification and analysis.

Figure 16 (a) shows the IMAC chromatogram of EGF27-EGFP purification in BL21-Gold(DE3) *E. coli* cells. *Figure 16 (b)* shows the SDS-PAGE gel image of the IMAC elutions with two dark bands (marked by an asterisk), one above and one below the 30 kDa marker. The band above 30 kDa is most likely the fusion protein, and the one below is likely EGFP-His. The

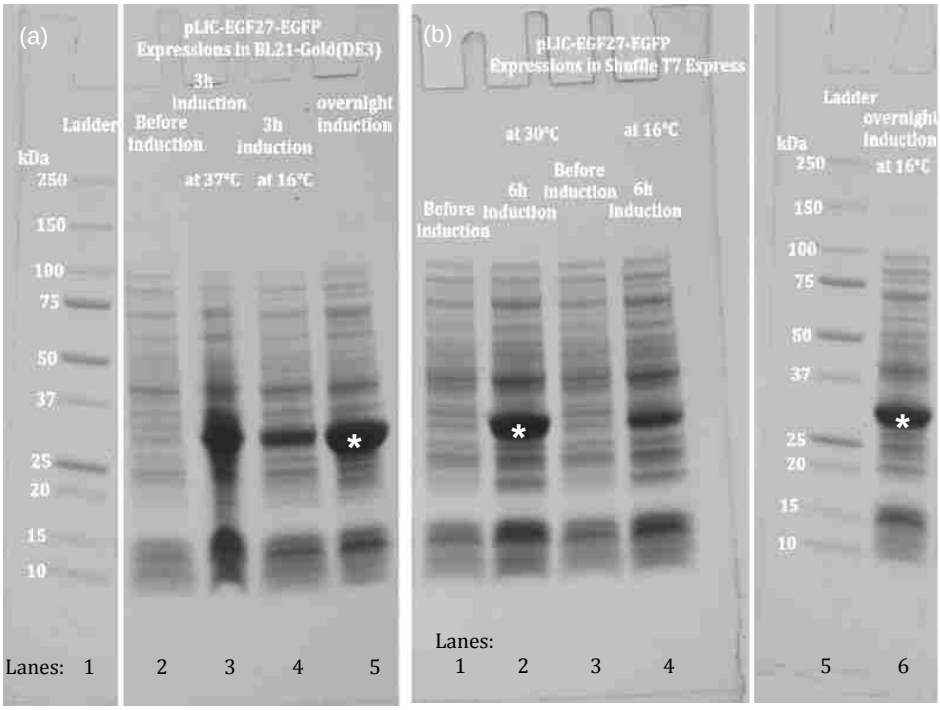

Figure 15. SDS-PAGE gel images of EGF27-EGFP expression cultures. (a) Expressions in BL21-Gold(DE3) cells. Samples include time points before induction (lane 2) and after induction (lane 3 = 3h at 37°C, lane 4 = 3h at 16°C, and lane 5 = overnight or approximately 16h at 16°C). (b) Expressions in Shuffle T7 Express cells. Samples include time points before induction (lane 1 and 3) and after induction (lane 2 = 6h at 30°C, lane 4 = 6h at 16°C, and lane 6 = overnight or approximately 16h at 16°C). The asterisk indicates overexpression of the protein.

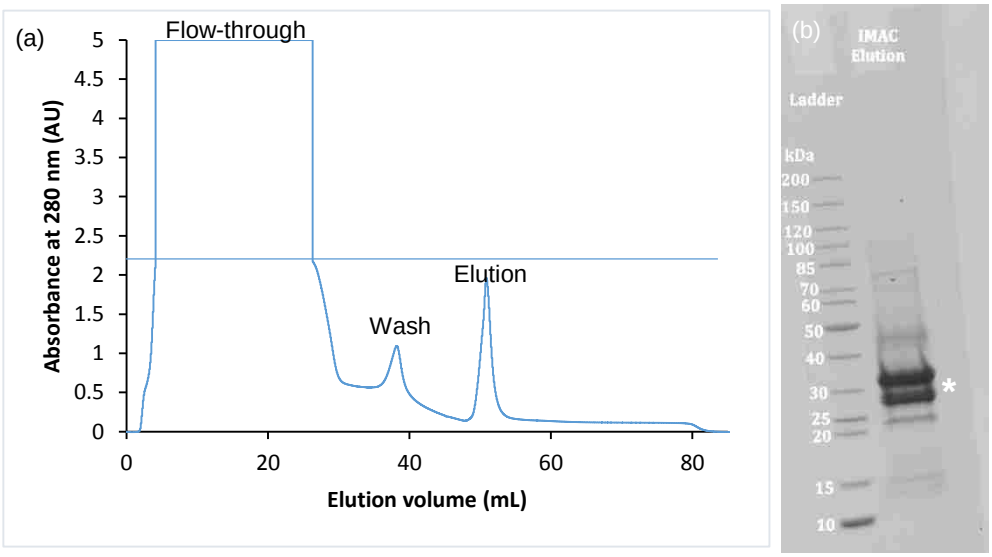

Figure 16. (a) Profinity IMAC nickel affinity chromatography purification of EGF27-EGFP in BL21-Gold(DE3) *E. coli* cells. The absorbance signal was cut-off between 2 to 2.5 AU, so the signal observed up to 5 AU was possibly due to a problem with the new UV lamp. (b) SDS-PAGE gel image of the IMAC purified EGF27-EGFP.

observance of two bands without the addition of TEV protease is quite unusual and indicates that

the fusion protein is being cleaved by an *E. coli* protease.

Figure 17 shows the SDS-PAGE gel image of EGF27-EGFP fusion protein reaction with TEV protease (AcTEV). 10 µL of AcTEV protease (100 units) was added to 300 µg of the fusion protein and incubated at 30°C for 16 hours. Ni Sepharose 6 Fast Flow affinity chromatography purification by gravity flow was performed, and the flow-through, washes, and elution samples

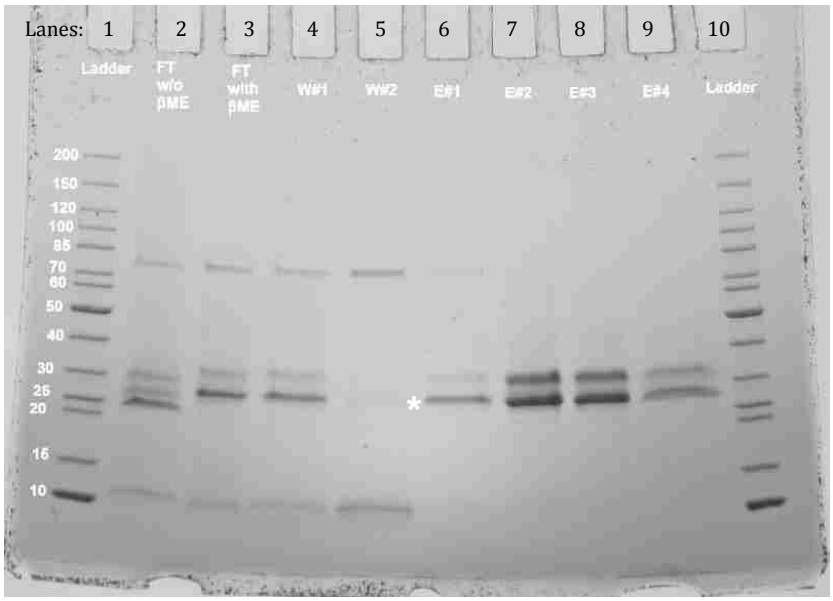

Figure 17. SDS-PAGE gel image of the nickel affinity chromatography gravity purification of pLIC-EGF27-TEV-EGFP-His after TEV cleavage using AcTEV protease. Lane 1 and 10 (protein ladder), lane 2-3 (flow-through with and without βME), lane 4-5 (washes), and lane 6-9 (elutions).

were analyzed by SDS-PAGE. TEV protease cleaves the fusion protein into EGF27 and EGFP-His proteins. EGFP-His has a 6X histidine tag, which has an affinity for nickel and thus binds to the nickel media. EGF27, which is our target protein, does not have any affinity tags attached to it after TEV cleavage, so it does not bind to the nickel media. EGFP-His is eluted from the nickel media by passing buffer containing (~0.5 M) imidazole. The gel image in *Figure 17* indicates that the TEV cleavage reaction worked, which is shown by the appearance of a dark band between 25 and 30 kDa in the elution samples (lanes 6-9, marked by an asterisk). The band at 30 kDa in the elution samples is uncleaved protein, indicating that the TEV protease does not cleave this fusion protein very efficiently. The appearance of a faint band around 10 kDa in the flow-through

sample (FT) is the target protein (EGF27). The amount of EGF27 produced by this method is very low and does not produce enough protein for NMR studies.

pelB-EGF27-TEV-EGFP:

pelB-EGF27-TEV-EGFP-6XHis (pelB-EGF27-EGFP) was expressed at 16°C overnight after induction with 0.5 mM IPTG in BL21-Gold(DE3) and Shuffle T7 Express competent *E. coli* cells using the pLIC-pelB-EGF27-TEV-EGFP-His vector. *Figure 18 (a)* and *(b)* are the SDS-PAGE gel images of the cultures in BL21-Gold(DE3) and Shuffle T7 Express cells, respectively, showing the expression level of the cultures. The MW of the pelB-EGF27-EGFP fusion protein is 38.7 kDa before the periplasmic proteolytic digestion of the pelB leader sequence. After pelB digestion, the MW of the fusion protein is 35.5 kDa. After TEV cleavage, the MW of EGF27 without pelB is 7.5

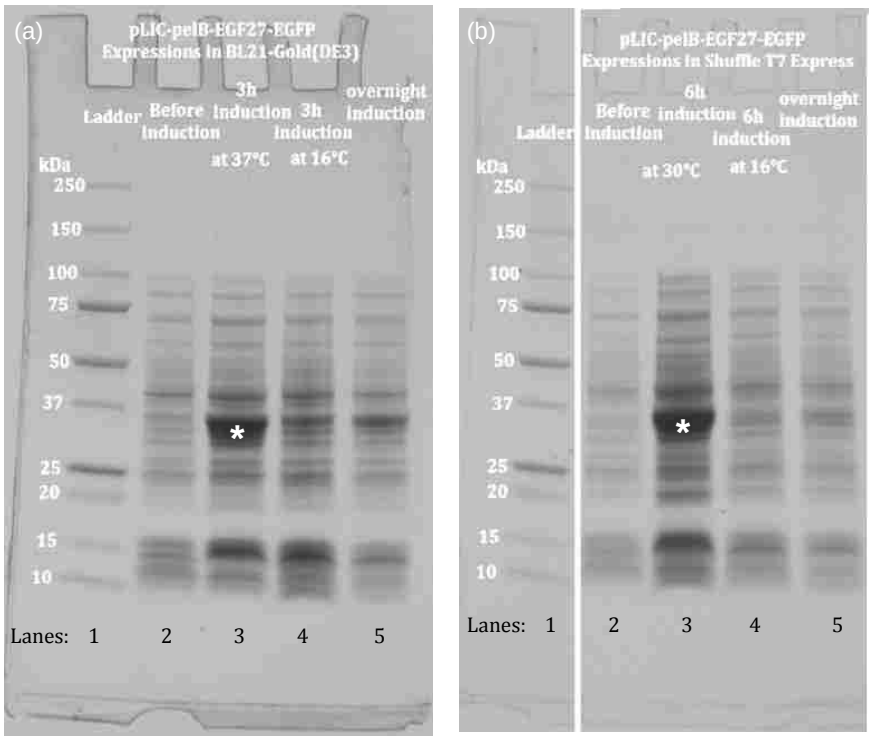

Figure 18. SDS-PAGE gel images of pelB-EGF27-EGFP expression cultures. (a) Expressions in BL21-Gold(DE3) cells. Samples include time points before induction (lane 2) and after induction (lane 3 = 3h at 37°C, lane 4 = 3h at 16°C, and lane 5 = overnight or approximately 16h at 16°C). (b) Expressions in Shuffle T7 Express cells. Samples include time points before induction (lane 2) and after induction (lane 3 = 6h at 30°C, lane 4 = 6h at 16°C, and lane 5 = overnight or approximately 16h at 16°C). The asterisk indicates overexpression of the protein.

kDa, EGF27 with pelB is 10.7 kDa, and the EGFP-His is 27.9 kDa. As compared to the

background *E. coli* proteins on the gel image in *Figure 18* above, the fusion protein is seen to be overexpressed (marked by an asterisk). BL21-Gold(DE3) cell expressions at 37°C for 3 hours and Shuffle cell expressions at 30°C for 6 hours after induction with IPTG had the highest expression of pelB-EGF27-TEV-EGFP-6XHis protein. The dark band of the expressed protein on the gel is seen a little below the 37 kDa marker. The gel migration of EGF27-EGFP in *Figure 15* should be the same as in *Figure 18*, if the pelB leader sequence is cleaved *in vivo*. The migration is clearly different, ~35 kDa in *Figure 18* and ~31 kDa in *Figure 15*, indicating that pelB-EGF27-EGFP is cleaved in the periplasm *in vivo*. Because the pelB sequence was not cleaved, the protein was not purified.

EGF27 sub-cloning and expression:

EGF27 was sub-cloned from the pET20b-pelB-EGF27-6XHis plasmid into the pET-28a plasmid at restriction sites NcoI and XhoI to make pET28a-EGF27-6XHis. Because a one-pot cloning reaction was used, the possible ligation products include the desired pET28a-EGF27 and the undesired empty pET-28a, pET20b-EGF27, and pET20b-pET28a hybrid. The post-ligation reaction with NdeI linearizes the undesired products, reducing the number of ligation-competent plasmid. Growing the transformed cells on agar plates with kanamycin selects for colonies transformed with either pET28a-EGF27-6XHis or empty pET-28a. pET 3' and pET 5' primers were used to amplify the plasmid from individual colonies by polymerase chain reaction (PCR). The amplification product from pET28a-EGF27-6XHis is 357 base pairs (bp) and from empty pET-28a is 313 bp. DNA agarose gel electrophoresis analysis of the PCR products (*Figure 19*) show that colonies C#1, C#2, C#3, C#4, and C#8 produce bands between 300-400 bp (marked by an asterisk) near the desired product size of 357 bp. Plasmid was purified from C#8, sequenced by the GeneLab at the LSU School of Veterinary Medicine, and verified to be pET28a-EGF27-6XHis.

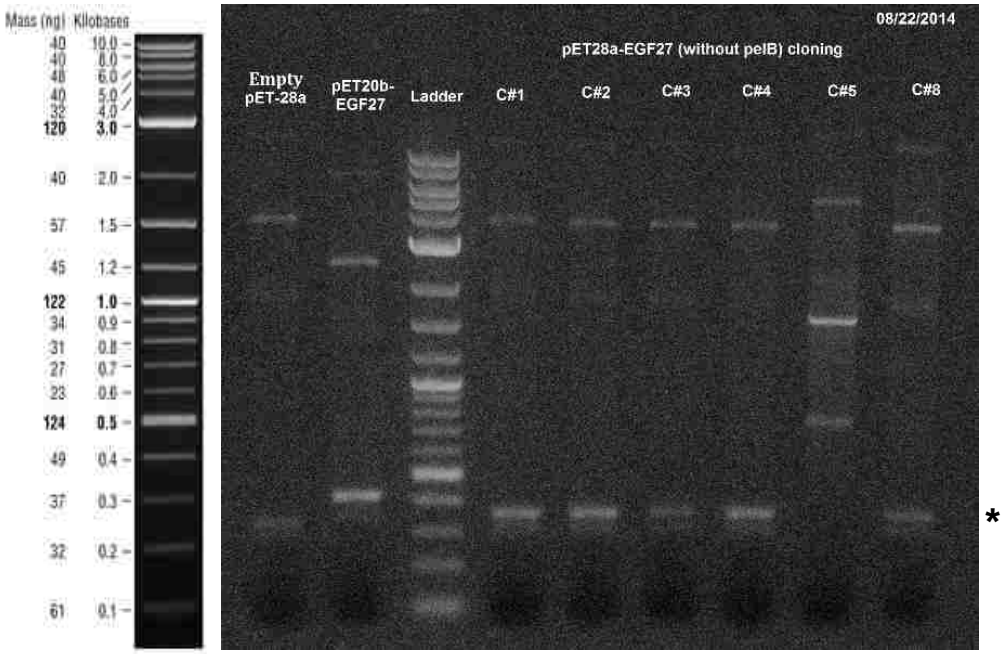

Figure 19. DNA agarose gel electrophoresis showing colony PCR of the sub-cloned pET28a-EGF27 plasmid. Lane 1 (empty pET-28a vector plasmid, control), lane 2 (pET20b-pelB-EGF27, control), lane 3 (2-log DNA ladder), lane 4-9 (individual colonies selected from the transformation plate of the sub-cloned pET28a-EGF27 plasmid).

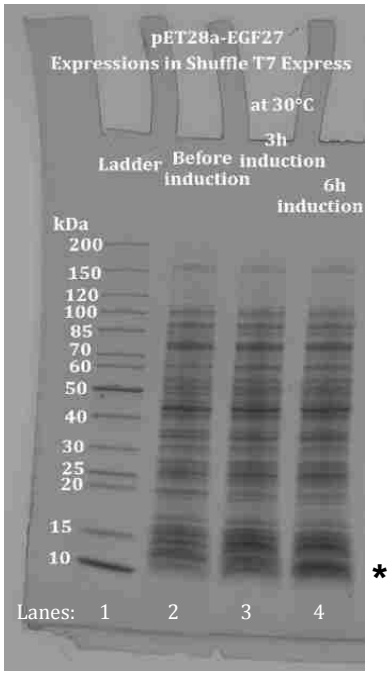

Figure 20. SDS-PAGE gel image of EGF27 expression culture in LB media; expressions in Shuffle T7 Express cells. Samples include time points before induction (lane 2) and after induction (lane 3 = 3h at 30°C, and lane 4 = 6h at 30°C).

pET28a-EGF27-6XHis (EGF27) was expressed in LB and minimal media at either 30°C for 4-6h or 16°C overnight after induction with 0.4 mM IPTG in BL21-Gold(DE3), Shuffle T7 Express, and Shuffle T7 Express *lysY E. coli* cells using the pET28a-EGF27 vector. *Figure 20* shows the SDS-PAGE gel image of the culture in Shuffle T7 Express cells (expressions in LB media), showing the expression level of the culture. The molecular weight of EGF27 is 7.1 kDa. Looking at the results from the gel, the slightly darker band after 6h of induction (marked by an asterisk) maybe over-expressed EGF27-6XHis.

Figure 21 (a), *(b)* and *(c)* shows the SDS-PAGE gel images of the cultures in BL21-Gold(DE3), Shuffle T7 Express, and Shuffle T7 Express *lysY* cells, respectively (expressions in ^{15}N labeled minimal media). The slightly darker bands after 4h of induction (marked by asterisks) maybe EGF27-6XHis.

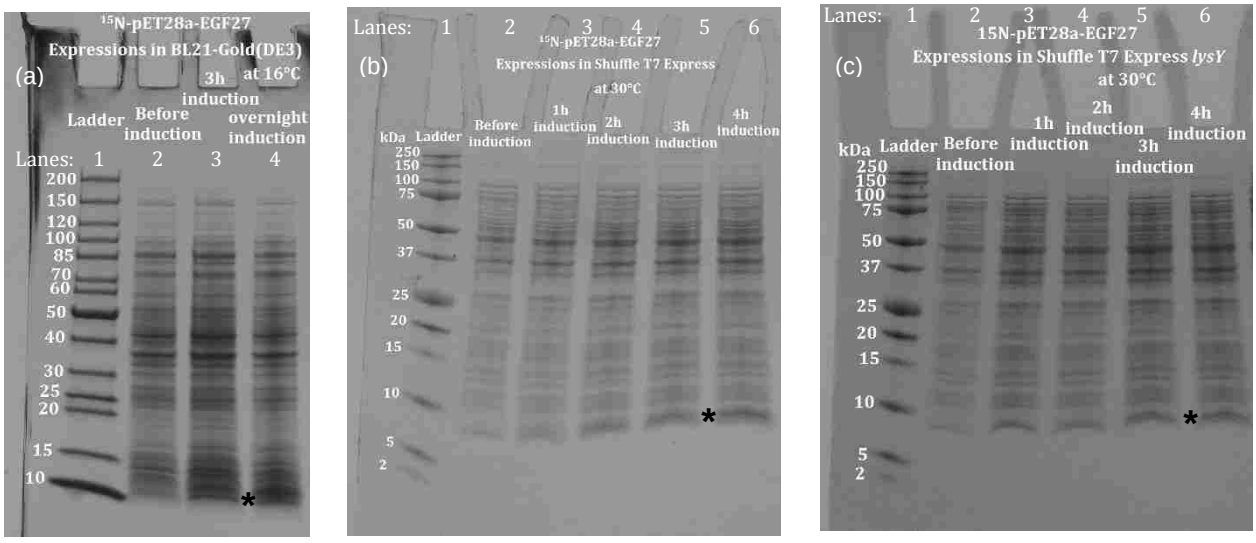

Figure 21. SDS-PAGE gel images of EGF27 expression cultures in ^{15}N labeled minimal media. (a) Expressions in BL21-Gold(DE3) cells. Samples include time points before induction (lane 2) and after induction (lane 3 = 3h at 16°C, and lane 4 = overnight or approximately 16h at 16°C). (b) Expressions in Shuffle T7 Express cells. Samples include time points before induction (lane 2) and after induction (lane 3 = 1h, lane 4 = 2h, lane 5 = 3h, and lane 6 = 4h at 30°C). (c) Expressions in Shuffle T7 Express *lysY* cells. Samples include time points before induction (lane 2) and after induction (lane 3 = 1h, lane 4 = 2h, lane 5 = 3h, and lane 6 = 4h at 30°C).

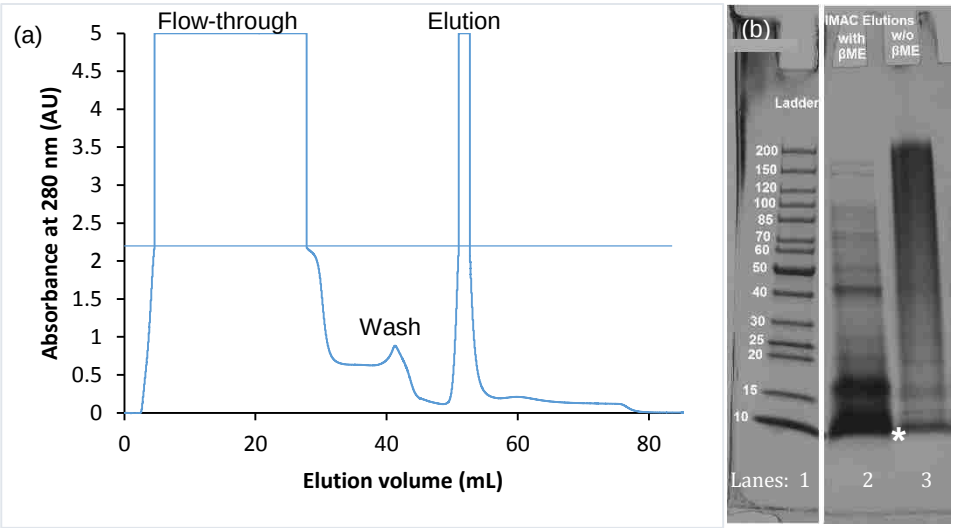

Figure 22. (a) Profinity IMAC nickel affinity chromatography purification of EGF27 in BL21-Gold(DE3) cells expressed in ^{15}N labeled minimal media. The absorbance signal was cut-off between 2 to 2.5 AU, so the signal observed up to 5 AU was possibly due to a problem with the new UV lamp. (b) SDS-PAGE gel image of IMAC purified EGF27 with βME (lane 2) and without βME (lane 3).

Figure 22 (a) shows the IMAC chromatography purification of ^{15}N labeled EGF27 in BL21-Gold(DE3) purification. *Figure 22 (b)* shows the SDS-PAGE gel image of the elution sample with and without βME, lanes 2 and 3, respectively. The IMAC elution sample, in the absence of βME, appears as a smear throughout the lane with a faint band a little above the 10 kDa marker. This smear suggests that most of the protein is in multimeric form and highly cross-linked with intermolecular disulfide bonds.

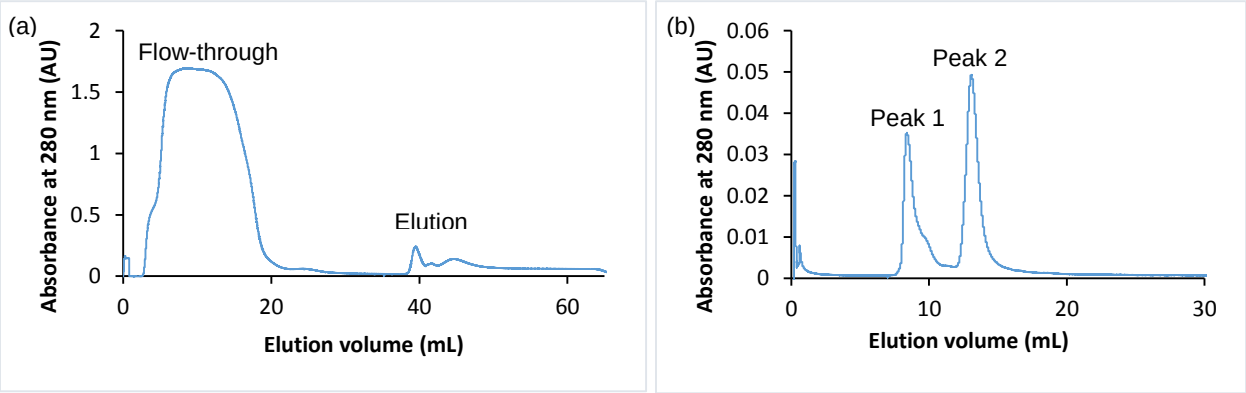

Figure 23. (a) Profinity IMAC nickel affinity chromatography purification of EGF27 in Shuffle T7 Express cells expressed in LB media. (b) Bio-Rad ENrich SEC 70 gel filtration chromatography purification of the IMAC purified EGF27 sample.

Figure 23 (a) shows the IMAC chromatography purification of EGF27-6XHis from a culture of Shuffle T7 Express in LB media. The chromatogram shows that the eluted protein (40 mL) is in low abundance. The protein concentration was determined using the Coomassie (Bradford) protein assay, and the yield was found to be 1.45 mg. The IMAC elutions were concentrated to about 2 mL using an Amicon 3 kDa molecular weight cut-off (MWCO) centrifugal filter unit and desalted using a Bio-Rad P-4 column. The protein yield after desalting (20 mM MES, 100 mM NaCl, pH 6.5 buffer) was determined again and was found to be 240 µg. The protein was likely lost through precipitation while concentrating the samples in the centrifugal filter. *Figure 23 (b)* shows the gel filtration chromatography analysis of the IMAC purified EGF27 sample using a Bio-Rad ENrich SEC 70 gel filtration column. For this analysis, 250 µL containing 120 µg of IMAC eluted protein was injected onto the gel filtration column. Two peaks eluted and each of these were analyzed by SDS-PAGE.

Figure 24 (a) shows the SDS-PAGE gel image of the sample eluted from the IMAC column (*Figure 23 (a)*) with and without βME. *Figure 24 (b)* shows the SDS-PAGE of the samples from the gel filtration column (*Figure 23 (b)*). Lanes 2-3 in *Figure 24 (b)* are the peak 1 samples of

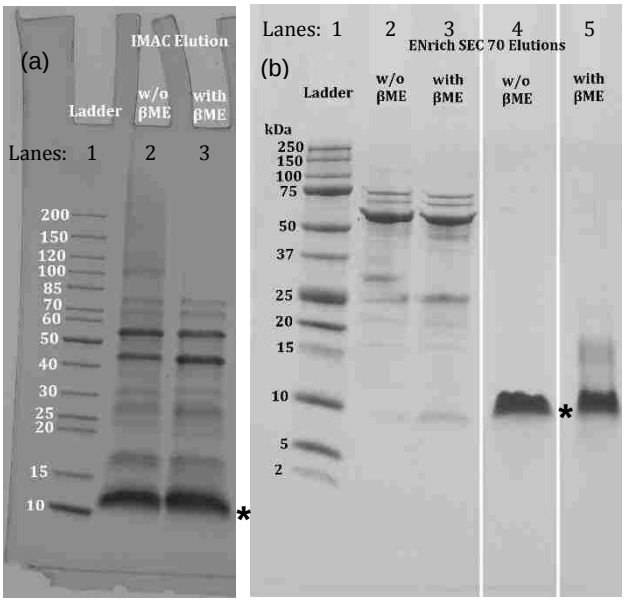

Figure 24. SDS-PAGE gel images of the EGF27 purification with and without βME using (a) Profinity IMAC and (b) ENrich SEC 70.

Figure 23 (b) with and without βME. Peak 1 is mostly a higher molecular weight protein (~60 kDa). Lanes 4-5 are the peak 2 samples with and without βME. The dark band of the protein appears a little below the 10 kDa marker both in the absence and presence of βME (*Figure 24 (b)*). The gel filtration and SDS-PAGE results taken together suggest that peak 2 in the gel filtration chromatogram is EGF27 in its monomeric form and is not cross-linked. The molecular weight of the sample was further confirmed by MALDI MS (discussed in section 2.3.3). Additionally, the folding isoforms of this protein was further analyzed using RP-HPLC (discussed in section 2.3.2).

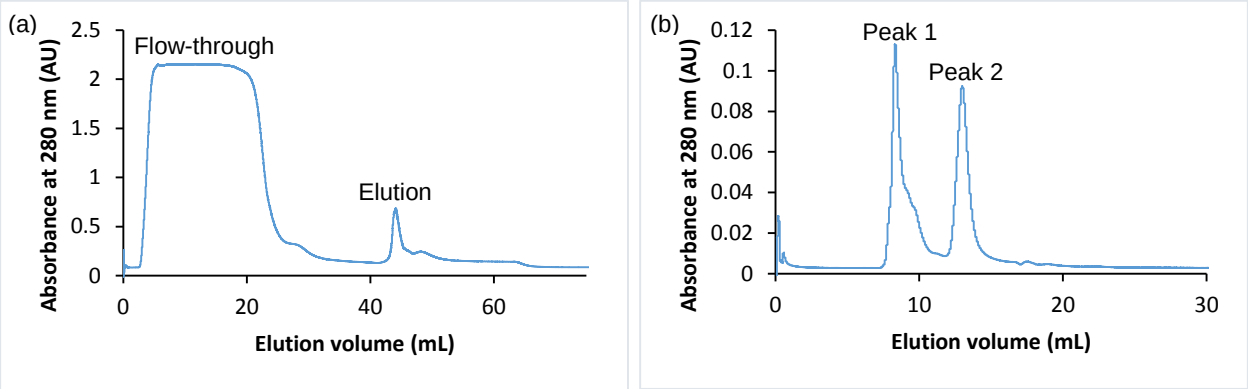

Figure 25. (a) Profinity IMAC nickel affinity chromatography purification of EGF27 in Shuffle T7 Express *lysY* cells expressed in LB media. (b) Bio-Rad ENrich SEC 70 gel filtration chromatography purification of the IMAC purified EGF27 sample.

Figure 25 (a) and *(b)* shows the IMAC and gel filtration chromatograms from purification of EGF27-6XHis Shuffle T7 Express *lysY* cells cultured in LB media. Peak 2 on the chromatogram in *Figure 25 (b)* was confirmed to be monomeric EGF27-6XHis by SDS-PAGE analysis and MALDI MS. The monomeric EGF27-6XHis (peak 2 elution) concentration was determined using Coomassie protein assay and the yield was found to be 100 µg, which was similar to the yield from Shuffle T7 Express cells. The reason for such a low yield in the end was because the protein was lost while concentrating using the molecular weight cut-off centrifugal filters.

2.3.2 Reversed-phase HPLC

The EGF27-6XHis samples purified with gel filtration were analyzed by reversed phase high performance liquid chromatography (RP-HPLC) using a Purity Semi-Prep HPLC C18 column. A linear gradient from 0 to 75% solvent B (100% acetonitrile containing 0.1% TFA) in solvent A (water with 0.1% TFA) for 45 minutes at a 2 mL/min flow rate was applied to elute the folding isoforms of the protein. The protein sample was treated with 15 mM dithiothreitol (DTT) or 15 mM calcium chloride. All samples were adjusted to 0.1% TFA before analysis. DTT reduces disulfide bonds formed between the cysteine residues and unfolds the protein. EGF27-6XHis has a predicted calcium binding site, so adding calcium may alter the retention time of EGF27-6XHis to aid in identifying the correct peak. RP-HPLC elution of EGF27-6XHis was recorded at 214 nm. *Figure 26 (a)* shows that in the absence of DTT, the sample elutes at 26.1 min and in presence of DTT, the peak migrates and elutes at 28.6 min. The peak at 26.1 minute most likely has all

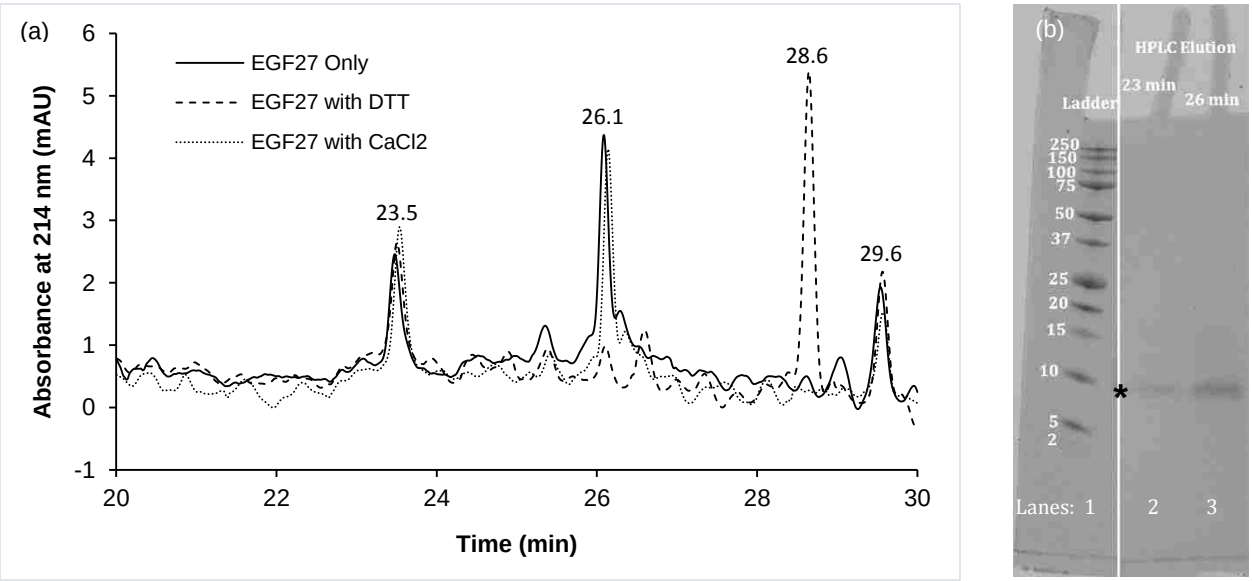

Figure 26. (a) Reversed-phase HPLC elution profile in presence and absence of DTT or $CaCl_2$. A peak at 26.1 min disappears and appears at 28.6 min when DTT was added to EGF27-6XHis sample. (b) SDS-PAGE gel image of the RP-HPLC samples eluted at 23 min (lane 2) and 26 min (lane 3).

three disulfide bonds correctly formed, and the addition of DTT reduces them, so the peak shifts to 28.6 min. Therefore, the sharp peak eluting at 26.1 min is most likely the properly folded

EGF27-6XHis, and the peak eluting next to it around 28.6 min is most likely its reduced form. These results suggest that the shift in retention time with DTT identifies disulfide-bonded EGF27-6XHis. The peaks did not shift noticeably in the absence or presence of calcium chloride, so this shows that the addition of calcium did not affect the retention time of EGF27-6XHis significantly. Peaks eluting at 23.5 min and 29.6 min remain unchanged in absence or presence of DTT or $CaCl_2$. Fractions of eluents from individual peaks were collected separately based on the 214 nm absorbance profile and freeze dried on a lyophilizer. The samples eluting at 23 min and 26 min were analyzed by SDS-PAGE, *Figure 26 (b)*, and matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF/TOF MS) to confirm the molecular weight of EGF27-6XHis (discussed in section 2.3.3).

2.3.3 Mass Spectrometry

The samples obtained from RP-HPLC were spotted on the MALDI target plate (Bruker Daltonics MTP 384 massive target T) and analyzed by Bruker UltrafleXtreme MALDI-TOF/TOF MS in the LSU Mass Spectrometry Facility. The molecular weight of EGF27-6XHis is 7109.76 daltons as determined from the protein sequence using Swiss Institute of Bioinformatics – ExPASY Bioinformatics Resource Portal. *Figure 27* shows the MALDI-MS spectra with mass to charge (*m/z*) ratio 7109, confirming the molecular weight of the purified protein to be correct. Hence, from these observations and analyses, the protein purified is EGF27-6XHis. MALDI-MS analysis of the RP-HPLC elutions at 23 min and 26 min also confirms both to be EGF27-6XHis peaks. The peak eluting at 23 min does not shift after the addition of DTT though, unlike the peak eluting at 26 min. It is not clear what isoform of EGF27-6XHis elutes at 23 min. In the next step, we isotopically labeled EGF27-6XHis using ^{15}N-NH_4Cl in the expression media and analyzed the "*foldedness*" of the protein with ^{1}H-^{15}N HSQC NMR.

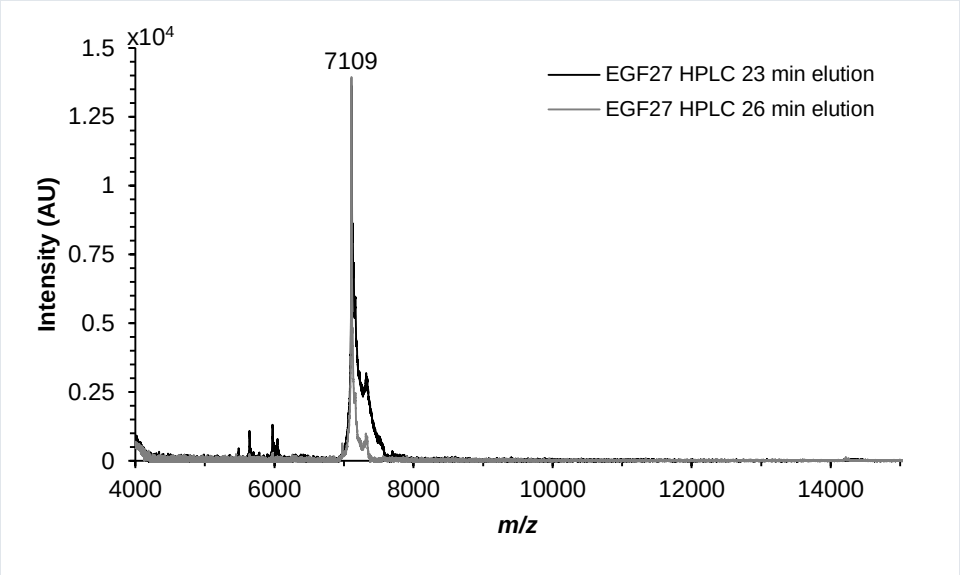

Figure 27. MALDI mass spectrometry spectra showing mass to charge (*m/z*) ratio to be 7109 of the EGF27-6XHis sample.

2.3.4 ^{1}H-^{15}N HSQC NMR

The EGF27-6XHis was expressed in minimal media with isotopic labeling using ^{15}N-NH_4Cl as previously described (in section 2.2.3). The protein was purified using nickel affinity chromatography followed by gel filtration chromatography as previously discussed. The peak eluting later on the gel filtration chromatogram was deduced to be the properly folded EGF27 from previous analyses and were pooled. The molecular weight was confirmed by running a sample on an SDS-PAGE gel. The ^{15}N-EGF27 protein sample was concentrated using a 3 kDa MWCO centrifugal filter. The NMR sample was prepared in a 250 µL total volume containing 10% D_2O, 90% H_2O, and 0.5 µL of 1% (w/v) DSS in D_2O for chemical shift referencing. The concentration of the NMR sample was determined to be 0.5 mM. A D_2O matched 5 mm Shigemi tube was used to hold the sample. A gradient sensitivity enhanced HSQC (gNhsqc) experiment was run with 16 scans and 256 number of increments at 27°C temperature.

Each ^{1}H atom attached to the ^{15}N atom on the backbone or the side chain of the protein corresponds to each peak in the ^{1}H-^{15}N HSQC spectrum. *Figure 28* shows the HSQC spectrum of 0.5 mM ^{15}N-EGF27-6XHis in 10% D_2O. The peaks in the spectra are not well dispersed indicating that the protein is cross-linked or is not correctly folded.

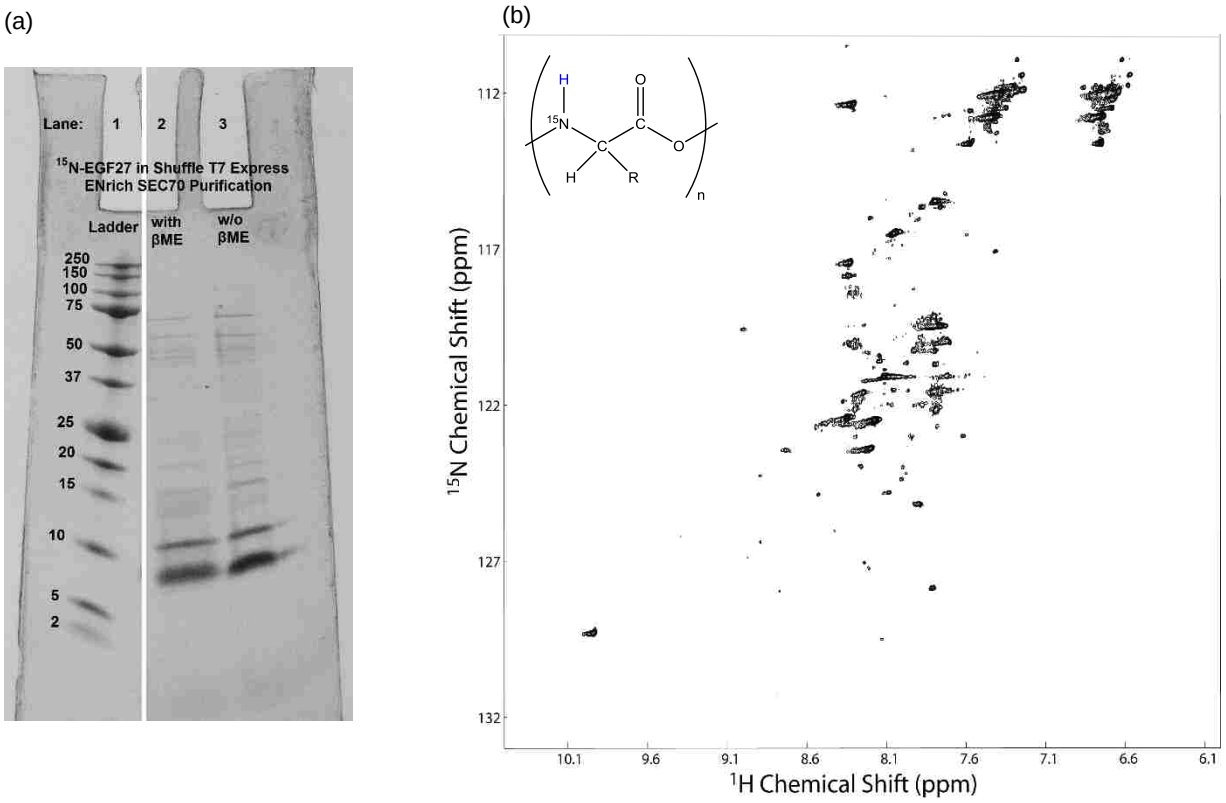

Figure 28. (a) SDS-PAGE gel image of the ^{15}N-EGF27 purification with βME (lane 2) and without βME (lane 3) using Enrich SEC 70. (b) ^{1}H-^{15}N HSQC spectrum of ^{15}N labeled EGF27 showing mostly undispersed peaks.

2.4 Discussion and Conclusions

The goal of this project was to express and purify milligram quantities of properly folded monomeric EGF repeats and determine the three dimensional structure of the protein using NMR spectroscopy. In order to obtain a good spectrum for protein structural studies using NMR spectroscopy, a protein concentration of at least 0.5 mM is required. Proteins with molecular weights ranging from 7 to 9 kDa require about a milligram of properly folded protein in 250 µL total sample volume to give a concentration of 0.5 mM.

We started the expressions in LB media using three tandem EGF repeats in a single construct for EGF26-27-28 and EGF25-26-27. After purification, the final yield of the monomeric EGF26-27-28 protein was found to be very low (in microgram quantities) and not enough to prepare samples for NMR. The tri-repeat sequence was aggregating possibly due to inter-molecular disulfide bonds. EGF25-26-27, in particular, is susceptible to intermolecular disulfide bond formation due to the presence of an extra cysteine residue in the EGF25 sequence. In order to reduce the chances of inter-repeat cross-linking and increase the chance of proper disulfide-mediated folding, we cloned two EGF domains together instead of three domains.

EGF25-26-C21G was cloned by mutating the extra cysteine residue in EGF25 to a glycine residue. This protein was expressed in BL21-Gold(DE3) cells and upon purification, we observed that the protein was in multimeric form. Thus, this sample was not suitable for NMR analysis. In order to overcome the problem of protein cross-linking during expression and purification steps, we used an enhanced green fluorescent protein as a fusion with a TEV cleavage site in between them.

The use of the green fluorescent protein as a fusion partner is one technique for promoting proper protein folding. It also acts as a reporter of proper folding.[41] EGF25-26-C21G, pelB-EGF27, and EGF27 without pelB were cloned separately with the green fluorescent protein as a fusion. The amount of target protein produced by this method after the TEV cleavage reaction was relatively low and was not enough to produce the isotopically labeled protein in milligram quantities required for our NMR studies.

During the green fluorescent protein fusion trial expressions, we noticed that the EGF27 construct without the pelB leader sequence was expressing significantly higher compared to the one with the pelB sequence. In the next step, we cloned a single EGF repeat (EGF27) without the pelB leader sequence and without the green fluorescent fusion partner. This construct was expressed in both BL21-Gold(DE3) and Shuffle T7 Express cells. After purification, we found that the protein expressed in Shuffle cells yielded significantly higher amount of monomeric protein.

2.5 Future work

We have found that EGF27-6XHis expressed in Shuffle cells produce correctly folded, disulfide bonded protein with an increased yield. In our initial protein purification attempts for EGF27-6XHis expressed in Shuffle cells, we found that the protein forms oligomers while concentrating during sample preparation for NMR. Thus, in the next steps we need to find a way to keep the protein stable in its natively folded state without aggregation after purification. Once we successfully produce milligram quantities of properly folded protein, we will be able to run NMR experiments required for elucidating the three dimensional structure of the protein.

REFERENCES

1. Rana, N. A.; Haltiwanger, R. S., Fringe benefits: Functional and structural impacts of *O*-glycosylation on the extracellular domain of Notch receptors. *Current Opinion in Structural Biology* **2011,** *21* (5), 583-589.

2. Kopan, R.; Ilagan, M. X., The canonical Notch signaling pathway: Unfolding the activation mechanism. *Cell* **2009,** *137* (2), 216-33.

3. National Institutes of Health. U.S. National Library of Medicine. What are proteins and what do they do? http://ghr.nlm.nih.gov/handbook/howgeneswork/protein (accessed 15 Feb).

4. Clancy, S.; Brown, W., Translation: DNA to mRNA to protein. *Nature Education* **2008,** *1* (1), 101.

5. National Institutes of Health. National Human Genome Research Institute. Talking glossary of genetic terms. http://www.genome.gov/glossary/ (accessed 16 Feb).

6. Keck MS & Proteomics Resource. Yale School of Medicine. Post-translational modifications (PTMs). http://keck.med.yale.edu/proteomics/technologies/posttransmodifications/index.aspx (accessed 17 Feb).

7. Thermo Fisher Scientific Inc. Overview of post-translational modifications. https://www.lifetechnologies.com/us/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-post-translational-modification.html (accessed 16 Feb).

8. Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E.; editors., *Essentials of Glycobiology*. 2nd ed.; Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press: 2009.

9. Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease. *Cell* **2006,** *126* (5), 855-67.

10. Schimpl, M.; Zheng, X.; Borodkin, V. S.; Blair, D. E.; Ferenbach, A. T.; Schüttelkopf, A. W.; Navratilova, I.; Aristotelous, T.; Albarbarawi, O.; Robinson, D. A.; Macnaughtan, M. A.; van Aalten, D. M. F., O-GlcNAc transferase invokes nucleotide sugar pyrophosphate participation in catalysis. *Nature Chemical Biology* **2012,** *8* (12), 969-974.

11. Bedford, M., Methylation of proteins. In *Encyclopedic reference of genomics and proteomics in molecular medicine*, Springer Berlin Heidelberg: 2006; pp 1070-1075.

12. LifeTein LLC Modifications: N-Terminal Acetylation. http://www.lifetein.com/Peptide-Synthesis-Amidation-Acetylation.html (accessed 17 Feb).

13. Kershaw, N. J.; Church, N. L.; Griffin, M. D.; Luo, C. S.; Adams, T. E.; Burgess, A. W., Notch ligand Delta-like1: X-ray crystal structure and binding affinity. *The Biochemical Journal* **2015**.

14. Purow, B. W.; Haque, R. M.; Noel, M. W.; Su, Q.; Burdick, M. J.; Lee, J.; Sundaresan, T.; Pastorino, S.; Park, J. K.; Mikolaenko, I.; Maric, D.; Eberhart, C. G.; Fine, H. A., Expression of Notch-1 and its ligands, Delta-like-1 and Jagged-1, is critical for glioma cell survival and proliferation. *Cancer Research* **2005,** *65* (6), 2353-63.

15. Cordle, J.; Redfieldz, C.; Stacey, M.; van der Merwe, P. A.; Willis, A. C.; Champion, B. R.; Hambleton, S.; Handford, P. A., Localization of the delta-like-1-binding site in human Notch-1 and its modulation by calcium affinity. *The Journal of Biological Chemistry* **2008,** *283* (17), 11785-93.

16. Takeuchi, H.; Haltiwanger, R., Enzymatic analysis of the protein *O*-glycosyltransferase, Rumi, acting toward epidermal growth factor-like (EGF) repeats. In *Glycosyltransferases*, Brockhausen, I., Ed. Humana Press: 2013; Vol. 1022, pp 119-128.

17. Xu, A.; Lei, L.; Irvine, K. D., Regions of Drosophila Notch that contribute to ligand binding and the modulatory influence of Fringe. *The Journal of Biological Chemistry* **2005,** *280* (34), 30158-65.

18. Sharma, A.; Rangarajan, A.; Dighe, R. R., Antibodies against the extracellular domain of human Notch1 receptor reveal the critical role of epidermal-growth-factor-like repeats 25-26 in ligand binding and receptor activation. *The Biochemical Journal* **2013,** *449* (2), 519-30.

19. Cordle, J.; Johnson, S.; Tay, J. Z.; Roversi, P.; Wilkin, M. B.; de Madrid, B. H.; Shimizu, H.; Jensen, S.; Whiteman, P.; Jin, B.; Redfield, C.; Baron, M.; Lea, S. M.; Handford, P. A., A conserved face of the Jagged/Serrate DSL domain is involved in Notch trans-activation and cis-inhibition. *Nature Structural & Molecular Biology* **2008,** *15* (8), 849-57.

20. Chillakuri, C. R.; Sheppard, D.; Lea, S. M.; Handford, P. A., Notch receptor-ligand binding and activation: insights from molecular studies. *Seminars in Cell & Developmental Biology* **2012,** *23* (4), 421-8.

21. Okajima, T.; Xu, A.; Irvine, K. D., Modulation of Notch-ligand binding by protein *O*-fucosyltransferase 1 and Fringe. *The Journal of Biological Chemistry* **2003,** *278* (43), 42340-5.

22. Rand, M. D.; Grimm, L. M.; Artavanis-Tsakonas, S.; Patriub, V.; Blacklow, S. C.; Sklar, J.; Aster, J. C., Calcium depletion dissociates and activates heterodimeric Notch receptors. *Molecular and Cellular Biology* **2000,** *20* (5), 1825-35.

23. Rand, M. D.; Lindblom, A.; Carlson, J.; Villoutreix, B. O.; Stenflo, J., Calcium binding to tandem repeats of EGF-like modules. Expression and characterization of the EGF-like modules of human Notch-1 implicated in receptor-ligand interactions. *Protein Science : A Publication of the Protein Society* **1997,** *6* (10), 2059-71.

24. Gridley, T., Notch signaling and inherited disease syndromes. *Human Molecular Genetics* **2003,** *12 Spec No 1*, R9-13.

25. Downing, A. K.; Knott, V.; Werner, J. M.; Cardy, C. M.; Campbell, I. D.; Handford, P. A., Solution structure of a pair of calcium-binding epidermal growth factor-like domains:

Implications for the Marfan syndrome and other genetic disorders. *Cell* **1996,** *85* (4), 597-605.

26. Hambleton, S.; Valeyev, N. V.; Muranyi, A.; Knott, V.; Werner, J. M.; McMichael, A. J.; Handford, P. A.; Downing, A. K., Structural and functional properties of the human Notch-1 ligand binding region. *Structure* **2004,** *12* (12), 2173-2183.

27. Boswell, E. J.; Kurniawan, N. D.; Downing, A. K., Calcium-binding EGF-like domains. In *Handbook of Metalloproteins*, John Wiley & Sons, Ltd: 2006.

28. Rampal, R.; Arboleda-Velasquez, J. F.; Nita-Lazar, A.; Kosik, K. S.; Haltiwanger, R. S., Highly conserved *O*-fucose sites have distinct effects on Notch1 function. *The Journal of Biological Chemistry* **2005,** *280* (37), 32133-40.

29. Lee, T.; Jafar-Nejad, H., *O*-glucose glycans in *Drosophila* Notch signaling. In *Glycoscience: Biology and Medicine*, Endo, T.; Seeberger, P. H.; Hart, G. W.; Wong, C.-H.; Taniguchi, N., Eds. Springer Japan: 2014; pp 1-8.

30. Moloney, D. J.; Panin, V. M.; Johnston, S. H.; Chen, J.; Shao, L.; Wilson, R.; Wang, Y.; Stanley, P.; Irvine, K. D.; Haltiwanger, R. S.; Vogt, T. F., Fringe is a glycosyltransferase that modifies Notch. *Nature* **2000,** *406* (6794), 369-375.

31. Stanley, P., Regulation of Notch signaling by glycosylation. *Current Opinion in Structural Biology* **2007,** *17* (5), 530-535.

32. Morgan, W. D.; Birdsall, B.; Frenkiel, T. A.; Gradwell, M. G.; Burghaus, P. A.; Syed, S. E.; Uthaipibull, C.; Holder, A. A.; Feeney, J., Solution structure of an EGF module pair from the *Plasmodium falciparum* merozoite surface protein 1. *Journal of Molecular Biology* **1999,** *289* (1), 113-22.

33. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Research* **2000,** *28* (1), 235-42.

34. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004,** *25* (13), 1605-12.

35. Braakman, I.; Bulleid, N. J., Protein folding and modification in the mammalian endoplasmic reticulum. *Annual Review of Biochemistry* **2011,** *80*, 71-99.

36. D'Souza, B.; Miyamoto, A.; Weinmaster, G., The many facets of Notch ligands. *Oncogene* **2008,** *27* (38), 5148-5167.

37. Pei, Z.; Baker, N. E., Competition between Delta and the *Abruptex* domain of Notch. *BMC Developmental Biology* **2008,** *8*, 4.

38. Mass Spectrometry Laboratory. School of Chemical Sciences. University of Illinois at Urbana-Champaign. Matrix-Assisted Laser Desorption Ionization. http://scs.illinois.edu/massSpec/ion/maldi.php (accessed 14 Mar).

39. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **1995,** *6* (3), 277-93.

40. Johnson, B.; Blevins, R., NMR View: A computer program for the visualization and analysis of NMR data. *Journal of Biomolecular NMR* **1994,** *4* (5), 603-614.

41. Waldo, G. S.; Standish, B. M.; Berendzen, J.; Terwilliger, T. C., Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnology* **1999,** *17* (7), 691-5.

APPENDIX

Table of Contents

A) SEQUENCES:

1) Mouse Notch1 (mN1) EGF1-36 nucleotide sequence:

The mN1 EGF1-36 nucleotide sequence is shown below. The first three codons of each EGF repeats from EGF1 to EGF36 are shown in bold alphabets.

ggcttgagatgctcccagccaagtgggacctgcctgaatggaggtaggtgcgaagtggccagcggcactg
aagcctgtgtctgcagcggagcctttgtgggccaacgatgccag**gactccaat**ccttgcctcagcacacc
gtgtaagaatgctggaacgtgccacgttgtggaccatggtggcactgtggattatgcctgcagctgtccc
ctgggtttctctgggcccctctgcctgacacct**ctggacaac**gcctgcctggccaacccctgccgcaatg
ggggcacctgtgacctgctcactctcacagagtacaagtgccgctgcccaccagggtggtcaggaaaatc
atgtcag**caggctgac**ccctgtgcctccaaccctgtgccaatggtggccagtgcctgccctttgagtct
tcatacatctgtcgctgcccgcctggcttccatggccccacctgcaggcaa**gatgttaat**gagtgcagcc
agaaccctgggctgtgccgccatggaggcacctgccacaatgagatcggctcctatcgctgtgcctgccg
tgccacccatactggtccccactgtgaactg**ccctatgtg**ccctgcagcccctcaccctgccagaatgga
ggcacctgccgtcctacaggggacaccacccacgagtgtgcctgcttgccaggttttgctggacagaact
gtgaagaa**aatgtggat**gactgtccaggaaacaactgcaagaatgggggtgcctgtgtggacggcgtgaa
tacctacaattgccgctgcccaccggagtggacgggtcagtactgtacagag**gatgtggac**gaatgtcag
ctcatgcccaatgcctgccagaatggcggaacctgccacaacacacacggcggctacaactgtgtgtgtg
tcaatgggtggactggcgaggactgcagtgag**aacattgat**gactgtgccagtgccgcctgtttccaggg
tgccacttgccacgaccgtgtggcttccttctactgcgaatgtccgcatgggcgcacaggtctgctgtgc
cac**ctcaacgat**gcgtgcatcagcaacccctgcaacgagggctccaactgtgacaccaaccctgtcaacg
gcaaagccatctgcacctgcccctcggggtacacagggccagcctgcagccag**gacgtggat**gagtgtgc
tctgggtgccaacccttgtgagcacgcaggcaaatgcctcaacacactgggttcttttgagtgccagtgt
ctacagggctacacgggaccccgctgtgagatt**gatgttaat**gagtgcatctccaacccatgtcagaatg
atgccacttgcctggaccagattggggagttccaatgcatatgtatgccaggttatgaaggtgtatactg
tgaaatc**aacacggat**gagtgcgccagcagcccctgtctgcacaatggccactgcatggacaagatcaat
gagttccaatgtcagtgccccaaaggcttcaacgggcacctgtgccagtat**gatgtggat**gagtgtgcca
gcacaccatgcaagaacggtgccaagtgcctggatgggcccaacacctatacctgcgtgtgtacagaagg
ttacacagggacccactgcgaagtg**gacattgac**gagtgtgaccctgacccctgccactatggttcctgt
aaggatggtgtggccacctttacctgcctgtgccagccaggctacacaggccatcactgtgagacc**aaca**
tcaatgagtgccacagccaaccgtgccgccatgggggcacctgccaggaccgtgacaactcctacctctg
cttatgcctcaagggaaccacagggcccaactgtgagatc**aacctggat**gactgcgccagcaacccctgt
gactctggcacctgtctggacaagattgatggctacgaatgtgcctgtgaaccaggctacacaggaagca
tgtgtaacgtc**aacattgac**gaatgtgcgggcagcccctgccacaacgggggcacttgtgaggatggcat
cgcgggcttcacttgccgctgccccgagggctaccatgaccccacgtgcctgtccgag**gtcaacgag**tgc
aacagtaacccctgcatccacggagcttgccgggatggcctcaatgggtacaagtgtgactgtgcccctg
ggtggagtggaacaaactgtgacatc**aacaacaac**gagtgtgagtccaacccttgtgtcaacggtggcac
ctgcaaggacatgaccagtggctacgtatgcacctgccgagaaggcttcagtggccctaattgccagacc
aacatcaacgaatgtgcctccaacccctgcctgaaccaggggacctgcattgatgatgtcgctggataca
agtgcaactgtcctctgccatatacaggagccacgtgtgaggtg**gtgttggcc**ccatgtgctaccagccc
ctgcaaaaacagcggggtatgcaaggagtctgaagactatgagagtttttcctgtgtctgtcccacaggc
tggcaaggtcaaacctgcgaggtt**gacatcaat**gagtgtgtgaaaagcccatgtcgccatggggcctcct
gccagaacaccaatggcagctaccgctgcctctgccaggccggctatacaggtcgcaactgtgagagt**ga**
catcgatgactgccgccccaacccgtgtcacaatgggggttcctgcaccgatggcatcaacacagccttc
tgcgactgcctgcccggcttccagggtgccttctgtgaggag**gacatcaat**gaatgtgccagcaatccct
gccaaaatggtgccaattgcactgactgtgtggacagctacacatgtacctgccccgtgggcttcaatgg

catccactgcgagaac**aacacacct**gactgtactgagagctcctgcttcaatggtggtacctgtgtggat
ggtatcaactccttcacctgtctgtgtccacctggcttcacgggcagctactgtcagtat**gatgtcaat**g
agtgtgattcacggccctgtctgcacggtggtacctgccaagacagctatggtacttataagtgtacctg
cccacagggctacactggtctcaactgccagaac**cttgtgcgc**tggtgcgactcggctccctgcaagaat
ggtggcaggtgctggcagaccaacacgcagtaccactgtgagtgccgcagcggctggactggcgtcaact
gcgacgtg**ctcagtgtg**tcctgtgaggtggctgcacagaagcgaggcattgacgtcactctcctgtgcca
gcatggagggctctgtgtggatgagggagataaacattactgccactgccaggcaggctacacgggcagc
tactgtgaggac**gaggtggac**gagtgctcacctaacccctgccagaatggagctacctgcactgactatc
tcggcggcttttcctgcaagtgtgtggctggctaccatgggtctaactgctcagag**gagatcaac**gagtg
cctgtcccagccctgccagaatgggggtacctgcattgatctgaccaactcctacaagtgttcctgcccc
cgggggacacagggtgtacactgtgagatc**aatgttgat**gactgccatccccccctgaccctgcctccc
gaagccccaagtgcttcaacaatggcacctgtgtggaccaggtgggtggctatacctgcacctgcccacc
aggcttcgtcggggagcggtgtgagggt**gatgtcaat**gaatgtctctccaaccctgtgacccacgtggc
acccagaactgtgtgcagcgtgttaatgacttccactgcgagtgccgggctggccacactggacgccgct
gtgagtca**gtcatcaat**ggctgcaggggcaaaccttgcaagaatgggggtgtctgtgccgtggcctccaa
caccgcccgtggattcatctgtaggtgccctgcgggcttcgagggtgccacatgtgagaat**gatgcccgc**
acttgtggcagcttacgctgcctcaacggtggtacatgcatctcgggcccacgtagtcccacctgcctat
gcctgggatccttcaccggccctgagtgccagttccca**gccagcagc**ccctgtgtgggtagcaacccctg
ctacaatcagggcacctgtgagcccacatccgagaacccctttctaccgctgtctatgccctgccaaattc
aacgggctactgtgccacatcctgg

2) mN1 EGF1-36 amino acid sequence:

mN1 EGF1-36 sequence has 1408 amino acid residues. The first amino acid residue of each EGF repeat is shown in bold alphabet. Cysteine (C) at position 943 is highlighted in grey.

GVAALGSGSTGDAAQPARRARRTYEAS**G**LRCSQPSGTCLNGGRCEVASGTEACVCSGAFVGQRCQ**D**SNPC
LSTPCKNAGTCHVVDHGGTVDYACSCPLGFSGPLCLTP**L**DNACLANPCRNGGTCDLLTLTEYKCRCPPGW
SGKSCQ**Q**ADPCASNPCANGGQCLPFESSYICRCPPGFHGPTCRQ**D**VNECSQNPGLCRHGGTCHNEIGSYR
CACRATHTGPHCEL**P**YVPCSPSPCQNGGTCRPTGDTTHECACLPGFAGQNCEE**N**VDDCPGNNCKNGGACV
DGVNTYNCRCPPEWTGQYCTE**D**VDECQLMPNACQNGGTCHNTHGGYNCVCVNGWTGEDCSE**N**IDDCASAA
CFQGATCHDRVASFYCECPHGRTGLLCH**L**NDACISNPCNEGSNCDTNPVNGKAICTCPSGYTGPACSQ**D**V
DECALGANPCEHAGKCLNTLGSFECQCLQGYTGPRCEI**D**VNECISNPCQNDATCLDQIGEFQCICMPGYE
GVYCEI**N**TDECASSPCLHNGHCMDKINEFQCQCPKGFNGHLCQY**D**VDECASTPCKNGAKCLDGPNTYTCV
CTEGYTGTHCEV**D**IDECDPDPCHYGSCKDGVATFTCLCQPGYTGHHCET**N**INECHSQPCRHGGTCQDRDN
SYLCLCLKGTTGPNCEI**N**LDDCASNPCDSGTCLDKIDGYECACEPGYTGSMCNV**N**IDECAGSPCHNGGTC
EDGIAGFTCRCPEGYHDPTCLSE**V**NECNSNPCIHGACRDGLNGYKCDCAPGWSGTNCDI**N**NNECESNPCV
NGGTCKDMTSGYVCTCREGFSGPNCQT**N**INECASNPCLNQGTCIDDVAGYKCNCPLPYTGATCEV**V**LAPC
ATSPCKNSGVCKESEDYESFSCVCPTGWQGQTCEV**D**INECVKSPCRHGASCQNTNGSYRCLCQAGYTGRN
CES**D**IDDCRPNPCHNGGSCTDGINTAFCDCLPGFQGAFCEE**D**INECASNPCQNGANCTDCVDSYTCTCPV
GFNGIHCEN**N**TPDCTESSCFNGGTCVDGINSFTCLCPPGFTGSYCQY**D**VNECDSRPCLHGGTCQDSYGTY
KCTCPQGYTGLNCQN**L**VRWCDSAPCKNGGRCWQTNTQYHCECRSGWTGVNCDV**L**SVSCEVAAQKRGIDVT
LLCQHGGLCVDEGDKHYCHCQAGYTGSYCED**E**VDECSPNPCQNGATCTDYLGGFSCKCVAGYHGSNCSE**E**
INECLSQPCQNGGTCIDLTNSYKCSCPRGTQGVHCEI**N**VDDCHPPLDPASRSPKCFNNGTCVDQVGGYTC
TCPPGFVGERCEG**D**VNECLSNPCDPRGTQNCVQRVNDFHCECRAGHTGRRCES**V**INGCRGKPCKNGGVCA
VASNTARGFICRCPAGFEGATCEN**D**ARTCGSLRCLNGGTCISGPRSPTCLCLGSFTGPECQFP**A**SSPCVG
SNPCYNQGTCEPTSENPFYRCLCPAKFNGLLCHI**L**DTRGGPEQKLISEEDLNSAVDHHHHL

3) EGF26-27-28 nucleotide sequence:

aacacacctgactgtactgagagctcctgcttcaatggtggtacctgtgtggatggtatcaactccttca
cctgtctgtgtccacctggcttcacgggcagctactgtcagtat**gatgtcaat**gagtgtgattcacggcc
ctgtctgcacggtggtacctgccaagacagctatggtacttataagtgtacctgcccacagggctacact
ggtctcaactgccagaac**cttgtgcgc**tggtgcgactcggctccctgcaagaatggtggcaggtgctggc
agaccaacacgcagtaccactgtgagtgccgcagcggctggactggcgtcaactgcgacgtg

4) EGF26-27-28-6XHis translated amino acid sequence from pET-28b vector:

EGF26-27-28 sequence is underlined below. EGF26-28 has amino acid residues from E961 to S1078 based on 1408 residues of EGF1-36.

MGENNTPDCTESSCFNGGTCVDGINSFTCLCPPGFTGSYCQYDVNECDSRPCLHGGTCQDSYGTYKCTCP
QGYTGLNCQNLVRWCDSAPCKNGGRCWQTNTQYHCECRSGWTGVNCDVLSLEHHHHHH

Theoretical pI/MW: 5.63 / 14118.48

5) EGF25-26-27 nucleotide sequence:

gacatcaatgaatgtgccagcaatccctgccaaaatggtgccaattgcactgactgtgtggacagctaca
catgtacctgccccgtgggcttcaatggcatccactgcgagaac**aacacacct**gactgtactgagagctc
ctgcttcaatggtggtacctgtgtggatggtatcaactccttcacctgtctgtgtccacctggcttcacg
ggcagctactgtcagtat**gatgtcaat**gagtgtgattcacggccctgtctgcacggtggtacctgccaag
acagctatggtacttataagtgtacctgcccacagggctacactggtctcaactgccagaac

6) EGF25-26-27-6XHis translated amino acid sequence from pET-28b vector:

EGF25-26-27 sequence is underlined below. EGF25-27 has amino acid residues from E923 to V1040 based on 1408 residues of EGF1-36.

MGEEDINECASNPCQNGANCTDCVDSYTCTCPVGFNGIHCENNTPDCTESSCFNGGTCVDGINSFTCLCP
PGFTGSYCQYDVNECDSRPCLHGGTCQDSYGTYKCTCPQGYTGLNCQNLVLEHHHHHH

Theoretical pI/MW: 4.53 / 13824.04

7) EGF25-26-C21G nucleotide sequence:

gacatcaatgaatgtgccagcaatccctgccaaaatggtgccaattgcactgactgtgtggacagctaca
catgtacctgccccgtgggcttcaatggcatccactgcgagaac**aacacacct**gactgtactgagagctc
ctgcttcaatggtggtacctgtgtggatggtatcaactccttcacctgtctgtgtccacctggcttcacg
ggcagctactgtcagtat

8) EGF25-26-C21G-6XHis translated amino acid sequence from pET-28a vector:

EGF25-26-C21G sequence is underlined below. EGF25-26-C21G has amino acid residues from E924 to Q999 based on 1408 residues of EGF1-36. Cysteine (C) at position 943 based on EGF1-36 or at position 21 based on EGF25-26 sequence is mutated to Glycine (G), which is highlighted in grey.

MEDINECASNPCQNGANCTDGVDSYTCTCPVGFNGIHCENNTPDCTESSCFNGGTCVDGINSFTCLCPPG
FTGSYCQKHHHHHH

Theoretical pI/MW: 4.94 / 8987.75

9) EGF27 nucleotide sequence:

gatgtcaatgagtgtgattcacggccctgtctgcacggtggtacctgccaagacagctatggtacttata
agtgtacctgcccacagggctacactggtctcaactgccagaac

10) EGF27-6XHis translated amino acid sequence from pET-28a vector:

EGF27 sequence is underlined below. EGF27 has amino acid residues from D1001 to W1042 based on 1408 residues of EGF1-36.

MDIGINSDPNSDVNECDSRPCLHGGTCQDSYGTYKCTCPQGYTGLNCQNLVRWAAALEHHHHHH

Theoretical pI/MW: 5.87 / 7109.76

11) EGF25-26-C21G-TEV-EGFP-6XHis translated amino acid sequence from pLIC-EGFP vector:

EGF25-26-C21G-TEV-EGFP sequence is underlined below.

MGTSNPFKLMEDINECASNPCQNGANCTDGVDSYTCTCPVGFNGIHCENNTPDCTESSCFNGGTCVDGIN
SFTCLCPPGFTGSYCQKLGLINLENLYFQSVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYG
KLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRA
EVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLAD
HYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKHHHHHHHH

Theoretical pI/MW: 5.50 / 38536.20

Protein sequences after TEV protease:

EGF25-26:

TEV protease recognition sequence is underlined below.

MGTSNPFKLMEDINECASNPCQNGANCTDGVDSYTCTCPVGFNGIHCENNTPDCTESSCFNGGTCVDGIN
SFTCLCPPGFTGSYCQKLGLINLENLYFQ

Theoretical pI/MW: 3.92 / 10559.71

EGFP-His:

SVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGH
KLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSK
DPNEKRDHMVLLEFVTAAGITLGMDELYKHHHHHHHH

Theoretical pI/MW: 6.07 / 27994.50

12) <u>pLIC-EGF27-TEV-EGFP-6XHis translated amino acid sequence from pLIC-EGFP vector</u>:

EGF27 sequence is underlined below.

MGTSNPFKLM<u>DVNECDSRPCLHGGTCQDSYGTYKCTCPQGYTGLNCQNLVRW</u>RKLGLINLENLYFQSVSK
GEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSR
YPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEY
NYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNE
KRDHMVLLEFVTAAGITLGMDELYKHHHHHHHH

Theoretical pI/Mw: 6.21 / 35474.03

<u>Protein sequences after TEV protease:</u>

EGF27:

TEV protease recognition sequence is underlined below.

MGTSNPFKLMDVNECDSRPCLHGGTCQDSYGTYKCTCPQGYTGLNCQNLVRWRKLGLINL<u>ENLYFQ</u>

Theoretical pI/Mw: 7.57 / 7497.55

EGFP-His:

SVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGH
KLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSK
DPNEKRDHMVLLEFVTAAGITLGMDELYKHHHHHHHH

Theoretical pI/Mw: 6.07 / 27994.50

13) <u>pelB-EGF27-TEV-EGFP-6XHis translated amino acid sequence from pLIC-EGFP vector</u>:

EGF27 sequence is underlined below.

MGTSNPFKLMKYLLPTAAAGLLLLAAQPAMAMDIGINSDPNS<u>DVNECDSRPCLHGGTCQDSYGTYKCTCP
QGYTGLNCQNLVRW</u>RKLGLINLENLYFQSVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGK
LTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAE
VKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADH
YQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKHHHHHHHH

Theoretical pI/Mw: 6.12 / 38697.82

Periplasmic proteolytic digestion of pelB:

EGF27 sequence is underlined below.

MDIGINSDPNSDVNECDSRPCLHGGTCQDSYGTYKCTCPQGYTGLNCQNLVRWRKLGLINLENLYFQSVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFS
RYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLE
YNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPN
EKRDHMVLLEFVTAAGITLGMDELYKHHHHHHHH

Theoretical pI/Mw: 5.95 / 35510.90

Protein sequences after TEV protease:

EGF27:

TEV protease recognition sequence is underlined below.

MDIGINSDPNSDVNECDSRPCLHGGTCQDSYGTYKCTCPQGYTGLNCQNLVRWRKLGLINLENLYFQ

Theoretical pI/Mw: 4.94 / 7534.42

pelB-EGF27:

TEV protease recognition sequence is underlined below.

MGTSNPFKLMKYLLPTAAAGLLLLAAQPAMAMDIGINSDPNSDVNECDSRPCLHGGTCQDSYGTYKCTCP
QGYTGLNCQNLVRWRKLGLINLENLYFQ

Theoretical pI/Mw: 6.51 / 10721.34

EGFP-His:

SVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGH
KLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSK
DPNEKRDHMVLLEFVTAAGITLGMDELYKHHHHHHHH

Theoretical pI/Mw: 6.07 / 27994.50

B) COPYRIGHT MATERIAL – PERMISSION FOR REUSE:

ELSEVIER LICENSE
TERMS AND CONDITIONS

Mar 12, 2015

This is a License Agreement between Amid Paudyal ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Amid Paudyal
Customer address	427 Choppin Hall, Dept. of Chemistry BATON ROUGE, LA 70820
License number	3562081468024
License date	Feb 04, 2015
Licensed content publisher	Elsevier
Licensed content publication	Current Opinion in Structural Biology
Licensed content title	Fringe benefits: Functional and structural impacts of O-glycosylation on the extracellular domain of Notch receptors
Licensed content author	Nadia A Rana,Robert S Haltiwanger
Licensed content date	October 2011
Licensed content volume number	21
Licensed content issue number	5
Number of pages	7
Start Page	583
End Page	589
Type of Use	reuse in a thesis/dissertation

Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figures 1, 2
Title of your thesis/dissertation	Expression and characterization of mouse Notch1 Abruptex EGF repeats in E. coli
Expected completion date	May 2015
Estimated size (number of pages)	50
Elsevier VAT number	GB 494 6272 12
Price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	**0.00 USD**

VITA

Mr. Amid Paudyal is from Kathmandu, Nepal. After completing his high school from the National School of Sciences in Kathmandu, he came to USA for his higher education. He completed his Bachelor of Science degree in Chemistry Pre-Medicine from Nicholls State University in Thibodaux, Louisiana in 2010. He got exposure into research while working with Prof. Henry Foust, Civil Engineer, at Nicholls in a project over a course of a summer in a summer undergraduate research experience program. With Prof. Foust, he researched to develop a mathematical model that relate to the salinity intrusion in Bayou Lafourche using vector auto-regression analysis method. In 2012, he joined Prof. Megan Macnaughtan's research group at LSU.

Amid is currently in office as a Secretary for the LSU Chemistry Graduate Student Council (CGSC). He is also a student member of the American Chemical Society (ACS) since 2012. He has volunteered for several activities such as chem demo for Pre-K children at Trinity Day School in Baton Rouge, chem demo at LSU and Nicholls, and Fourcheon beach dune restoration project. Amid will graduate with a Master of Science degree in Analytical Chemistry from Louisiana State University in May 2015.