10-1-2020

Expression, purification, and glycosylation of epidermal growth factor-like repeat 27 from mouse NOTCH1

Justin A. Grennell
Louisiana State University

Kendra D. Jenkins
Louisiana State University

Huimin Zhong
Louisiana State University

Amid Paudyal
Louisiana State University

Kelvin B. Luther
University of Georgia

See next page for additional authors

Recommended Citation

Grennell, J., Jenkins, K., Zhong, H., Paudyal, A., Luther, K., Haltiwanger, R., & Macnaughtan, M. (2020). Expression, purification, and glycosylation of epidermal growth factor-like repeat 27 from mouse NOTCH1. *Protein Expression and Purification, 174* https://doi.org/10.1016/j.pep.2020.105681

Expression, Purification, and Glycosylation of Epidermal Growth Factor-like Repeat 27 from Mouse NOTCH1

Justin A. Grennell,[†] Kendra D. Jenkins,[†] Huimin Zhong,[†] Amid Paudyal,[†] Kelvin B. Luther,[‡] Robert S. Haltiwanger,[‡] Megan A. Macnaughtan[†*]

[†] Department of Chemistry, Louisiana State University, Baton Rouge, LA 70803, United States

[‡] Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, United States

Corresponding Author

* Louisiana State University, Department of Chemistry, Choppin Hall, Baton Rouge, LA 70803. E-mail: macnau@lsu.edu.

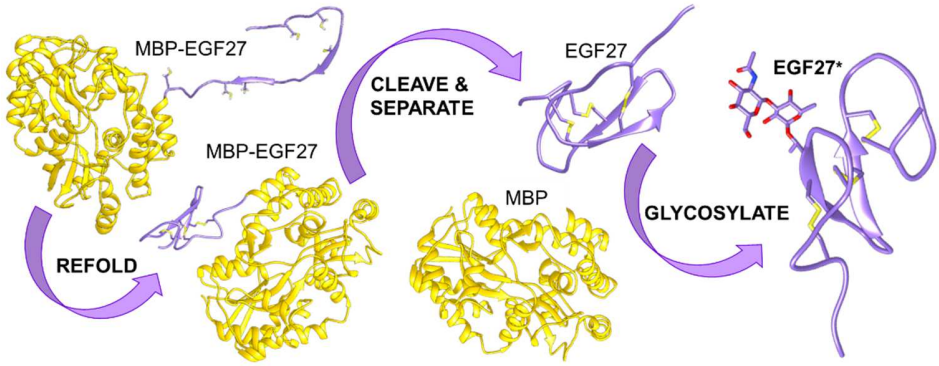
ABSTRACT

Notch receptors have large extracellular domains containing up to 36 tandem epidermal growth factor-like (EGF) repeats, which facilitate cell signaling by binding ligands on neighboring cells. Notch receptors play major roles in a variety of developmental processes by controlling cell fate decisions. Each EGF repeat consists of about 40 amino acids with 3 conserved disulfide bonds. Many of the EGF repeats are modified by *O*-linked fucose glycans, and more than half have calcium-binding sites, but the sequences of the EGF repeats vary giving distinct roles to each repeat. EGF repeat 27 (EGF27) from mouse NOTCH1 is modified with *O*-fucose and is 1 of 7 repeats that is differentially modified by specific Fringe enzymes, which are

known to regulate NOTCH1 activation and ligand binding. To better understand the role of EGF27 in NOTCH1 function and regulation, the 3-dimensional structures of EGF27 and its glycoforms are being pursued. *E. coli* cells were used to produce EGF27 in sufficient quantities for nuclear magnetic resonance analysis. Previous attempts to express the repeat alone and refold the repeat under a steady redox environment were unsuccessful due to low yields and extensive mixed-disulfide bond cross-linking. A new strategy using a cleavable maltose binding protein fusion tag increased the solubility and yield of EGF27. With the fusion tag, EGF27 was refolded to produce the correct disulfide bond arrangement, which was verified enzymatically with the glycosyltransferases, Protein *O*-fucosyltransferase 1 (POFUT1) and Lunatic Fringe.

KEYWORDS: Notch, EGF, glycosylation, Fringe, *Abruptex*, NMR

GRAPHICAL ABSTRACT



ABBREVIATIONS

ECD, extracellular domain; LBD, ligand binding domain; EGF, epidermal growth factor-like; Fuc, fucose; GlcNAc, *N*-acetylglucosamine; Gal, galatose; NeuAc, *N*-acetylneuraminic acid; Xyl, xylose; Glc, glucose; NMR, nuclear magnetic resonance; POFUT1, Protein *O*-fucosyltransferase 1; TEV, tobacco etch virus protease; MBP, maltose binding protein; His_6, 6X-

histidine tag; IPTG, isopropyl β-D-1-thiogalactopyranoside; Tris, tris(hydroxymethyl)aminomethane hydrochloride; Ni-IMAC, nickel-charged immobilized metal affinity chromatography; FPLC, fast protein liquid chromatography system; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; TFA, trifluoroacetic acid; MALDI, matrix-assisted laser desorption ionization; MS, mass spectrometry; LFNG, Lunatic Fringe; DTT, dithiothreitol; HPLC, high performance liquid chromatography; EICs, extracted ion chromatograms.

HIGHLIGHTS

- Recombinant expression of mouse NOTCH1 EGF27 in *E. coli* using M9 medium
- TEV cleavable MBP fusion avoids intermolecular disulfide cross-linking
- Steady redox refolding increases the yield of the correct disulfide isomer
- Refolded EGF27 is a substrate of Protein *O*-fucosyltransferase 1 and Lunatic Fringe
- Suitable yields for biophysical and structural biology studies

1. Introduction

Notch receptors are transmembrane proteins that play an integral role in the development of metazoans [1]. Notch receptors are comprised of a large extracellular domain (ECD) including a negative regulatory region, a transmembrane domain, and an intracellular domain. The signaling of Notch is activated by receptor–ligand interactions between the ECD and ligands of the Delta or Serrate/Jagged families on neighboring cells. Upon activation, proteolysis releases the Notch intracellular domain, which translocates to the nucleus for the transcription of Notch targets. Two regions in the ECD that have been implicated in ligand binding and Notch function are the *ligand binding domain* (LBD) and the *Abruptex* region (Fig. 1) [2-6]. Recent structural studies reveal direct contacts between residues in the LBD with ligand [6, 7]. The *Abruptex* region is defined by a series of mutations in *Drosophila* Notch that result in hyperactivation of Notch activity [5].

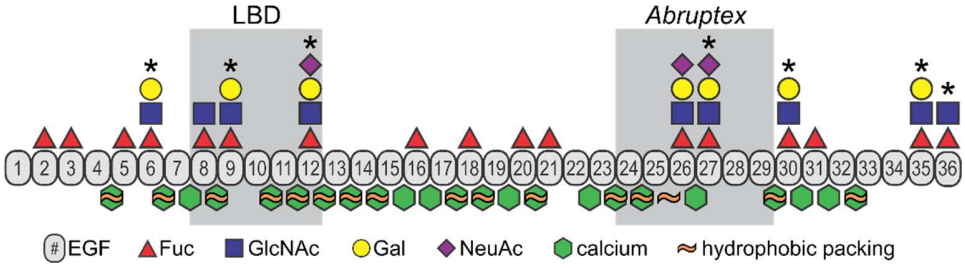


Fig. 1. Domain map of mouse NOTCH1 ECD. Epidermal growth factor-like (EGF) repeats are represented by numbered ovals, and the *ligand binding domain* (LBD) and *Abruptex* regions are indicated with gray boxes. EGF repeats modified with *O*-fucosylation at the consensus sequence, $C^2X_4(\underline{S/T})C^3$, where C^n is the n^{th} conserved cysteine in the EGF repeat sequence, are indicated by red triangles (Fuc, fucose) [8, 9]. Elongation of *O*-fucose at selective EGF repeats with *N*-acetylglucosamine (GlcNAc, blue square), galactose (Gal, yellow circle), and *N*-acetylneuraminic acid (NeuAc, purple diamond) is based on data from Kakuda and Haltiwanger, 2017 [9]. Asterisks indicate *O*-fucosylated EGF repeats that show different glycan elongation as a result of specific Fringe activity [9]. The green hexagons indicate EGF repeats with a calcium binding consensus sequence of $(D/N)X(D/N)(D/Q)C^1X_nC_3X(D/N)$ [10]. An orange tilde (~) indicates tandem repeats that contain interdomain hydrophobic packing residues with a consensus sequence of $(F/Y/W)X_4C^{6'}X_nC^3X(D/N)X(I/L/V/P)$, where $C^{n'}$ and C^n are the conserved cysteine residues in the *N*-terminal and *C*-terminal repeat sequences, respectively [11].

Notch receptor ECDs contain up to 36 tandem epidermal growth factor-like (EGF) repeats. EGF repeats are relatively small domains that contain about 40 amino acids with six cysteines forming three disulfide bonds in a conserved arrangement: C^1–C^3, C^2–C^4, and C^5–C^6, where C^n is the n^{th} conserved cysteine in the repeat sequence. The degree of rigidity between tandem EGF repeats is variable and attributed to calcium binding and inter-repeat hydrophobic packing. Calcium binding sites are in conserved positions along the ECD and occur in the linker region between many of the tandem EGF repeats (Fig. 1, green hexagons) [10, 12]. Structural studies have shown that calcium binding rigidifies the linker between repeats, and inter-repeat hydrophobic packing (Fig. 1, orange tildes) can further stabilize the linker region [11, 13, 14]. The combination of a calcium binding site and inter-repeat hydrophobic packing was shown to increase the affinity for calcium [13, 15]. In the absence of a calcium binding consensus sequence, the linker region is flexible [16]. The highest concentration of predicted flexibility in mouse NOTCH1 occurs at both ends of the ECD and in the *Abruptex* region (Fig. 1). Several Notch EGF repeats also contain conserved consensus sequences for three types of *O*-linked glycans: *O*-fucose, *O*-glucose, and *O*-*N*-acetylglucosamine [17-19]. The *O*-glucose form is typically the trisaccaharide, Xyl-α1,3-Xyl-α1,3-Glc (Xyl, xylose; Glc, glucose), and the *O*-fucose form (Fig. 1, red triangles) can be elongated to di-, tri-, and tetrasaccharide (NeuAc-α2,3/6-Gal-β1,4-GlcNAc-β1,3-Fuc) forms.

Glycosylation of the EGF repeats directly affects the regulation of the Notch signaling pathway [20-22]. Many of the EGF repeats (17 of 36) in mouse NOTCH1 are modified by *O*-fucose, and a subset of these are elongated with a β1,3-linked GlcNAc by one or more of the Fringe enzymes, Lunatic, Manic, or Radical Fringe [9]. Modification by each Fringe results in specific elongation profiles of the NOTCH1 ECD with differences at 7 EGF repeats (Fig. 1,

asterisks). The Fringe enzymes are known modulators of Notch signaling [9, 23]. Mutations of the *O*-fucose sites in the LBD and the *Abruptex* region directly affect NOTCH1 activity and reduce the ability of Fringes to modulate its activity [9, 24]. While the mutations in the *Abruptex* region had variable effects depending on the cells used for the Notch activity assays, loss of the *O*-fucose modification on EGF12 (part of the LBD) had significant effects on Notch function in cell-based assays and *in vivo* [9, 23-25]. The recent co-crystal structures of NOTCH1 LBD and portions of DELTA-LIKE 4 or JAGGED1 show that the *O*-fucose on EGF12 is in direct contact with ligand, providing a structural explanation for the importance of this *O*-fucose in Notch function [6, 7].

Less is known about the *Abruptex* region since no structures are currently available for any of the EGF repeats in this part of the Notch ECD. Our group is interested in studying the 3-dimensional structure of this region to understand the role of glycosylation on Notch function. Our first target is mouse NOTCH1 EGF27, which is modified by *O*-fucose and elongated with a β1,3-GlcNAc by Fringes in cells [9, 24]. Lunatic Fringe (LNFG) and Manic Fringe modify the *O*-fucose resulting in a mixture of trisaccharide and tetrasaccharide glycans on EGF27 (Fig. 1) [9]. Radical Fringe does not modify *O*-fucose on EGF27, and elimination of the *O*-fucose site alters NOTCH1 activity in cell based assays [9, 24]. Nuclear magnetic resonance (NMR) spectroscopy is a great tool for studying glycosylated proteins, which tend to be dynamic and not crystallize well. For NMR studies, proteins are typically produced with ^{15}N- and ^{13}C-isotopes through metabolic labeling using *E. coli*. For EGF repeats, *E. coli* expression does not always produce the protein with the correct disulfide bond arrangement and can result in a mixture of disulfide isoforms and highly cross-linked protein (via intermolecular disulfide bonds). Bacterial expression of EGF27 results in two major disulfide isoforms, but the correct form is selectively

glycosylated by Protein *O*-fucosyltransferase 1 (POFUT1) and UDP-glucose:protein *O*-glucosyltransferase 1, POGLUT1/Rumi [26, 27], giving a means of identifying the native isoform. In addition, the small size of the EGF repeats (5-7 kDa) can make purification and sample concentrating difficult. The yield of protein has to be high (hundreds of micrograms to milligrams per liter culture) due to the cost of labeled media. Our group has previously attempted to produce properly folded EGF27 using a variety of strategies, including varying the *N*-terminal sequence, using enhanced green fluorescence protein as a cleavable fusion tag, resolubilization and refolding methods, and varying the *E. coli* expression strain and growth conditions [28, 29]. In this article, we present a new method for producing EGF27 in high yield with the native disulfide bond arrangement by using a tobacco etch virus protease (TEV) cleavable maltose binding protein (MBP) fusion and redox refolding.

2. Materials and methods

2.1. Materials

The recombinant expression plasmid, pMHTΔ238, was purchased from the DNASU Plasmid Repository (clone #TvCD00084286) [30] and encodes the self-cleaving *N*-terminal fusion protein of MBP, a 7X-histidine tag, and a mutated and truncated form of TEV protease [31]. TEV protease was expressed from the pMHTΔ238 plasmid in Rosetta 2(DE3) *E. coli* cells as described in Blommel and Fox with a few modifications [31] (detail in the Supplementary Data). pET-28a(+) vector (catalog #69864-3) and Rosetta 2(DE3) *E. coli* cells (catalog #71397-3) were purchased from Novagen (EMD Millipore). pMAL-c5X vector (catalog #N8108S) was purchased from New England Biolabs. All water used was supplied from a Millipore Direct-Q 3 ultrapure water system.

2.2. EGF27 expression vectors

The pMAL-TEV-EGF27 expression plasmid was produced through a series of cloning experiments starting with pET20b-EGF27 described in Takeuchi *et al.* [26]. The expression plasmid, pET28-EGF27, was produced by subcloning *EGF27* from pET20b-EGF27 using NcoI and XhoI restriction sites [29]. pET28-EGF27 encodes the EGF27 domain (mouse NOTCH1 (UniProtKB Q01705) amino acids 1021-1062) with a *C*-terminal 6X-histidine tag (His_6). The pMAL-TEV-EGF27 plasmid was produced by amplifying *EGF27* and the *C*-terminal His_6 sequences from pET28-EGF27 using the primers: 5'-CGTCCATATGGAAAACCTGTACTTC-CAGTCGGATGTCAATGAGTGT and 3'-CGCAGTCGACGTTAGCAGCCGGATCTCA. The forward primer included an NdeI restriction site followed by the sequence that encodes the tobacco etch virus recognition sequence, and the reverse primer included a SalI restriction site. The amplicon and pMAL-c5x were cut with NdeI and SalI restriction enzymes followed by ligation. The resulting pMAL-TEV-EGF27 plasmid encodes the fusion protein referred to as MBP-EGF27 and includes the cytosolic maltose binding protein followed by a TEV protease cleavage site, the EGF27 domain (amino acids 1021-1062), and a *C*-terminal His_6. All plasmid sequences were confirmed by sequencing in the Louisiana State University GeneLab.

2.3. MBP-EGF27 protein expression and purification

MBP-EGF27 was expressed in Rosetta 2(DE3) *E. coli* cells in minimal M9 medium with ampicillin (50 μg/mL) and chloramphenicol (35 μg/mL) for antibiotic selection. Minimal M9 medium is necessary to produce metabolically labeled protein for NMR studies. Even though magnetic isotopes were not incorporated here, testing the yield from minimal M9 medium is

necessary due to the significant differences often observed between expression in LB versus M9 media. The strategy described by Marley *et al.* [32], where the expression cells are grown in rich medium and then exchanged into minimal medium before induction, was implemented. LB medium (1 L) was inoculated with 50 mL of starter culture grown from a single colony at 37 °C with 250 rpm shaking for 16 h. The 1 L cultures were incubated at 37 °C with 250 rpm shaking until an optical density of 0.5 - 0.6 at 600 nm was reached. The LB medium was exchanged to minimal M9 medium [M9 salts (50 mM Na_2HPO_4, 20 mM KH_2PO_4, 10 mM NaCl, 18 mM (1.0 g/L) NH_4Cl), 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1X BME vitamins, 3.0 mM (1.0 mg/L) thiamine hydrochloride, 1X trace metal solution [33], and 22 mM (4.0 g/L) glucose], and the culture was incubated at 37 °C with 250 rpm shaking for 1 h. The expression of MBP-EGF27 was induced by adding 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) to the minimal medium. The cultures were incubated at 16 °C with 250 rpm shaking for 16 h. The cells were harvested by centrifugation at 3,000 x *g* for 10 min at 4 °C and suspended in 20 mL of Tris buffer (125 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris) and 175 mM NaCl at pH 7.0). The cells were either stored at -80 °C or processed for protein purification.

Suspension of bacterial cell pellets (20 mL) from the expressions were prepared for lysis by adding 1.5 mL of CelLytic B cell lysis reagent, ½ SigmaFAST EDTA-free protease inhibitor tablet, 1 μL (250 units) of endonuclease, and approximately 5 mg of phenylmethane sulfonyl fluoride. The cells were lysed using a French pressure cell followed by centrifugation at 25,000 x *g* for 30 min at 4 °C to obtain soluble protein in the cleared lysate. The cleared lysate was loaded onto 5 mL of nickel-charged immobilized metal affinity chromatography (Ni-IMAC) resin in a 1 cm inner diameter column at a flow rate of 1 mL/min using a fast protein liquid chromatography (FPLC) system. The captured proteins were washed with 30 mL of Tris wash

buffer (50 mM Tris, 300 mM NaCl, 5% glycerol, and 5 mM imidazole at pH 7.0), and eluted with 20 mL Tris elution buffer (50 mM Tris, 300 mM NaCl, 5% glycerol, and 250 mM imidazole at pH 7.0). The eluted proteins were detected using absorbance at 280 nm and pooled from 2 mL fractions (typically 8 - 10 mL). The Bradford protein assay was used to analyze the protein concentration following the 96-well plate instructions (Pierce Comassie Plus Assay Kit, ThermoFisher Scientific). The Tris elution buffer was exchanged to 10 mM ammonium bicarbonate buffer (1 L reservoir) using dialysis tubing at 12 °C for 16 h with stirring.

Samples of the expression cultures, purification steps, and purified proteins were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Sample aliquots (15 μL) were mixed with 15 μL of 2X Laemmli buffer with 5% (by volume) of β-mercaptoethanol, heated at 85 °C for 5 min, and loaded into 4-20% Tris/glycine gels. The gels were run at 100 V for 60 min in running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3 buffer) followed by staining with brilliant blue R-250 solution (1.0 g/L brilliant blue R-250 dye, 50% methanol, 40% water, and 10% acetic acid) for 1 h. After de-staining in 50% water, 40% methanol, 10% acetic acid, the gels were imaged using a Bio-Rad Gel Doc EZ System.

2.4. Refolding by disulfide isomerization and cleavage of the MBP-EGF27 protein

Non-native disufide bonds in EGF27 within the MBP-EGF27 fusion protein were dissociated with reducing agent and reformed under a steady redox environment [34]. The disulfide bonds were reduced by adding 15 mM dithiothreitol (DTT) to 3.0 mg/mL of MBP-EGF27 in 10 mM ammonium bicarbonate buffer and incubating at 12 °C for 30 min. The DTT-treated MBP-EGF27 protein (8 – 10 mL) was placed in dialysis tubing and suspended in 100 mL of redox refolding buffer (5 mM reduced glutathione, 0.5 mM oxidized glutathione, 10 mM ammonium bicarbonate, and 0.5 M L-arginine at pH 5.8) overnight at 12 °C with stirring.

The cleavage of the MBP-EGF27 protein with TEV protease was performed at an enzyme: substrate ratio of 1:100 (mass:mass). An appropriate volume of TEV protease stock solution (Supplemental Data) was added to the refolded MBP-EGF27, and the mixture was incubated overnight at 4 °C. An aliquot of the mixture was analyzed by SDS-PAGE analysis, as described earlier, to assess cleavage.

2.5. Purification, analysis, and validation of EGF27 isomers

EGF27 was purified from the mixture containing MBP and TEV protease by reverse phase high performance liquid chromatography (HPLC). The mixture was prepared for HPLC by adding 0.1% TFA and 5% acetonitrile. Aliquots of the mixture (2 mL) were injected into a semi-prep C18 column (300 Å, 5 μm, 10 x 250 mm) and eluted with linear gradients of solvent A (0.1% trifluoroacetic acid (TFA) in water) and solvent B (0.1% TFA in acetonitrile). The elution profile was 0% to 45% solvent B at a flow rate of 1 mL/min followed by 45% to 90% solvent B at a flow rate of 2 mL/min for 60 min. The eluent was monitored with a diode-array detector at 210 nm and 280 nm and collected as 100 μL aliquots in 96-well plates. Peaks eluting between 35 – 55 min were collected and lyophilized. Peak areas were calculated with the PeakFit software using the second derivative method with Gaussian peak shapes, variable amplitudes and widths, cubic baseline correction, and 1% smoothing.

Matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS) was used to determine the molecular weight of the species in each collected HPLC sample. On a MALDI target plate, 1 μL of the sample was mixed with 1 μL of saturated sinapinic acid in 50% acetonitrile/0.1% TFA in water. The mixture was co-crystallized by evaporation at room temperature. Data acquisition was performed on a Bruker UltrafleXtreme MALDI TOF/TOF in

the reflectron positive ion mode with 1000 shots signal-averaged and a molecular weight range of 4000 – 7000 Da.

2.6. Validation of EGF27 isomer structure by in vitro *glycosylation*

Production and purification of POFUT1-myc-His_6 and Lunatic Fringe (LFNG)-Myc-His_6 enzyme were performed as previously described in Luo *et al.* and Luther *et al.* [35, 36]. The *in vitro* production of glycosylated EGF27 for MS analysis was performed as previously described in Takeuchi *et al.* [26]. Briefly, in a total volume of 30 µL, 10.5 µg of EGF27 acceptor substrate was mixed with 750 ng of POFUT1 and 1.5 µg of LFNG with 200 µM each of UDP-GlcNAc and GDP-fucose donors. The reaction was buffered with 50 mM HEPES at pH 6.8 with 10 mM $MnCl_2$. The reaction was incubated for 6 h in a water bath at 37 °C at which point a further 750 ng of POFUT1 and 1.5 µg of LFNG were added and incubated for 6 h. A final addition of 1.5 µg of POFUT1 and 3 µg of LFNG was added, followed by incubation at 37 °C overnight. The assay was stored at -20 °C awaiting MS analysis. The *in vitro* radioactive glycosylation of EGF27 with either GDP-[2-^{3}H]-fucose or UDP-[2-^{3}H(N)]GlcNAc donor (American Radiolabeled Chemicals) was performed as described previously in Luther *et al.* and Rampal *et al.* [36, 37]. The fucosylation reaction contained 5 µg of EGF27, 1.5 µg of POFUT1 and 100 nCi of GDP-[2-^{3}H]-fucose. The GlcNAcylation reaction contained 5 µg of EGF27, 1.5 µg of POFUT1, 3 µg of LFNG, 200 µM GDP-fucose, and 45 nCi of UDP-[2-^{3}H(N)]GlcNAc. The reactions were incubated overnight at 37 °C, diluted with 900 µL of 50 mM EDTA, and stored at -20 °C until purification and analysis. The products were purified over a C18 cartridge (Agilent), washed with 9 mL of H_2O, eluted in 1 mL of 80% ethanol, mixed with 4 mL of ScintiSafe Plus 50% liquid scintillation cocktail (Fisher Scientific), vortexed, and left in the dark

overnight. The samples were scintillation counted for 2 minutes on a Perkin Elmer TriCarb 2910 TR Liquid Scintillation Analyzer.

MS data acquisition for EGF27 was performed as previously described in Takeuchi *et al.* [38]. Production of extracted ion chromatograms (EIC) was performed using Xcalibur QualBrowser (ThermoFisher Scientific). All observed peaks from the isotopic distribution of the +7 *m/z* peak are included in the EIC and the peak values include ±10 ppm error.

3. Results and discussion

3.1. Expression, purification, and cleavage of the MBP-EGF27 fusion protein

The MBP-EGF27 fusion protein was expressed from pMAL-c5x-TEV-EGF27 in Rosetta 2(DE3) *E. coli* cells and purified using Ni-IMAC. Fig. 2 shows the SDS-PAGE analysis of the expression, purification, and TEV protease cleavage of MBP-EGF27. Samples from the culture before and after induction with IPTG (Fig. 2, lanes 1 and 2) show the appearance of a dark band at 50 kDa, which corresponds to the overexpressed MBP-EGF27 fusion protein. The protein is highly soluble as shown by the lysate sample (Fig. 2, lane 3). Ni-IMAC purification yielded highly pure protein with a typical yield of 20 mg per L culture. The elution of MBP-EGF27 is shown in Fig. 2, lane 5 (band i). Samples of MBP-EGF27 before and after TEV protease digestion are shown in Fig. 2, lanes 6 and 7, respectively. These samples show nearly complete digestion of the MBP-EGF27 protein with TEV protease (band iii) to produce MBP (band ii) and cleaved EGF27 (band iv).

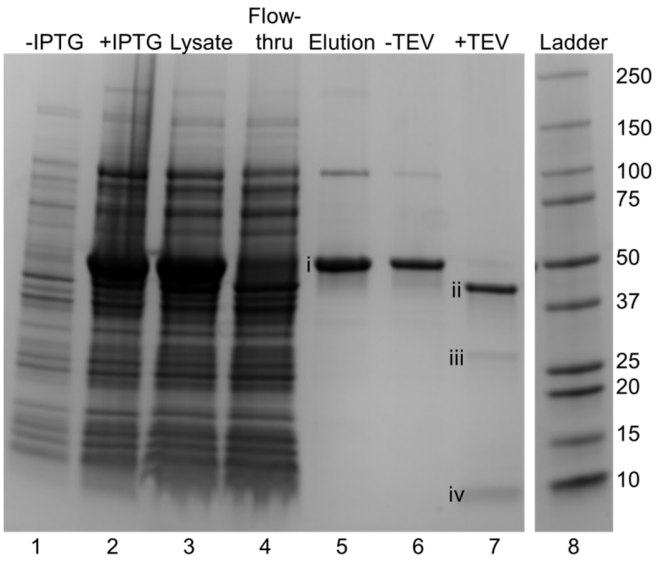


Fig. 2. SDS-PAGE analysis of the expression, purification, and TEV protease digestion of MBP-EGF27. Lane numbers are labeled along the bottom. Lanes 1 and 2 are the whole cell lysate before and after induction with IPTG showing a dark band near 50 kDa in lane 2 corresponding to MBP-EGF27. Purification of MBP-EGF27 using Ni-IMAC is shown with samples of the cleared lysate (lane 3), the lysate flow-through from the Ni-IMAC resin (lane 4), and the elution from the Ni-IMAC resin with imidazole buffer (lane 5). In lane 5, the MBP-EGF27 band is labeled (i). Lanes 6 and 7 are the purified MBP-EGF27 protein before and after digestion with TEV protease. Lane 8 is Bio-Rad Precision Plus protein standard with the molecular weights listed to the right in kDa. The migration position of MBP-EGF27 (49.8 kDa), MBP (43.8 kDa), TEV (28.1 kDa), and EGF27 (6.0 kDa) are indicated with roman numerals, i – iv, respectively. The gel images originated from the same gel; the image was split to remove irrelevant lanes for clarity. The full gel image is included as Fig. S1.

3.2. Separation and analysis of cleaved EGF27

EGF27 can form different conformational isomers based on the presence and arrangement of the three disulfide bonds. The native disulfide bond arrangement of EGF repeats is C^1–C^3, C^2–C^4, and C^5–C^6. For the EGF27 domain, the disulfide bonds should be C1025-C1036, C1030-C1045, and C1047-C1056. When an EGF repeat is expressed in *E. coli*, numerous isomers are possible especially considering different numbers and arrangements of intramolecular disulfide bonds and oligomers formed through intermolecular disulfide bonds.

The EGF27 isomers (and MBP) were separated using reverse-phase HPLC, and the proteins were identified using MALDI MS and SDS-PAGE analyses.

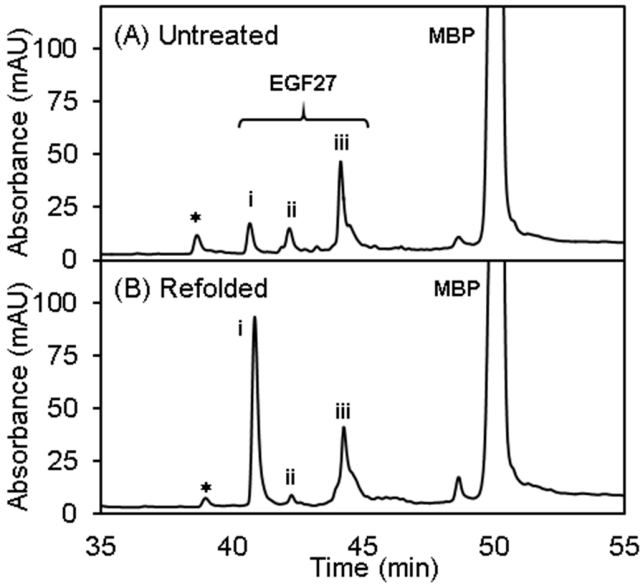


Fig. 3. HPLC separation and analysis of cleaved EGF27. Mixtures of EGF27 and MBP were separated using reverse-phase HPLC. (A) The chromatogram of the untreated sample shows 5 major peaks. The peaks labeled i, ii, and iii were confirmed to contain EGF27 based on MALDI MS analysis. The peak labeled with an asterisk is unknown because it had a lower *m/z* than was expected for EGF27. The large peak at 50.5 min was confirmed to be MBP by SDS-PAGE analysis (Fig. S1, lane 8). (B) When the MBP-EGF27 fusion protein is reduced and refolded, the cleaved EGF27 and MBP mixture shows the same 5 peaks as the untreated sample but with different relative intensities for peaks i-iii. The refolded sample is enriched in the EGF27 isomer corresponding to peak i.

Fig. 3A shows the HPLC chromatogram of the EGF27 and MBP mixture before redox refolding. The presence of EGF27 was confirmed in the peaks labeled i, ii, and iii by MALDI MS analysis (Fig. 4). The resolution of the MALDI mass spectra is not high enough to confidently identify the number of disulfide bonds in each isomer, but the *m/z* values at the peak maxima fall within the range of EGF27 with 2 or 3 disulfide bonds (*m/z* 6048.6 or 6046.6, respectively). Peaks ii and iii have additional higher molecular weight components at *m/z* of 6598.0 (+551.3) and 6353.6 (+306.2). It is not clear whether these proteins are adducts,

modified proteins, or protease fragments. One possibility for the peak at *m/z* 6353.6 is the addition of glutathione (+305) by the activity of an *E. coli* glutathione S-transferase. The peak with a *m/z* 6598.0 may be a product of endogenous protease digestion within the TEV recognition sequence leaving extra amino acids at the *N*-terminus (LYFQ-) with a calculated *m/z* of 6598.8. The HPLC peak at 37.5 min (labeled with an asterisk in Fig. 3A) gave a MALDI MS signal at a *m/z* of 4910.8 (Fig. S2); the peak is likely a truncated form of EGF27 because the *m/z* corresponds with EGF27 minus the *C*-terminal tag (-ALEHHHHHH) with a calculated *m/z* of 4911.0. The HPLC peak at 50.5 min did not give a MALDI MS signal within the range of 4000 – 7000 Da, but SDS-PAGE analysis indicated that the peak is MBP (Fig. S1, lane 8).

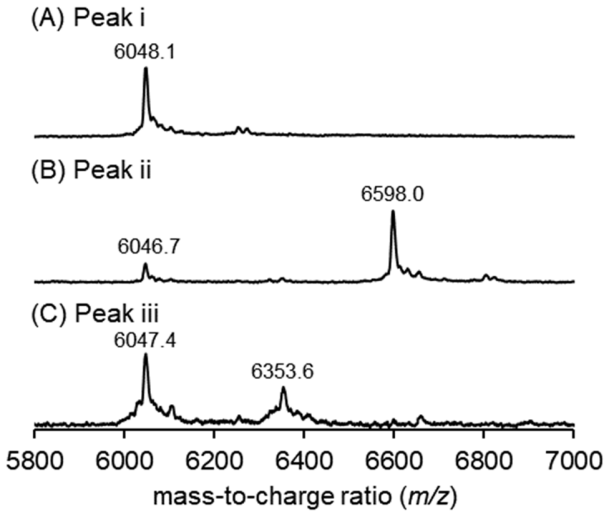


Fig. 4. MALDI-MS analysis of HPLC peaks. MALDI mass spectra of samples from the HPLC separation of the untreated sample are shown for (A) peak i, (B) peak ii, and (C) peak iii. The calculated *m/z* for EGF27 with 3 disulfide bonds is 6046.6. Peak i contains EGF27, and peaks ii and iii contain EGF27 and species at higher *m/z*.

To identify which EGF27 isomers contain disulfide bonds, the EGF27 and MBP mixture was treated with DTT and analyzed with HPLC (Fig. S3). All EGF27 isomers (i, ii, and iii) shifted in retention time after treatment with DTT, confirming at least one disulfide bond in each. The fact that the peaks do not collapse into a single peak, but shift to several peaks with

shoulders, indicates that even with the disulfide bonds dissociated, there is a mixture of species, which is consistent with the MALDI-MS analysis.

Since at least three isomers of EGF27 were produced in *E. coli* and the yield of EGF27 may be limited by the presence of oligomers of EGF27 (due to intermolecular disulfide bonds), the MBP-EGF27 fusion protein was reduced and refolded under steady redox conditions. Initial optimization of the redox buffer was performed using EGF27 expressed from pET20b-EGF27 (with no MBP fusion) [28]. The success of forming intramolecular disulfide bonds versus intermolecular disulfide bonds depended on the concentration of EGF27 such that very low concentrations (10 μg/mL) were needed [28]. The refolded EGF27 protein needed to be concentrated 300-fold for NMR analysis, which resulted in significant loss of protein. To overcome this limitation, EGF27 was reduced and refolded as part of the fusion protein, MBP-EGF27. Even at a concentration of 3.0 mg/mL, the refolded MBP-EGF27 remained soluble. The relatively large MBP likely prevents intermolecular disulfide bond formation by maintaining EGF27-solubility during refolding and sterically hindering the intermolecular interactions of EGF27 molecules.

To evaluate whether different isomers result from refolding, the fusion protein was cleaved with TEV protease and purified by HPLC. Fig. 3B shows the HPLC chromatogram of the refolded sample. Compared to Fig. 3A, peak i is enriched by 8-fold upon refolding while peak ii decreases and peak iii remains unchanged (Table S1), indicating that the isomer in peak i is the most stable one under the redox conditions used. The MBP is the same intensity in both samples (Table S1 and Fig. S4), which is expected since it does not contain cysteine residues.

3.3. Confirmation of EGF27 disulfide bonds through glycosylation studies

Li *et al.* have shown through co-crystallization of POFUT1 and EGFs that only a serine or a preferred threonine in an appropriate consensus sequence between cysteines 2 and 3 of a properly folded EGF can be modified by this enzyme [39]. LFNG can then modify the fucosylated EGF with a GlcNAc [37]. Thus, we used modification by these enzymes as a stringent test to confirm that the major EGF27 isomer (Fig. 3B, peak i) has the native disulfide bond arrangement [27, 37, 39]. EGF27 is a native substrate of POFUT1 and LFNG and is modified at T1035 with *O*-fucose by POFUT1 followed by the elongation of the glycan with a β1,3-GlcNAc by LFNG [9]. Single point assays were performed with ^{3}H-labeled donor substrates (GDP-^{3}H-fucose or UDP-^{3}H-GlcNAc) to evaluate the EGF27 isomer as a substrate for the glycosyltransferases. Fig. S5 shows robust incorporation of ^{3}H-fucose into EGF27 when incubated with POFUT1 and GDP-^{3}H-fucose, and incorporation of ^{3}H-GlcNAc when incubated with POFUT1, GDP-fucose, LFNG, and UDP-^{3}H-GlcNAc, indicating that the major EGF27 isomer has the correct disulfide bond arrangement.

To confirm the purity of the EGF27 isomer and extent of glycosylation, the same reactions were performed without radioactivity and assessed by mass spectrometry. Shown in Fig. 5 are the extracted ion chromatograms (EICs) for EGF27 untreated (Fig. 5A), EGF27 incubated with POFUT1 and GDP-fucose (Fig. 5B), and EGF27 incubated with POFUT1, GDP-fucose, LFNG, and UDP-GlcNAc (Fig. 5C). It was observed that the protein was able to be fully glycosylated in both its mono- and di-saccharide forms, further confirming that the EGF27 isoform from HPLC peak-i is in fact in its correct form with all the disulfide bonds intact. Moreover, the absence of unmodified EGF27 in the mono- and di-saccharide samples indicates that the EGF27 isomer is pure without traces of other co-eluting isomers.

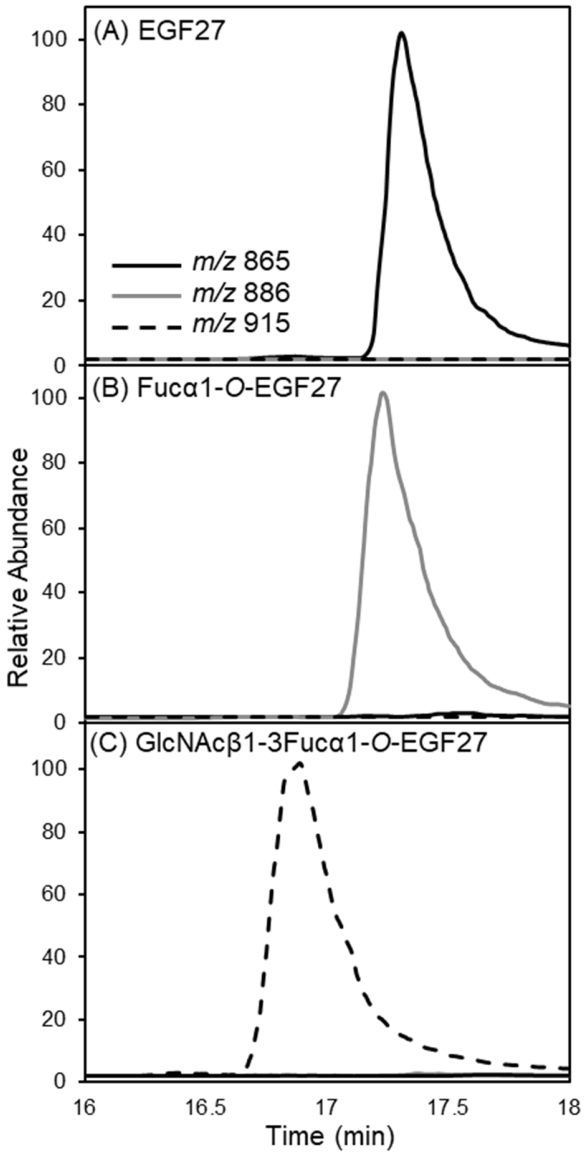


Fig. 5. Extracted ion chromatograms for the +7 charge state of (A) the unmodified form, (B) the *O*-fucose monosaccharide form, and (C) the GlcNAc-Fucose disaccharide form of EGF27. Each chromatogram includes three EICs at *m/z* of 865 (unmodified EGF27, black line), 886 (Fucα1-*O*-EGF27, gray line), and 915 (GlcNAcβ1-3Fucα1-*O*-EGF27, dashed black line).

4. Conclusions

In this study, a novel method to produce recombinant mouse NOTCH1 EGF27 with its native disulfide bond arrangement was established. Developing this method was necessary to

pursue 3-dimensional NMR structural studies of EGF27. For these studies, the yield of properly folded EGF27 from *E. coli* expression must be on the order of hundreds of micrograms per liter expression culture using minimal M9 medium. Various constructs of EGF27, including varying the *N*-terminal sequence, expressing multiple EGF repeats in tandem, and using an enhanced green fluorescent protein fusion, were tested for expression in multiple *E. coli* cell lines over many years [28, 29]. The yields were either low or recovering the small EGF27 protein proved difficult. One challenge that was overcome by using the MBP fusion was the loss of EGF27 during concentration of the protein following the redox refolding step. The MBP fusion allows for high concentrations (3.0 mg/mL) of MBP-EGF27 to be reduced and refolded without the formation of intermolecular disulfide bonds. The MBP-EGF27 protein produced in *E. coli* consists of at least three isoforms of EGF27 as determined by varying retention times in reverse phase HPLC and MALDI-MS analysis. The native disulfide arrangement of the EGF27 isoform eluting in peak-i was validated by glycosylation assays using POFUT1 and LFNG. Redox refolding increased the yield of the properly folded EGF27 by 8-fold. We anticipate that this expression and purification method will be applicable to other single and tandem EGF repeat domains as we continue to pursue structural characterization of the NOTCH1 *Abruptex* region. With the increased yield of EGF27, structural studies of EGF27 with and without its native glycosylation are being pursued.

Author contributions

A.P. produced the expression plasmid, pET28-EGF27, and performed protein expression tests. H.Z. optimized the methods to express, refold, and purify EGF27 and proposed the idea to express EGF27 as a TEV-cleavable fusion protein with MBP to improve the protein yield.

K.D.J. and J.A.G. performed protein expression tests with different cell lines to maximize yields of the MBP-EGF27 fusion protein. J.A.G. optimized the expression, refolding, cleavage, and purification of the properly folded EGF27 isomer. R.S.H. and K.B.L. designed and performed the glycosylation reactions and LC-MS analysis. J.A.G., K.B.L., R.S.H., and M.A.M. contributed to writing the manuscript.

Acknowledgements

This work was supported by funding from the National Science Foundation grant CHE-1413576 to M.A.M. and the National Institutes of Health grant GM061126 to R.S.H. J.A.G., K.D.J., H.Z., and A.P. thank the Louisiana State University Department of Chemistry for providing support through teaching assistantships. We thank Dr. Tamara Chouljenko and Dr. Ted Gauthier from LSU's AgCenter Biotechnology Laboratory for their support and advice.

Appendix A. Supplementary data

TEV protease expression and purification, SDS-PAGE analysis of MBP-EGF27 (Fig. S1), MALDI-MS analysis of an HPLC peak (Fig. S2), HPLC chromatograms of EGF27 and MBP (Fig. S3), HPLC peak areas (Table S1), HPLC chromatograms of refolded EGF27 (Fig. S4), and radioactive glycosylation assay (Fig. S5).

References

[1] R. Kopan, M.X.G. Ilagan, The canonical Notch signaling pathway: Unfolding the activation mechanism, Cell 137 (2009) 216-233. https://doi.org/10.1016/j.cell.2009.03.045.

[2] I. Rebay, R.J. Fleming, R.G. Fehon, L. Cherbas, P. Cherbas, S. Artavanistsakonas, Specific EGF repeats of Notch mediate interactions with Delta and Serrate: Implications for Notch as a multifunctional receptor, Cell 67 (1991) 687-699. https://doi.org/10.1016/0092-8674(91)90064-6.

[3] M.R. Kelley, S. Kidd, W.A. Deutsch, M.W. Young, Mutations altering the structure of epidermal growth-factor like coding sequences at the Drosophila Notch locus, Cell 51 (1987) 539-548. https://doi.org/10.1016/0092-8674(87)90123-1.

[4] K. Bruckner, L. Perez, H. Clausen, S. Cohen, Glycosyltransferase activity of Fringe modulates Notch-Delta interactions, Nature 406 (2000) 411-415. https://doi.org/10.1038/35019075.

[5] J.F. de Celis, S.J. Bray, The *Abruptex* domain of Notch regulates negative interactions between Notch, its ligands and Fringe, Development 127 (2000) 1291-1302.

[6] V.C. Luca, B.C. Kim, C. Ge, S. Kakuda, W. Di, M. Roein-Peikar, R.S. Haltiwanger, C. Zhu, T. Ha, K.C. Garcia, Notch-Jagged complex structure implicates a catch bond in tuning ligand sensitivity, Science 355 (2017) 1320-1324. https://doi.org/10.1126/science.aaf9739.

[7] V.C. Luca, K.M. Jude, N.W. Pierce, M.V. Nachury, S. Fischer, K.C. Garcia, Structural basis for Notch1 engagement of Delta-like 4, Science 347 (2015) 847-853. https://doi.org/10.1126/science.1261093.

[8] N.A. Rana, R.S. Haltiwanger, Fringe benefits: Functional and structural impacts of *O*-glycosylation on the extracellular domain of Notch receptors, Curr. Opin. Struct. Biol. 21 (2011) 583-589. https://doi.org/10.1016/j.sbi.2011.08.008.

[9] S. Kakuda, R.S. Haltiwanger, Deciphering the Fringe-mediated Notch code: Identification of activating and inhibiting sites allowing discrimination between ligands, Dev. Cell 40 (2017) 193-201. https://doi.org/10.1016/j.devcel.2016.12.013.

[10] P.A. Handford, B. Korona, R. Suckling, C. Redfield, S.M. Lea, Structural insights into Notch receptor-ligand interactions, Adv. Exp. Med. Biol. 1066 (2018) 33-46. https://doi.org/10.1007/978-3-319-89512-3_2.

[11] S. Hambleton, N.V. Valeyev, A. Muranyi, V. Knott, J.M. Werner, A.J. McMichael, P.A. Handford, A.K. Downing, Structural and functional properties of the human Notch-1 ligand binding region, Structure 12 (2004) 2173-2183. https://doi.org/10.1016/j.str.2004.09.012.

[12] E.J. Boswell, N.D. Kurniawan, A.K. Downing, Calcium-binding EGF-like domains, Encyclopedia of Inorganic and Bioinorganic Chemistry, John Wiley & Sons, Ltd, 2011.

[13] J. Cordle, C. Redfield, M. Stacey, P.A. van der Merwe, A.C. Willis, B.R. Champion, S. Hambleton, P.A. Handford, Localization of the Delta-like-1-binding site in human Notch-1 and its modulation by calcium affinity, Journal of Biological Chemistry 283 (2008) 11785-11793.

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

Implications for the Marfan syndrome and other genetic disorders, Cell 85 (1996) 597-605. https://doi.org/http://dx.doi.org/10.1016/S0092-8674(00)81259-3.

[15] J.A. Davis, P.A. Handford, C. Redfield, The N1317H substitution associated with Leber congenital amaurosis results in impaired interdomain packing in human CRB1 epidermal growth factor-like (EGF) domains, J Biol Chem 282 (2007) 28807-28814. https://doi.org/10.1074/jbc.M704015200.

[16] W.D. Morgan, B. Birdsall, T.A. Frenkiel, M.G. Gradwell, P.A. Burghaus, S.E.H. Syed, C. Uthaipibull, A.A. Holder, J. Feeney, Solution structure of an EGF module pair from the Plasmodium falciparum merozoite surface protein 1, Journal of Molecular Biology 289 (1999) 113-122.

[17] D.J. Moloney, L.H. Shair, F.M. Lu, J. Xia, R. Locke, K.L. Matta, R.S. Haltiwanger, Mammalian Notch1 is modified with two unusual forms of *O*-linked glycosylation found on epidermal growth factor-like modules, J. Biol. Chem. 275 (2000) 9604-9611. https://doi.org/10.1074/jbc.275.13.9604.

[18] A. Matsuura, M. Ito, Y. Sakaidani, T. Kondo, K. Murakami, K. Furukawa, D. Nadano, T. Matsuda, T. Okajima, *O*-linked *N*-acetylglucosamine is present on the extracellular domain of Notch receptors, J. Biol. Chem. 283 (2008) 35486-35495. https://doi.org/10.1074/jbc.M806202200.

[19] H. Takeuchi, M. Schneider, D.B. Williamson, A. Ito, M. Takeuchi, P.A. Handford, R.S. Haltiwanger, Two novel protein *O*-glucosyltransferases that modify sites distinct from POGLUT1 and affect Notch trafficking and signaling, Proc. Natl. Acad. Sci. USA 115 (2018) E8395-E8402. https://doi.org/10.1073/pnas.1804005115.

[20] Y. Urata, H. Takeuchi, Effects of Notch glycosylation on health and diseases, Dev. Growth Diff. 62 (2020) 35-48. https://doi.org/10.1111/dgd.12643.

[21] S. Varshney, P. Stanley, Multiple roles for *O*-glycans in Notch signalling, FEBS Lett. 592 (2018) 3819-3834. https://doi.org/10.1002/1873-3468.13251.

[22] B.C. Holdener, R.S. Haltiwanger, Protein *O*-fucosylation: Structure and function, Curr. Opin. Struct. Biol. 56 (2019) 78-86. https://doi.org/10.1016/j.sbi.2018.12.005.

[23] A. Pandey, B.M. Harvey, M.F. Lopez, A. Ito, R.S. Haltiwanger, H. Jafar-Nejad, Glycosylation of specific Notch EGF repeats by O-Fut1 and Fringe regulates Notch signaling in *Drosophila*, Cell Rep. 29 (2019) 2054-2066 e2056. https://doi.org/10.1016/j.celrep.2019.10.027.

[24] R. Rampal, J.F. Arboleda-Velasquez, A. Nita-Lazar, K.S. Kosik, R.S. Haltiwanger, Highly conserved *O*-fucose sites have distinct effects on Notch1 function, J. Biol. Chem. 280 (2005) 32133-32140. https://doi.org/10.1074/jbc.M506104200.

[25] S. Varshney, H.X. Wei, F. Batista, M. Nauman, S. Sundaram, K. Siminovitch, A. Tanwar, P. Stanley, A modifier in the 129S2/SvPasCrl genome is responsible for the viability of *Notch1*[12f/12f] mice, BMC Dev. Biol. 19 (2019) 1-11. https://doi.org/10.1186/s12861-019-0199-3.

[26] H. Takeuchi, J. Kantharia, M.K. Sethi, H. Bakker, R.S. Haltiwanger, Site-specific *O*-glucosylation of the epidermal growth factor-like (EGF) repeats of Notch: Efficiency of glycosylation is affected by proper folding and amino acid sequence of individual EGF repeats, J. Biol. Chem. 287 (2012) 33934-33944. https://doi.org/10.1074/jbc.M112.401315.

[27] Y. Wang, M.W. Spellman, Purification and characterization of a GDP-fucose:polypeptide fucosyltransferase from Chinese hamster ovary cells, J. Biol. Chem. 273 (1998) 8112-8118. https://doi.org/10.1074/jbc.273.14.8112.

[28] H. Zhong, M.A. Macnaughtan, Expression, purification and characterization of EGF27 from the Notch1 *Abruptex* region, LSU Master's Theses 4352 (2017).

[29] A. Paudyal, M.A. Macnaughtan, Expression and characterization of mouse Notch1 *Abruptex* EGF repeats in *E. coli*, LSU Master's Theses 825 (2015).

[30] C.Y. Seiler, J.G. Park, A. Sharma, P. Hunter, P. Surapaneni, C. Sedillo, J. Field, R. Algar, A. Price, J. Steel, A. Throop, M. Fiacco, J. LaBaer, DNASU plasmid and PSI:Biology-Materials repositories: Resources to accelerate biological research, Nucleic Acids Res. 42 (2014) D1253-D1260. https://doi.org/10.1093/nar/gkt1060.

[31] P.G. Blommel, B.G. Fox, A combined approach to improving large-scale production of tobacco etch virus protease, Protein Expres. Purif. 55 (2007) 53-68. https://doi.org/10.1016/j.pep.2007.04.013.

[32] J. Marley, M. Lu, C. Bracken, A method for efficient isotopic labeling of recombinant proteins, J. Biomol. NMR 20 (2001) 71-75. https://doi.org/10.1023/a:1011254402785.

[33] F.W. Studier, Protein production by auto-induction in high-density shaking cultures, Protein Expres. Purif. 41 (2005) 207-234. https://doi.org/10.1016/j.pep.2005.01.016.

[34] D. Lu, Z. Liu, Dynamic redox environment-intensified disulfide bond shuffling for protein refolding *in vitro*: Molecular simulation and experimental validation, J. Phys. Chem. B 112 (2008) 15127-15133. https://doi.org/10.1021/jp804649g.

[35] Y. Luo, R.S. Haltiwanger, *O*-Fucosylation of Notch occurs in the endoplasmic reticulum, J. Biol. Chem. 280 (2005) 11289-11294. https://doi.org/10.1074/jbc.M414574200.

[36] K.B. Luther, H. Schindelin, R.S. Haltiwanger, Structural and mechanistic insights into Lunatic Fringe from a kinetic analysis of enzyme mutants, J. Biol. Chem. 284 (2009) 3294-3305. https://doi.org/10.1074/jbc.M805502200.

[37] R. Rampal, A.S.Y. Li, D.J. Moloney, S.A. Georgiou, K.B. Luther, A. Nita-Lazar, R.S. Haltiwanger, Lunatic fringe, Manic fringe, and Radical fringe recognize similar specificity determinants in *O*-fucosylated epidermal growth factor-like repeats, J. Biol. Chem. 280 (2005) 42454-42463. https://doi.org/10.1074/jbc.M509552200.

[38] H. Takeuchi, H. Yu, H. Hao, M. Takeuchi, A. Ito, H. Li, R.S. Haltiwanger, *O*-Glycosylation modulates the stability of epidermal growth factor-like repeats and thereby regulates Notch trafficking, J. Biol. Chem. 292 (2017) 15964-15973. https://doi.org/10.1074/jbc.M117.800102.

[39] Z. Li, K. Han, J.E. Pak, M. Satkunarajah, D. Zhou, J.M. Rini, Recognition of EGF-like domains by the Notch-modifying *O*-fucosyltransferase POFUT1, Nat. Chem. Biol. 13 (2017) 757-763. https://doi.org/10.1038/nchembio.2381.