



The Synthesis of Protein Polymer Conjugates using the Human Regulatory Protein Galectin-3

Ling Lin, Amanda M. Pritzlaff, Hailie-Ann C. Lower, & Daniel A. Savin

University of Florida

Faculty mentor: Daniel A. Savin, Department of Chemistry

Abstract

Galectin-3 (gal3) is a human lectin protein that is known to interact with extracellular matrix proteins by regulating functions in both healthy and cancerous cells. The goal of this project is to conjugate polymers to gal3 to better study and control its functions *in vitro*. We hypothesize that a covalently attached polymer will sterically modulate gal3 function. In the project, we created two protein variants with polymer-reactive handles. The first construct is similar to wild-type gal3 with a cysteine in place of the 6th serine (S6C) which was created by site-directed mutagenesis (SDM). Maleimide-terminated polyethylene oxide (PEO, 5000 g/mol) was then attached to this mutant via thiol-Michael addition at the cysteine site. Attachment of polymer to the unstructured N-terminal domain (NTD) may increase the binding of the protein by sterically pulling the NTD away from the carbohydrate recognition domain (CRD). In addition, the NTD, which is implicated in undesired self-association, was removed for the second construct. The gal3 CRD only construct is shown to have a higher solubility in solution and an increased ligand-binding affinity. Ultimately, the two unique constructs will help us understand the structural role of the NTD in gal3 ligand-binding and self-association.

Keywords: galectin-3, Maleimide-terminated polyethylene oxide, cysteine thiol-Michael addition, polymer bioconjugate

Introduction

Galectin-3 (Gal3) is a structurally unique protein which is implicated in many important biological pathways such as growth and differentiation, inflammation, and cancer metastasis (Li et al., 2014). Gal3 is frequently found in the extracellular matrix (ECM), and the protein has an important function in healthy cells and metastatic cancerous cells. For instance, in many human carcinomas, an increased expression of gal-3 correlates with progressive tumor stages (Nangia-Makker et al., 2001). Understanding the structural role of the carbohydrate-recognition domain (CRD) and N-terminal domain (NTD) of gal3 can provide insights into gal3 modulation in the human body.

Gal3 is a member of the lectin family, of which all members have a conserved CRD. Uniquely, only gal3 has an unstructured NTD and the CRD. This removal of NTD has also been

shown to prevent gal3 oligomerization *in vitro* (Ochieng et al., 1998). The role of the NTD for biological function is still mysterious due to the inability to isolate oligomerized gal3 from living cells. For example, gal3 is monomeric up to high concentrations (100 μM), but can aggregate with ligands and agglutinate cells at much lower concentrations, which is promoted by its N-terminal non-CRD region *via* an oligomerization mechanism that is not yet completely understood (Johannes et al., 2018). Thus, it is hypothesized that NTD is involved in the self-association process. Interestingly, metalloproteinase, particularly MMP-2 and MMP-9, have shown to be capable of efficiently cleaving gal3 to generate a CRD and NTD fragments in the ECM (Ochieng et al., 1998). This suggests that cleavage of the NTD can occur in the body enzymatically. Therefore, the hypothesis that NTD is involved in the self-association process is tested by cleaving the NTD sequence out of the gal3 DNA sequence.

Herein we report polymer conjugates produced by covalently attaching polymer to two gal3 constructs. The gal3S6C construct is to the wide-type full length gal3, but with an inserted cysteine as a polymer-reactive handle. The CRD only construct, on the other hand, will have the NTD removed and a cysteine polymer-reactive handle. The cysteine is inserted using site-direct mutagenesis (SDM), a method that provides a site-specific location of the nucleophile for the conjugation reaction to occur. For instance, the thiol in the cysteine is the polymer-reactive handle used to covalently attach polymer to gal3 protein. Figure 1 below shows the conjugation of maleimide-terminated polyethylene oxide (PEO, 5000 g/mol) covalently attached to the sulfur position of cysteine via thiol-Michael addition.

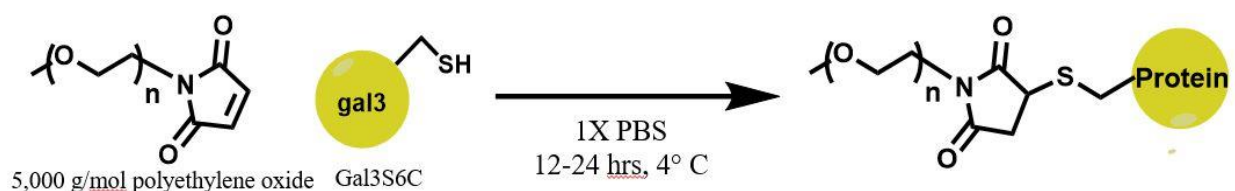


Figure 1. Conjugation scheme of PEO and gal3 construct at the sulfur position of cysteine via thiol-Michael addition

Successful completion of this work will benefit the polymer community by providing an understanding of how proteins with disordered regions behave in a bioconjugate.

Methods

Site Directed Mutagenesis:

SDM was used to create a substitution mutation in the full length and CRD only constructs of gal3 where one of the amino acids in the protein sequence was replaced by cysteine. SDM was also used to create an insertion mutation where an NcoI restriction site was added to the full-length gal3 DNA sequence between the N-terminal coding region and the CRD coding region. The initial plasmid DNA to which all mutations were made was a pET-21d(+) vector with the full length, wild type human gal3 gene between NcoI and XhoI restriction sites. This plasmid was a generous gift from The Hudalla Lab in the University of Florida Biomedical Engineering Department.

First, primers were designed to contain the desired mutation such as a substitution or insertion. Then, the primers were used in polymerase chain reactions (PCR) to amplify DNA with the desired mutation. The PCR products were analyzed with a 1% agarose gel containing ethidium bromide to confirm amplification and Sanger sequencing was used to confirm the mutation.

Gal3S6C denotes a serine to cysteine mutation at the 6th amino acid position, part of the NTD. The sequence of the designed overlap primers are shown below:

Gal3S6C Forward Primer 5'-AATTTTTTGCCTCCATGATGCGTTATCTGG-3'
 Gal3S6C Reverse Primer 5'-CATGGAGGCAAAAAT TGTCTGCGACCATGG T-3'

A serine to cysteine substitution was applied in the sequence using the underlined nucleotides "TGC" which codes for a cysteine.

The gal3 CRD only construct is a gal3 protein without the NTD. A second NcoI restriction site was inserted in the pET-21d(+)-gal3 DNA so the NTD-coding sequence would be removed by restriction digest with NcoI. The second NcoI site was inserted between the DNA regions coding for the NTD and CRD. This resulted in the NTD region flanked with NcoI sites. Non-overlap primers were initially designed for this mutation and used in PCR but it was unsuccessful. Thus, overlap primers then designed and used for PCR. The sequence of the designed overlap primers are shown below:

Overlap NcoI Insertion Forward Primer 5'-CCCCTCCATGGCTGGGCCACTGATT-3'
 Overlap NcoI Insertion Reverse Primer 5'-CCAGCCATGGAGGGGCGCCATAGGG-3'

Insertion occurred where the sequence of NcoI site was introduced to the primer as shown in the nucleotides underlined above. “CCATGG” codes for an NcoI restriction site. After three cycles of PCR, the NcoI sequence is predicted to be incorporated into the DNA sequence of gal3. A total of 30 PCR cycles were completed to further amplify the modified gal3 DNA sequence. The PCR sample was analyzed with agarose gel electrophoresis and the brightest band near the expected kilobase (kb) reference on the ladder was then excised from the gel to extract the DNA. The amplified DNA was extracted using the Wizard® SV Gel and PCR Clean-Up System. The extracted DNA then goes through a series of reactions to have the sequence for the NTD removed. (Refer to “Series of Reaction for the Removal of NTD”)

After the CRD only construct was produced, SDM was used to substitute threonine at the 134th position with cysteine. The sequences in the designed overlap primers are shown below:

Gal3T134C CRD only construct Forward Primer 5’-
AGACCTCTGCAGTGCTTCATATAACC-3’

Gal3T134C CRD only construct Reverse Primer 5’- AGCACTGCAGAGGTCTATGTCAC-
3’

A serine to cysteine substitution was applied to the sequence using the underlined nucleotides “TGC” which codes for a cysteine.

Polymerase Chain Reaction

PCR was used to amplify DNA with targeted mutations inserted or substituted into the sequence. Table 1 shows the reagents used in a standard 50µL PCR and Table 2 shows the PCR cycling instruction.

Reagents	Control (µL)	PCR (µL)
H ₂ O	30.5	28.5
5x Phusion HF buffer	10	10
dNTPs (deoxyribonucleotide triphosphate)	4	4
Reverse Primer	2.5	2.5
Forward Primer	2.5	2.5
Template DNA	0	2
Phusion Hot Start II DNA Polymerase	0.5	0.5
<i>Note.</i> Thermo Scientific™ Phusion™ Hot Start II High-Fidelity DNA Polymerase was used. PCR was set up by following the guidelines from the manual provided by Thermo Scientific™. The reagents are added in the order shown in the table.		

Table 2. PCR cycling instructions

Stages	Time	Temperature (°C)	Cycle
1. Initial denaturation	2 minutes	95	1
2. Denaturation	20 seconds	95	30
Annealing	20 seconds	*	
Extension	2 minutes	72	
3. Final Extension	30 minutes	72	1
Completion	∞	4	-

Note. * The annealing temperature varies base on the melting temperature of the primers.

In every PCR setup, one control reaction was prepared to monitor contamination in the laboratory and ensure the proper functioning of the reagents. In addition, all the reagents were stored on ice during the PCR setup. All PCR results were analyzed by running a 1% agarose gel at 100 volts for 70 minutes.

Series of Reactions for the Removal of NTD

After the addition of a second NcoI site, the sequence for the NTD can be removed by NcoI digestion in order to create the CRD-only construct. As previously mentioned, the PCR product was run on an agarose gel and the amplified DNA was extracted using the Wizard® SV Gel and PCR Clean-Up System. The DNA extracted was first digested with DpnI, an enzyme to remove cell-derived template plasmid from the PCR products. This digestion was performed for all PCR products. The reagents for the DpnI digestion include 1 µL of DpnI enzyme, 5 µL of CutSmart buffer and 50 µL of PCR product. The reaction was incubated at 37°C for 80 minutes. After digestion, the DpnI enzyme was heat inactivated at 80°C for 20 minutes. The DpnI digested product was stored at 4°C.

The DpnI digested product was then digested with NcoI. The reagents for the digestion includes the DNA sample, autoclaved water, CutSmart Buffer and NcoI (buffer and enzyme from New England BioLabs). DNA sample (1 µg) was added to a 50 µL digestion reaction. The reaction was allowed to run for one hour at 37°C. After digestion, the NcoI enzyme was heat inactivated at 80°C for 20 minutes. The NcoI digested product was stored at 4°C.

Next, the NcoI digested product was run on a 1% agarose gel to separate the two DNA fragments, the CRD gene in the plasmid and the NTD coding region flanked with NcoI sites. The

larger CRD-only DNA fragment was excised from the gel with a scalpel under UV light and the DNA was extracted using the Wizard® SV Gel and PCR Clean-Up System. Then the CRD-only DNA fragment was ligated using DNA Ligase. “Ligation Protocol with T4 DNA Ligase (M0202)” from New England BioLabs was followed for such ligation reaction.⁷ The reagents include T4 DNA Ligase Buffer (10X), linear CRD-region containing plasmid, nuclease free water and T4 DNA Ligase. For a 20µL ligation reaction, 50 ng of vector DNA, 1 µL of T4-ligase, 2 µL of 10X T4 ligase buffer, and 11.3 µL of nuclease free water was used. The reaction was incubated at 16°C for 2 hours. After that, the DNA Ligase was heat inactivated at room temperature for 2 hours. The ligated product was stored at 4°C.

Overexpression and Purification

A pET-21d(+) plasmid with a gene coding for gal3S6C or gal3-CRD-T134C between the restriction sites XhoI and NcoI was transformed into OrigamiB(DE3) E. coli cells and overexpressed. The transformation protocol from the Novagen Competent Cells was adapted. Transformation was done by heat shocking chemically competent cells with approximately 20 ng of plasmid DNA in a water bath set to 42°C. The cells were then spread on a lysogeny broth (LB) agar plate containing ampicillin and kanamycin and incubated overnight at 37 °C. A colony from the overnight culture was used to inoculate and grow the cells in a 5 mL volume of LB medium with ampicillin and kanamycin (0.1 mg/mL each). This culture was incubated in a shaker at 37°C until the optical density (OD) at 600 nm (OD₆₀₀) was between the values of 0.5 to 1 (12 hours). Next, the 5mL culture was transferred to a larger 500 mL volume of LB with the same concentration of ampicillin and kanamycin and allowed to grow to OD₆₀₀~0.5 (4-6 hours). The cells were then induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a concentration of 3 mM. After 18 hours of expression at 32°C, the cells were centrifuged to collect the pellets, and the pellets were lysed with a microfluidizer to collect the soluble protein construct. The lysate was centrifuged for 20 minutes at 5,000 rpm and 4°C and the supernatant was used for further purification steps. The protein was purified with cobalt affinity chromatography. The pET-21d(+) includes a coding region for a C-terminal hexahistidine-tag which binds strongly to the cobalt-resin in the column, and the protein was eluted at different imidazole concentrations, 5 mM and 25 mM in 1X phosphate buffered saline (PBS). The first step in the protein purification protocol was to allow one flow-through of filtered supernatant where proteins bind to the column. The column was then washed twice with 50 mL of 1X PBS

followed by one 50 mL wash of 1X PBS with 500 mM sodium chloride. After the washes, 3 column volume (10 mL) of 1X PBS with 5 mM imidazole and 2 column volume of 1X PBS with 25 mM imidazole was collected. The purification products were analyzed in a 10% acrylamide SDS-PAGE gel. The gel was run at 85 Volts until the bands had reached the end of the stacking layer of the gel. After that, the voltage was increased to 130 Volts until all the bands reach the end of the gel.

Conjugation Scheme

The purified protein construct was covalently attached to maleimide-terminated polyethylene oxide at the sulfur position of cysteine via thio-Michael addition (See Figure 1 for more information). Such conjugation was performed in 1x PBS buffer solution at 4° C for 24 hours, and sample aliquots were collected at several time points for analysis of the reaction. The gal3S6C bioconjugate was first analyzed in an SDS-PAGE gel to compare the size difference before and after conjugation. Secondly, gal3S6C bioconjugate was further analyzed through a Matrix Assisted Laser Desorption/Ionization - Time of Flight Mass Spectrometry (MALDI-TOF) performed in linear-ion mode on a Bruker Autoflex instrument.

Results and Discussion

Polymer bioconjugate materials will help to define the structure-function relationship of gal3. This will aid in the understanding of the biological roles of gal3 such as its part in cancer progression and metastasis.

Galectin-3 S6C (Gal3S6C) Construct

Fractions of gal3S6C were collected using different imidazole concentrations as eluent, and the fractions were then analyzed by an SDS PAGE gel. The gel image is shown in Figure 2. Lane 7 in Figure 1 includes a band with the highest intensity which indicates that the first wash with 25 mM imidazole elutes the highest concentration of gal3 construct. In addition, Figure 1 also shows that the size of the Gal3S6C is around 25 kDa which is consistent with the theoretical monoisotopic molecular weight of 27.2 kDa as calculated with Swiss Institute of Bioinformatics ExPASy Tool (Gasteiger et al., 2015).

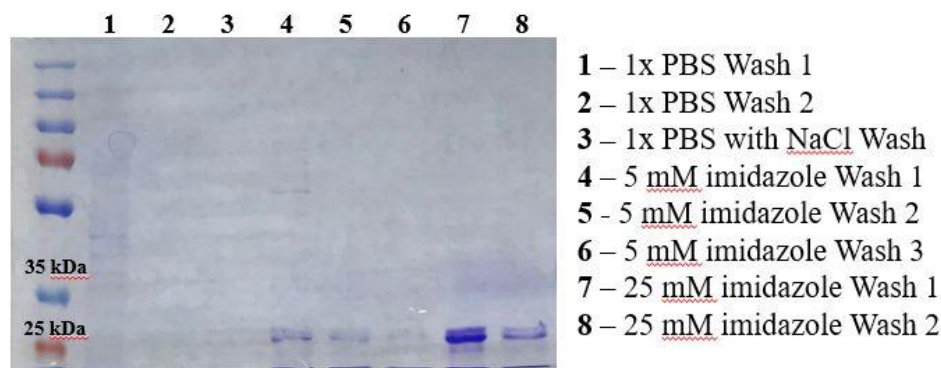


Figure 2. SDS PAGE gel image of Gal3S6C construct purification with cobalt affinity chromatography. Each lane represents different buffers conditions used to wash the column. Lane 1 and 2 represent an aliquot taken from 50 mL of 1x Phosphate buffered saline (PBS) eluate. Lane 3 represents an aliquot taken from 50 mL of 1x PBS with Sodium Chloride (NaCl) eluate. Lane 4 through 6 represent an aliquot taken from 15 mL of 5 mM imidazole eluate. Lane 7 and 8 represents an aliquot taken from 15 mL of 25 mM imidazole eluate.

Polymer conjugation can be used as a tool to understand the role of the NTD in gal3 function. Gal3S6C construct was generated and conjugated to PEO polymer. From the conjugation scheme presented in Figure 1, the cysteine mutations were placed in areas considered to be solvent accessible based on proximity to the N-terminus or the crystal structure of the protein. Conjugation of gal3S6C to a polymer is hypothesized to increase the solubility and causes a change in the binding affinity. Attachment of polymer to Gal3S6C construct at the NTD may increase the binding of the protein by sterically pulling the NTD away from the CRD while attachment of polymer to gal3 CRD only construct is hypothesized to decrease ligand-binding affinity.

With the completion of the conjugation, two samples of conjugated product and two samples of unconjugated protein were run on an SDS PAGE gel for analysis. Figure 3 below shows the gel image comparing samples before and after conjugation. On lane 1 and 2, the size of unconjugated gal3S6C is about 25 kDa. However, the size of the conjugate shows about 10kDa increase to 35 kDa. This is evidence to suggest that the conjugation reaction was indeed successful. About 50% conversion to gal3S6C conjugate was observed.

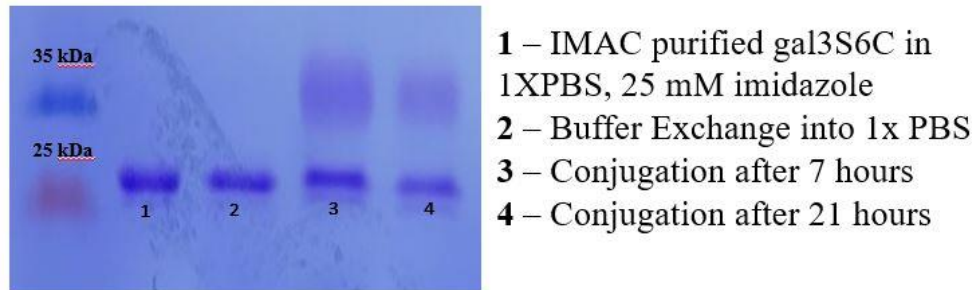


Figure 3. SDS PAGE gel for the comparison gal3S6C protein construct before and after conjugation. Lane 1 represents an aliquot taken from 15 mL of 25 mM imidazole eluate collected during gal3S6C protein purification. Lane 2 represents an aliquot taken from a 1x PBS buffer exchange of the gal3S6C protein. Lane 3 and 4 represents an aliquot taken at different time points of conjugation, 7 hour and 21 hours after conjugation respectively.

To further support that the PEO was indeed conjugated to the gal3S6C protein construct, MALDI-TOF was used to characterize the samples before and after conjugation. As shown in Figure 4, the bottom spectrum represents the protein sample before conjugation which only has one peak. However, a second peak appeared on the top spectrum that represents the protein sample after conjugation. The occurrence of the second peak with a 5.5 kDa mass shift from the original peak as well as the diminishing of the original peak indicates that the molecular weight of the protein sample has increased. Therefore, the MALDI-TOF is a compelling piece of evidence to support the formation of the gal3S6C protein conjugate.

To assess initial solution stability of the gal3S6C conjugates, a small sample of the conjugate solution in 1X PBS was allowed to warm to room temperature and some precipitation was observed. Further characterization is needed to identify the soluble portion.

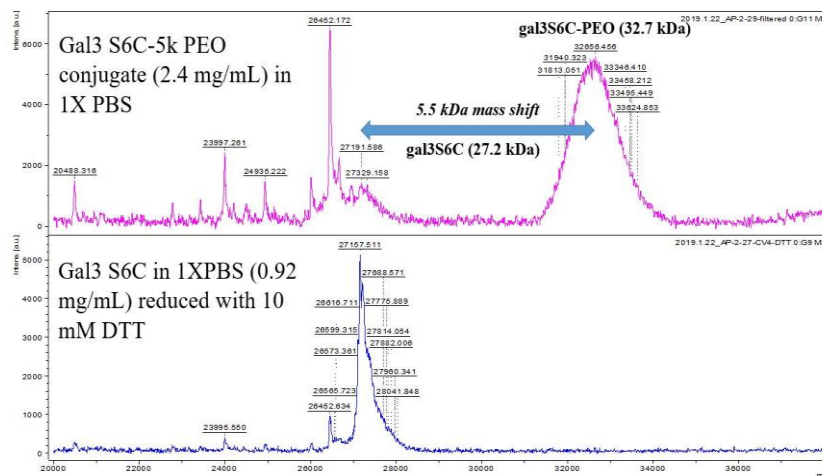


Figure 4. MALDI-TOF was used to characterize gal3S6C before and after the conjugation with 5,000 g/mol PEO.

Galectin-3 Carbohydrate Recognition Domain (CRD) only Construct

Primer was designed to include the second NcoI restriction site. The results of the PCR with different annealing temperatures were analyzed in an agarose gel (Figure 5). For lane 3 through lane 6, four bands were present. While the brightest band includes the amplified DNA product with the second NcoI site, all the other bands are PCR by-products. Note that the target PCR product (brightest bands) with the approximate size of 5kb are found higher up in the gel than the DNA template band (Lane 1, Figure 5). This would suggest a size difference between the amplified DNA and the DNA template. The apparent size difference is due to the different migration rates of linear PCR products and supercoiled circular plasmid templates (Voytas, 2001). Linear DNA travel slower than supercoiled circular plasmid so the DNA template can travel farther down the gel. In addition, since an increase in band brightness was observed from lane 1 to lane 3 through lane 6, it can conclude that the gal3 DNA sequence was amplified and the second NcoI restriction was successfully inserted.

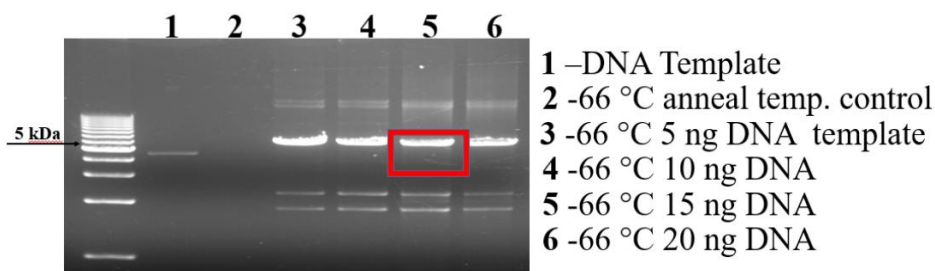


Figure 5. 1% Agarose gel image of PCR samples after the insertion of the second NcoI Restriction Site. Lane 1 represent the DNA template before PCR. Lane 2 is PCR control with does not contain any DNA. Lane 3 through 6 represents samples from PCR with different DNA template concentrations at the annealing temperature of 66°C.

The band in the red box presented in Figure 5 was cut out and the PCR amplified product was extracted using the Wizard® SV Gel and PCR Clean-Up System. Lane 5 was chosen because the DNA extracted had the highest DNA concentration and an A260/280 (absorbance maxima at 260 and 280 nm) ratio closest to 1.8 which suggested a pure DNA sample. The 5 kb band on lane 5 was chosen because it matches with the molecular weight of the CRD only fragment. The measurement of DNA concentration and A260/280 ratio was obtained with a NanoDrop Spectrophotometer. The PCR amplified product was then DpnI digested to remove template DNA and NcoI digested to cut the DNA product into 2 fragments, the CRD only fragment and NTD fragment. The NcoI digestion product was analyzed in an agarose gel and the gel image is shown in Figure 6. In order to maximize yield after gel-purification, 52 µL of the

NcoI digested DNA product was loaded. Contrary to what we predicted, only one band showed up in lane 1 of the gel. This band is the CRD only fragment with an approximate size of 5 kb. However, the smaller NTD-coding fragment was not visualized. One possibility that the NTD-coding fragment was not visible on the gel is due the small size of 0.33 kb of the fragment.

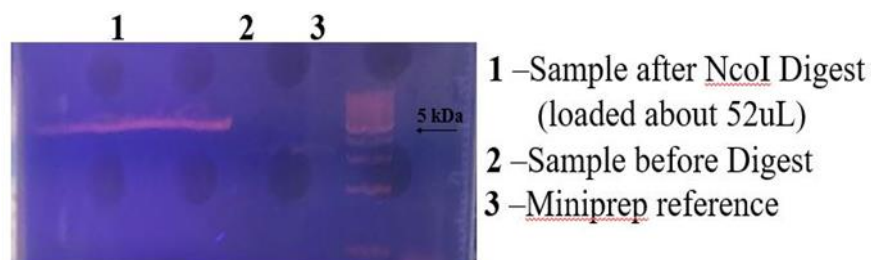


Figure 6. SDS PAGE gel image of NcoI-cut pET21-d(+)-gal3-NcoI-insert. Lane 1 represents the DNA sample of gal3 CRD only sequence with after NcoI digestion. Lane 2 represents the DNA sample of gal3 CRD only sequence before NcoI digestion. Lane 3 is a DNA sample extracted using Wizard® Plus Minipreps DNA Purification System.

The long band in lane 1 shown in Figure 6 was cut out and the NcoI digested DNA product was extracted using the Wizard® SV Gel and PCR Clean-Up System. Next, the CRD only DNA fragment was ligated using the “Ligation Protocol with T4 DNA Ligase (M0202)” from New England BioLabs (New England Biolabs, 2019). To remove the cell-derived plasmid template in the PCR amplified product, DpnI enzyme was used for the digestion reaction. Then, the DpnI digested product underwent a SDM through modified primer and PCR to substitute the threonine with cysteine at the 130 position. Figure 7 shows the PCR products at different annealing temperatures. The reason for the size difference between the PCR amplified products and the DNA template is the same as observed in Figure 5. Since an increase in band brightness was observed from lane 1 to lane 2 and lane 3, it can be concluded that the gal3 CDR only DNA sequence was amplified and SDM were successful.

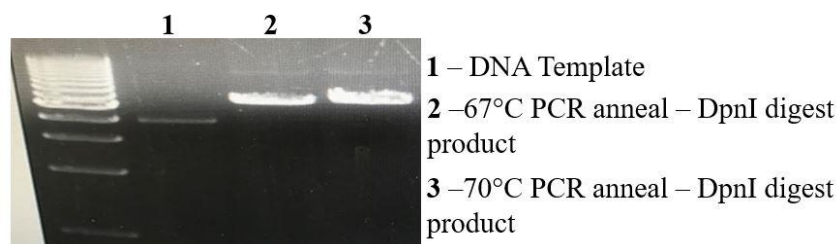


Figure 7. 1% Agarose gel image of pET-21d(+)-gal3CRD-T134C. Lane 1 represents DNA template before PCR. Lane 2 and 3 represents the PCR sample at 67°C and 70°C annealing temperature respectively.

The bands in lane 2 and 3 shown in Figure 7 were cut out and the DnpI digested PCR amplified products was extracted using the Wizard® SV Gel and PCR Clean-Up System. The sequence of the products was obtained through Sanger sequencing. Sequence alignment was used to compare the CRD only gal3 gene with the full length, wild type gal3 gene. The results show that there was no nucleotide bases at the beginning of the CRD only gal3 gene sequence which indicates that the NTD was successfully removed from the sequence. Furthermore, the sequencing results demonstrated the second NcoI restriction site was inserted into the CRD only gal3 gene as well as the change in nucleotide bases which causes a substitution of threonine at the 134th position with cysteine.

Conclusion/ Future Work

The two gal3 constructs, gal36SC and CRD, were successfully produced in the laboratory. Because the gal3S6C conjugate was observed to precipitate out of solution, the hypothesis that covalently attaching polymer to the unstructured NTD region of gal3 protein will increase solubility is unsupported in this experiment. Also, the conjugation product consisted of both the conjugated and unconjugated gal3S6C with about 50% conversion via SDS-PAGE. Ion exchange chromatography will be used to purify the conjugates from the unmodified protein.

In contrast, the CRD-only gal3 protein construct is prepared to be expressed and purified. The protein purification affinity chromatography procedure that was established to purify gal3S6C protein will be adapted to purify the CRD-only gal3 protein construct. The procedure will be adjusted as needed to collect fractions with high protein concentration as well as high protein purity. CRD-PEO conjugates will then be created using the same reaction scheme as for gal3S6C.

References

- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). *The proteomics protocols handbook* (J. M. Walker, Ed.). Totowa, NJ: Humana Press. Protein Identification and Analysis Tools on the ExPASy Server
- Johannes, L., Jacob, R., & Leffler, H. (2018). Galectins at a glance. Retrieved May 23, 2019, from <http://jcs.biologists.org/content/131/9/jcs208884>
- Li, L., Li, J., & Gao, J. (2014). Functions of Galectin-3 and Its Role in Fibrotic Diseases. *Journal of*

Pharmacology and Experimental Therapeutics, 351(2), 336-343. doi:10.1124/jpet.114.218370

Nangia-Makker, P., Honjo, Y., & Raz, A. (n.d.). Galectin-3 Binding and Metastasis. *Metastasis Research Protocols*, 149-157. doi:10.1385/1-59259-137-x:149

NEB, New England Biolabs. Ligation Protocol with T4 DNA Ligase (M0202). Retrieved March 22, 2019, from <https://www.neb.com/protocols/1/01/01/dna-ligation-with-t4-dna-ligase-m0202>

Ochieng, J., Green, B., Evans, S., James, O., & Warfield, P. (1998). Modulation of the biological functions of galectin-3 by matrix metalloproteinases. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1379(1), 97-106. doi:10.1016/s0304-4165(97)00086-x

Voytas, D. (2001). Agarose Gel Electrophoresis. *Current Protocols in Molecular Biology*. doi:10.1002/0471142727.