Western Kentucky University
TopSCHOLAR®

Honors College Capstone Experience/Thesis Projects

Honors College at WKU

2019

Investigation of Transient Expression System in Nicotiana bethamiana to Produce Novel TALEs and the Development of (Ds)DNA Detection with Quantam Dot-Labeled Proteins in Graphene Oxide Quenching Arrays

Colleen Jackson

Follow this and additional works at: https://digitalcommons.wku.edu/stu_hon_theses

Part of the Biochemistry Commons, Biotechnology Commons, and the Molecular Biology Commons

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Honors College Capstone Experience/Thesis Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.

INVESTIGATING A TRANSIENT EXPRESSION SYSTEM IN *NICOTIANA BENTHAMIANA* TO PRODUCE NOVEL TALES AND THE DEVELOPMENT OF (DS)DNA DETECTION WITH QUANTUM DOT-LABLED PROTEINS IN GRAPHENE OXIDE QUENCHING ARRAYS.

A Capstone Project Presented in Partial Fulfillment

of the Requirements for the Degree Bachelor of Arts and Sciences

with Honors College Graduate Distinction at

Western Kentucky University

By

Colleen Jackson

May 2019

*****

CE/T Committee:

Professor Moon-Soo Kim, Chair

Professor Claire Rinehart

Professor Christopher Keller

Copyright by
Colleen Jackson
2019

I dedicate this thesis to my mother, Amy Monday, and to Uriah Albrink and Dr. Jeremy Maddox, two friends and mentors that taught me never to be content with just knowing and not understanding.

ACKNOWLEDGEMENTS

I would like to acknowledge all the members of the Kim laboratory that have helped me complete my work on this project including, Dat Thinh Ha, Ji Young Shim, Wendy Cecil and Dr. Moon-Soo Kim herself. I would also like to thank the members of Chung-Ang's Therapeutic Protein Engineering Laboratory for their guidance in South Korea and in new biochemistry techniques.

ABSTRACT

Our objective was to develop a faster method of quantitively detecting double stranded (ds)DNA of pathogenic bacteria such as the Shiga Toxin 2 gene present in *E. coli* O157. Transcription Activator-Like Effectors (TALEs) are a new class of DNA-binding proteins which selectively bind to dsDNA with the 12th and 13th amino acids of each repeat, called repeat variable diresidues (RVDs). Novel TALE proteins were designed to target the stx2 gene and were cloned into existing AvrBs3 TALE protein in the pMAL c2x vector system for bacterial BL-21 *E. Coli* expression. The protein's DNA-binding region was then subcloned pEAQ vectors for expression in *N. Benthamiana* expression systems. Protein expression was comparable between systems, as plant expression had a higher yield at the cost of additional time and resources. After expression, TALE proteins were purified using affinity chromatography and characterized. TALE proteins were then labeled with CdSe/ZnS Quantum Dots (QDs) using EDC and NHS labeling method to create a stable peptide bond between the QD and protein. Labeling efficiency was very high ~90% with relatively low initial protein concentration. Labeled proteins were then used in Graphene oxide (GO) quenching and sensitivity arrays to observe if GO could quench QD signal and if target DNA could restore QD signal. In both experiments, QD were excited with 300 nm light and 515 nm endpoint fluorescence data was collected. Optimal sensitivity array conditions were determined to be 10 nM protein and 2 µg/ml GO, and DNA sensitivity arrays could be completed in under 30 minutes. It is important to further investigate the binding capabilities of the TALE proteins and the sensitivity of this system in the presence on nontarget DNA and complex biosamples.

VITA

EDUCATION

Western Kentucky University, Bowling Green, KY — May 2019

B.A. in Chemistry – Mahurin Honors College Graduate

Honors Capstone: *Investigating a Transient Expression System in Nicotiana benthamiana to produce Novel TALEs and the development of (ds)DNA detection with Quantum Dot-labeled proteins in Graphene oxide Quenching Arrays.*

Fleming County High School, Flemingsburg, KY — May 2015

PROFESSIONAL EXPERIENCE

Department of Chemistry, WKU — January 2019-
Teaching Assistant — May 2019

Department of Chemistry, WKU — May 2017-
Undergraduate Researcher — May 2019

Department of Biology, WKU — August 2017-
Student Ambassador — August 2018

AWARDS & HONORS

Magma Cum Laude, WKU, May 2019
International Research Experience for Students, NSF, Summer 2018
Faculty Undergraduate Student Engagement, WKU, August 2017
Lifetime Experience Grant, WKU, April 2017
Scholarship in Science, Technology, Engineering, and Mathematics Program, NSF, January 2016-May 2019

PROFESSIONAL MEMBERSHIPS

American Chemical Society (ACS)

INTERNATIONAL EXPERIENCE

Chung-Ang School of Medicine — June 2018-
National Science Foundation's International Research Experience — August 2018

for students

PUBLICATIONS

Slocum, D. W., Maulden, E. A., Whitley, P. E., Reinscheld, T. K., Jackson, C. S. and Maddox, J. B. (2017), *Anomalous ortho-Proton Acidities of the para-Haloanisoles*. Eur. J. Org. Chem., 2017: 6882-6884

PRESENTATIONS

C. S. Jackson, W. Cecil, K. Kim, K. Ko, M. S. Kim. (2019, April). *Expression of TALEs in plants and development of TALE arrays for diagnostic applications*. Poster presented at The American Chemical Society Annual National Spring Meeting, Orlando, FL

C. S. Jackson, W. Cecil, K. Kim, K. Ko, M. S. Kim. (2019, March). *Diagnostic Applications of Double-Stranded DNA Detection Utilizing Tale Proteins Expressed in Plants.* Poster presented at the WKU Student Research Conference. Bowling Green, KY.

C. S. Jackson, H. Evie, and B. Williams. (2018, March). *Cytotoxicity of Novel Platinum Compound in Models of Mammalian Cancer*. Poster presented at National Society of Toxicology Annual Meeting, San Antonio, TX

CONTENTS

Acknowledgements iv

Abstract v

Vita vi

List of Figures ix

Background 1

Introduction 6

Materials and Methods 11

Results and Discussion 18

Conclusion 24

References 26

Appendex: Supplemantal Figures 28

LIST OF FIGURES

Figure 1. Bound and unbound conformations of two naturally occurring TALE proteins 2

Figure 2. Schematic of naturally occurring TALE protein ... 2

Figure 3. Structure of 2D GO ... 9

Figure 4 System diagram showing QD-labeled protein signal and FRET in the presence or absence of target DNA ... 10

Figure 5. Design of novel TALE stx2 236 RVD sequence ... 12

Figure 6. Diagram of EDC and NHS chemistry for creating a stable peptide bond. ... 16

Figure 7. SDS-PAGE gel of stx2 236 protein from bacterial purification ... 18

Figure 8. SDS-PAGE gel of stx2 236 protein from bacterial purification ... 19

Figure 9 A. Quenching Efficiency for QD quenching array ... 22

Figure 9 B. Spectrum fluorescence for QD quenching array region of linearity ... 22

Figure 10 A. Endpoint fluorescence for 10 nM QD-TALE and 2 µg/ml GO with various DNA concentrations ... 23

Figure 10 B. Calculated Fluorescence Restoration for 10 nM QD-TALE and 2 µg/ml over increasing DNA concentrations. ... 23

BACKGROUND

Transcription activator-like effectors, or TALE proteins, are infectious plant proteins expressed by the pathogenic plant bacteria *Xanthomonas* and are secreted to infect a variety of host plants, including rice (Hopkins, 1992). Initially, these bacteria were investigated as a cause of plant disease and death in commercial crops before the discovery of TALE proteins and their deoxyribonucleic acid (DNA)-binding capabilities. Similarly to zinc finger proteins (ZFP), these proteins can directly bind to specific nucleotides or nucleotide groups in the major groove of double-stranded (ds)DNA. Unlike ZFP, the structure, the mechanism of nucleotide specificity, and the number of nucleotide targets are completely unique.

Much of the information that has been discovered, concerning the structure and function of TALE has been acquired through the investigation of the AvrBs3 family of TALE proteins. Studies into other TALE families differ mostly in host specificity and DNA sequence targets (Ackerveken, 1996). TALE proteins take on what is referred to as a "sunflower" conformation, with the majority of the unbound protein comprised of a series of paired left-handed α-helices curled gradually into a larger right-handed α-helix, (See Figure 1). This curve in the overall protein fits well in the right-handed α-helix of dsDNA. During binding the protein anneals itself into the major groove, adopting a tighter sunflower conformation and slightly distorting the dsDNA's shape (Dong, 2012).

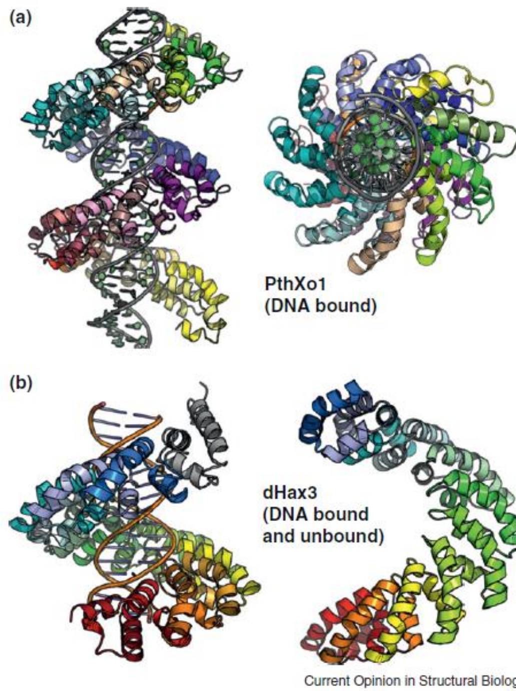
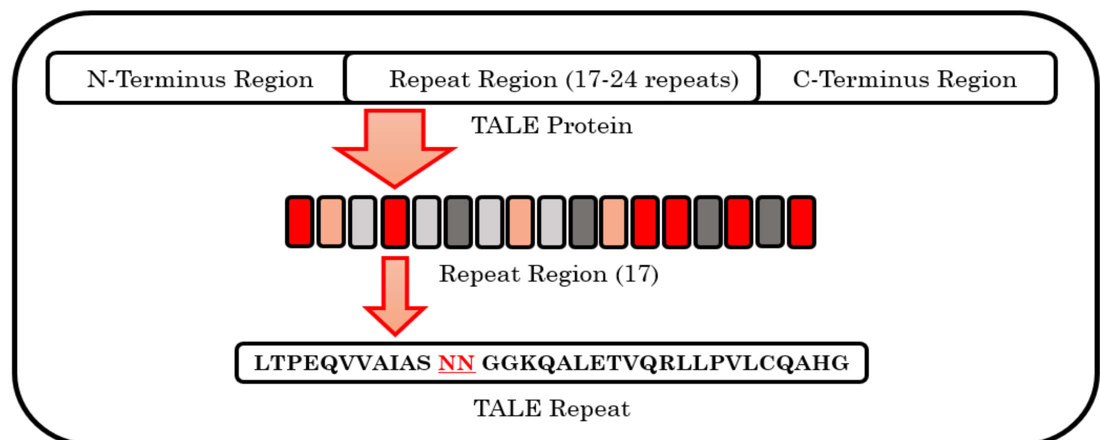


Figure 1 (Left): Bound and unbound conformations of two naturally occurring TALE proteins, PthXo1, with target DNA from Dong et al (2012).

Figure 2 (Below): Schematic of naturally occurring TALE protein, central repeat domain, and a single TALE repeat with single letter amino acids abbreviations containing two Asparagine residues (red) as the repeat variable diresidues (RVDs).



Natural TALE proteins generally consist of three parts, an N-terminus flanking region, C-terminus flanking region, and a central DNA-binding repeat region (See Figure 2). The very beginning of the N-terminus flanking region contains a T3S signal which allows for chaperoned protein secretion out of the bacterial cell, through the host plants' cell wall and eventually to the nucleus (Boch, 2010). This is critical as the compartmentalizing of the eukaryotic host cells requires the infectious prokaryotic cell to

utilize the host's own trafficking mechanism in order to transport the TALE where it will be possible to bind to the host DNA, i.e. the nucleus. Proteolytic cleavage studies, mass spectroscopy, and sodium dodecyl sulfate polyacrylamide (SDS-PAGE) gels by Gao et. al. seems to imply that proper folding of the N-terminus region and the central DNA binding domain results in additional stabilizing interactions between the two regions (2012). This can somewhat explain previous studies in TALE binding affinity where the length of the flanking terminal regions did influence TALE binding capabilities, even though TALE binding mechanism and specificity is controlled by the central repeat domain (CRD) (Meckler, 2013).

The DNA-binding CRD is the most well-understood region of the TALE gene and protein. It consists of multiple nearly identical 34 amino acid repeats, followed by a conserved half repeat consisting of the first 17 amino acids. Naturally, TALE CRDs have 17.5 to 24.5 tandem repeats, but engineered proteins are found to be stable with 1.5 to 33.5 repeats (Mak, 2013). Each whole repeat takes on a helix-turn-helix motif and in the center of the turn are two variable amino acids corresponding to the 12th and 13th positions (Dong, 2012). These anomalous residues, termed repeat variable di-residues (RVDs), control the nucleotide specificity and overall protein DNA-binding capabilities. Different RVDs have different degrees of selectivity and binding strength, many of which have been quantified using electrophoresis mobility shift assay (EMSA) (Meckler). Interestingly, all TALE proteins require a 5' thiamine, independently of the other RVDs before the protein can attempt to bind to the DNA. When the first nucleotide is identified, RVDs bind sequentially in the 5' to 3' direction, resulting in higher sensitivity to single nucleotide substitutions near the 5' end. A set of four specific RVDs have been identified for use in engineered

proteins, as they are the most selective of the known, naturally occurring RVDs for each of DNA's bases. Asparagine-Isoleucine (NI), Asparagine-Glycine (NG), Asparagine-Asparagine (NN), and Histidine-Aspartate (HD) are used in engineered proteins to bind adenine, thymine, guanine, and cytosine bases respectively (Meckler).

Finally, the C-terminus of several TALE proteins has been found to contain a nuclear localization signal to ensure that infectious protein reaches the site of its binding target. (Szurek, 2002). Also, Szurek et. al. identified in this region an acidic transcription activation domain, which are domains that affect gene expression in the host nucleus after the protein is fully bound to the target DNA strand (2001). Otherwise, there is no evidence that the length of the C-terminus region affects overall TALE binding, unlike the N-terminus region.

The unique structure and location of the RVDs in TALE proteins is not only interesting in the diversity of their host species, but because these proteins have a relatively simple binding motif with only one RVD corresponding to only one nucleotide specificity, making them extremely modular. In practice, designer TALE protein have been reverse engineered from the target sequence using the RVDs identified by EMSA. This can create custom proteins capable of locating a specific known gene sequence. Furthermore, labeling or cloning additional dyes, proteins, or nanoparticles onto these proteins has allowed for both the detection of dsDNA, and targeted genetic editing. For example, TALEs have been utilized with the protein β-lactamase to develop colorimetric dsDNA detection systems (Ooi, 2006), and with nonspecific nucleases to create TALENs, an alternative to clustered regularly interspaced short palindromic repeats (CRISPR)-based gene editing tools (Joung, 2013). Although TALEs have not yet been studied as thoroughly as CRISPR and ZFP, they

still stand independently as a viable option for designer proteins for specific dsDNA targeting.

INTRODUCTION

The rise of antibiotic resistant pathogens and improvements in the understanding of genetic biomarkers for medical conditions, such as cancer, have created a necessity for improvements in the rapid and sensitive detection of known DNA sequences. There are a few techniques that are currently employed to identify target DNA sequences including those that code for resistance in pathogens, such as DNA denaturation followed by hybridization, and culturing methods. Both methods require an appreciable detection time, ranging from a few hours for denaturation and hybridization, to a few days for some culturing methods, making them impractical for point-of-care (POC) or field detection. DNA denaturation and subsequent hybridization methods, such as polymerase chain reactions (PCR), require denaturing DNA into single strands and binding of synthesized nucleotide primers. The hybridization step requires sensitive adjustments to temperature and usually employ expensive lab equipment such as thermocyclers. Furthermore, the primers which bind to the DNA target must be designed to satisfy several specific requirements. These requirements significantly hinder the possible targets for hybridization methods. Finally, the formation of triple-helixes and non-specific binding are also major limitations in hybridization-based DNA detection methods as these forms can impede on detection or extension. By designing a dsDNA detection method that does not require DNA or RNA probes, these limitations may be circumvented.

The objective of this project was to investigate and develop a new, faster quantitative method of identifying specific DNA sequences that are known to be associated with antibiotic resistant and pathogenic strains of bacteria, for example, the Shiga toxin 2

(stx2) from *E. coli* O157. This common pathogen causes foodborne illness, which the Center for Disease Control and Prevention (CDC) estimated in 2011 to hospitalize about 3,600 individuals a year. The Shiga toxin causes severe gastrointestinal and urinary distress in humans and can sometimes lead to kidney failure or death. Our device would employ the dsDNA-binding capabilities of naturally occurring proteins, TALEs, as detection probes to bind to target dsDNA. The probes would be used in conjunction with a signal transducer, to produce a detectable and quantifiable signal change when the probe is bound to target DNA. Furthermore, to be useful as a new DNA detection system, it must have rapid and quantitative detection of DNA, possibly even in the presence of complex genomic DNA.

In conjunction with TALEs used as dsDNA probes, the signal transduction of our system employs two components, CdSe/ZnS quantum dots (QD) and graphene oxide (GO), which allows for easy development into single-channel multiplexing and an increase in system signal proportional to target DNA detection. QDs are metallic, semiconducting nanoparticles that can fluoresce at monochromatic wavelengths after excitation by a single shorter wavelength, depending on the size and composition of the nanoparticle This phenomenon is a consequence of quantum confinement, which can be described by the quantum mechanical model of particle-in-a-box. This model describes how the energy levels of an emitted photon are dependent upon the size of the region the energy is confined into (i.e. the nanoparticle's diameter). Therefore, the larger the QD, the smaller the energy difference between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), and the lower in energy the resulting emitted photon (Soloviev, 2000). Lately, QDs have been investigated as bio-labels by utilizing

surface carboxyl groups on some QDs which can react with other terminal functional groups, such as primary amines on the N-terminus of proteins to form stable covalent bonds (Xing, 2011). By having different sizes of QDs bond to different TALE proteins targeting a series of DNA sequences, a single short wavelength may be used to excite many of the QDs in a single analysis, resulting in a polychromatic output with multiple specific emission wavelengths corresponding to the presence of multiple specific DNA targets in a single sample.

To allow for direct detection of dsDNA, our application of this technique also involves fluorescent restoration, where QD-labeled TALE protein signal is quenched by proximal graphene oxide (GO) sheets through fluorescent resonance energy transfer (FRET). GO sheets are similar to graphene as they are a two-dimensional surface composed of a sheet of sp2 hybridized carbons, creating a large conjugated π-system located above and below the molecular plane. This portion of GO is very hydrophobic, and its π-system has the ability to absorb photons throughout the UV-Vis region. Unlike graphene, however, GO also contains a variety of very hydrophilic groups on the perimeter of the carbon ring structure, making it somewhat immiscible in aqueous solutions and usable in aqueous biological samples (Georgakilas, 2016).

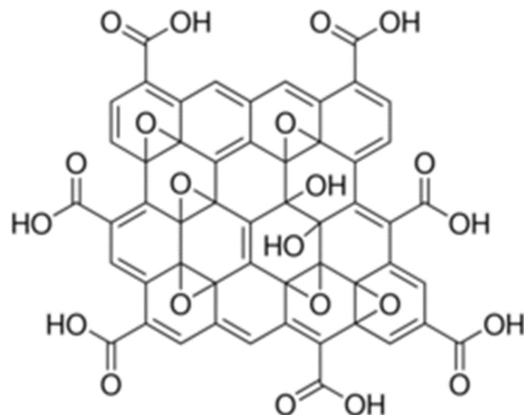


Figure 3. Structure of 2D Graphene Oxide. Image from: Kim S, et. el. PubChem 2019 Nucleic Acids Res. 2019 Jan 8; 47

In this system, FRET occurs when the QD-labeled protein adsorbs onto the GO surface through hydrophobic interactions, π-π stacking between the aromatic graphene rings and aromatic side chains available on the surface of the protein, and some hydrogen bonding between hydrophilic side chains on the protein and the hydrophilic groups on the edge of the GO sheets (Xing). These three interactions are relatively weak, making the adsorption of the protein very reversable. Because the protein is adsorbed onto the GO, the distance between the QD and GO is relatively short and therefore when the QD emits a lower energy photon, the photon is very likely to be absorbed and dispersed by the GO (Dong 2010). Essentially, the energy released by the QD via a photon is often transferred directly to the GO, resulting in a reduction of overall QD signal, called FRET signal quenching (Chou, 2015).

FRET in this system is dependent on the adsorption of the protein onto GO via relatively weak interactions that can be easily broken by a conformational change or physical methods of separation (Zhang, 2018). Our method utilizes this by introducing target hairpin dsDNA to the sample containing the TALE DNA-binding protein. TALE is

known to change into a denser, and shorter sunflower conformation when bound to its target DNA (Dong, 2012). When binding to target dsDNA, the affinity and covalent bonding between TALE-QD and the target dsDNA is significantly greater than the interactions between the protein and GO. Consequently, when target dsDNA is present, DNA binding induces a conformational change in the protein and the protein-GO interactions are weakened, often resulting in dissociation of the TALE-QD from the GO surface. In this case, the distance between the QD and GO increases, and the effect of FRET is reduced, restoring QD signal. Overall, this system detects target DNA through the restoration of QD emission, as restored signal increases directly with target DNA concentration.

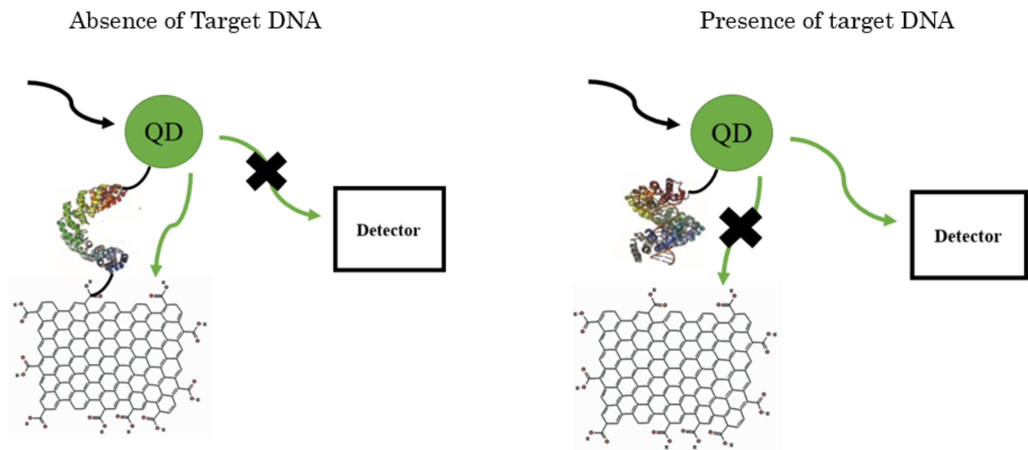


Figure 4. System diagram showing QD-labeled protein signal and FRET in the presence or absence of target DNA. In the absence of target DNA, hydrogen bonding, hydrophobic interactions and π-π stacking adsorbs unbound TALE-QD protein onto GO. In the presence of DNA, TALE-DNA binding induces a conformational change and the protein-QD is released from the GO.

MATERIALS AND METHODS

Engineering and Expression of TALE Proteins

First, TALE proteins were engineered and then expressed using two different protein expression systems in order to investigate whether *E. coli* or *N. benthamiana* could better express adequate amounts of custom TALE proteins for the development of multiple TALE dsDNA detection systems. TALE proteins were designed to be specific to various regions on the DNA sequence of the stx2 gene present in *E.coli* O157 (see Figure 3 Supplemental Figure 1). The N and C-terminal truncated regions were maintained by using the sequence of naturally occurring AvrBs3 protein (see Supplemental Figure 2). This was done to maintain natural protein functionality and localization signals. The two proteins targeted two different, but adjacent positions on the stx2 gene, and were made by replacing the naturally occurring AvrBs3 central binding region with a novel 17.5 repeat long central DNA-binding region, which were designed using four nucleotide specific RVDs identified in literature (Meckler) (see Figure 3 and Supplemental Figure 1). An 18 base pair (bp) long DNA target region, including a starting 5' thiamine, was chosen and used to design the protein because an 18 bp long region of DNA is long enough to be specific for a single site in the *E. coli* genome.

5' -> 3' stx2 DNA Sequence (238-258)	G	A	T	G	T	C	T	A	T	C	A	G	G	C	G	C	G
TALE (A) stx2 236 RVDs	NN	NI	NG	NN	NG	HD	NG	NI	NG	HD	NI	NN	NN	HD	NN	HD	NN

Figure 5: Design of novel TALE stx2 236 RVD sequence to target Shiga Toxin gene at two adjacent locations. Single letter amino acid abbreviations are used to identify proteins in the 12th and 13th positions of each TALE dsDNA-binding repeats.

The DNA coding for designer stx2 TALE DNA-binding regions were originally ordered in pUC vectors from Bio basic and then the central binding domain inserts were sub-cloned into multiple cloning sites in custom pMAL c2x cloning vectors containing AvrBs3 C and N-termini, using the *Stu*I and *Aat*II restriction sites (see Supplemental Figure 3). The stx2 inserts were ligated into the pMAL vector and were immediately transformed into 10-β *E. coli* cells (Invitrogen). Ligation was confirmed through PCR before isolating DNA from successful recombinant colonies and was transformed into BL21 *E. coli* (Invitrogen).

To prepare for plant expression using Peyret et. al.'s method (2013), in-fusion cloning (Takara Bio) was utilized to subclone the TALE protein insert from pMAL vector (located at the N-terminus of Maltose Binding Protein (MBP)), into the pEAQ vector at the *Age*I restriction site (see Supplemental Figure 4). This involved designing PCR oligonucleotides with 3' ends complementary to the insert and 5' end complementary to the vector as described by Takara Bio in-fusion protocol. The primers were also designed to contain the prokaryotic start codon and with an additional N-terminus glycine to

maintain TALE reading frame into a downstream double Histidine purification tag (His Tag). The PCR amplified inserts were isolated with gel extraction and were ligated into linearized and calf intestinal alkaline phosphatase (CIAP) treated pEAQ vector to prevent self-ligation. Ligation products were immediately transformed into DH5α *E. coli*. Recombinant DNA was isolated from both TALE stx2 236 and TALE stx2 255 and then the in-fusion cloning was confirmed through sequencing.

Correct TALE pEAQ plasmids were transformed into *A. tumefaciens* and then agro-infiltrated into *N. benthamiana* leaves for expression. After two weeks of transient plant expression, proteins were isolated from leaf samples. Plant protein expression and protein isolation were performed as outlined in literature by Park et. al. (2015). Finally, TALE proteins were purified through Nickel column purification which utilizes affinity chromatography through binding between the resin and His tags on the protein. Proteins were eluted from the columns with a solution containing high concentrations of imidazole (~500 mM). For long term storage of purified TALE proteins (i.e. more than two weeks), a buffer exchange was used to concentrate protein into TALE Storage buffer (480 mM Potassium chloride, 1.6 mM Disodium ethylenediaminetetraacetic acid (EDTA) dihydrate, 12mM Tris-Cl pH 7.5, 0.6 μM tris(2-carboxyethyl)phosphine (TCEP), 2.8 μM dithiothreitol (DTT)), containing two reducing agents, DTT and TCEP, which prevent oxidation of cystine residues in proteins and therefore maintain tertiary and quaternary protein structure and function for several additional weeks.

To compare yield with the new plant expression system, TALE A protein was also purified from pMAL c5x transformed *E. coli* using the isopropyl β-D-1-thiogalactopyranoside (ITPG) protein expression system. This system inactivates the LacI

repressor gene located upstream and allows for the expression of TALE and other connected plasmid proteins. In the pMAL vector, TALE proteins are flanked by MBP and a single His tag (see Supplemental Figure 3) to both increase solubility and to allow for alternate purification options, such as amylose or Nickel column chromatography. Transformed cells were used to inoculate 2 liter cultures of Luria-Bertani (LB) medium and were incubated at 37°C until it reached an optical density (OD) of 0.5~0.6. Cell were then cooled on ice until OD reached 0.6, where protein expression was induced by adding ITPG to create a final concentration of 0.1 mM.

Next, overnight (16 hours) expression occurred at 16°C to prevent inclusion body formation. Then cells were pelleted through centrifugation at 3500 rotations per minute (rpm) for 20 minutes, resuspended in TALE Lysis buffer (2 mM Sodium azide, 500 mM Sodium chloride, 5 mM Imidazole, 20 mM Tris-Cl pH 8), lysed using sonication (50% amplitude, for 7, 10 second intervals with 20 second breaks), and centrifuged to pellet insoluble cellular debris (at 15,000 rpm for 40 minutes). Cell lysate supernatant containing cytoplasmic, soluble proteins including TALE-MBP were collected and purified through a nickel column (Bio Rad). Column purification was conducted at 4°C with a bed height of 3 cm. Column was equilibrated with 75 ml of TALE Lysis buffer, then the sample was passed, followed by 100 ml of TALE Lysis buffer, 100 ml of TALE High-Salt wash (2 mM Sodium azide, 2 M Sodium chloride, 5 mM Imidazole, 20 mM Tris-Cl pH 8), and another 100 ml of TALE Lysis buffer. Finally, proteins were eluted with 10 ml of TALE elution buffer (2 mM Sodium azide, 500 mM Sodium chloride, 500 mM Imidazole, 20 mM Tris-Cl pH 8). *E. coli* expressed TALE proteins were also concentrated in TALE storage buffer using a 10,000 Da filter (Pierce, Rockford, IL, USA) and stored on ice until QD-labeling.

Graphene Oxide Array Design and Optimization

Next, polyethylene glycol (PEG)-coated CdSe/ZnS QDs (Creative Diagnostics, Shirley, NY, USA) with peak emission at 515 nm were covalently attached to the N-terminus of the TALE stx2 236 proteins using 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and N-hydroxysuccinimide (NHS) chemistry (Thermo Scientific, Rockford, IL, USA), in which the EDC crosslinker reacts with carboxyl functional groups on the surface of the QD to create an unstable intermediate. In the presence of NHS, this intermediate reacts to create a second, more stable intermediate, capable of reacting with available primary amines such as the N-terminus of proteins. Optimization of this method of QD-labeling proteins was conducted previously by Ha et. al on zinc finger proteins in the Kim lab at Western Kentucky University (WKU). The optimized 1:5:4000:8000 ratio of QD to proteins, EDC, and NHS, was used with engineered TALE proteins in phosphate-buffered saline (PBS) buffer. The QD, EDC, and NHS were equilibrated at room temperature for 20 minutes while mixing before adding TALE protein and mixing for an additional two hours. The QD-labeled proteins were then collected on a 100,000 Da filter (Pierce, Rockford, IL, USA) via centrifugation and unlabeled protein flow-through concentration was measured by Bradford assay to calculate labeled protein concentration and labeling efficiency.

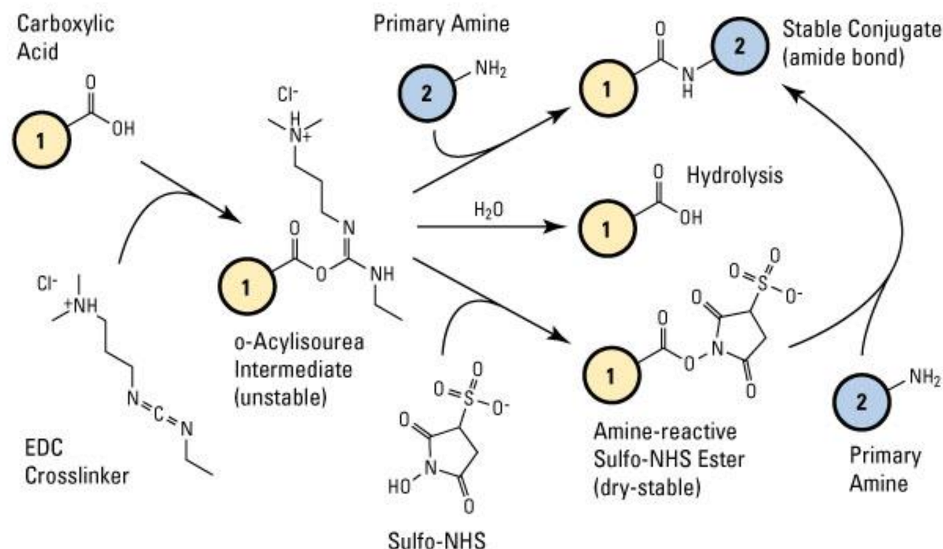


Figure 6. Diagram of EDC and NHS chemistry for creating a stable peptide bond between free carboxyl and amine groups. Image from Thermofisher Scientific.

To observe and confirm fluorescent quenching of QD-labeled protein by GO, a quenching array was performed utilizing several different concentrations of GO (ACS Material, Pasadena, CA, USA) while keeping the final concentration of QD-labeled protein constant at 20 nM. Arrays were constructed by serial dilution of a 1 mg/ml stock of GO, with experimental concentrations ranging from 10 to 1000 µg/ml. Stock GO and protein concentrations were diluted one-tenth into final reaction concentrations for the array. Each experiment in the array was completed in duplicate using black, opaque 94-well plates (Corning, Kennebunk, ME, USA). Reaction volumes were brought up to 100 µl with PBS and mixed by pipetting, before allowing 10 minutes of incubation. Quenching was then quantified by exciting QD with a 300 nm light and then a fluoresce spectrum scan from 400 to 650 nm and a 515 nm endpoint fluorescence was collected using a Synergy H1 multi-plate reader (BioTek Instruments, Winooski, VT, USA). Fluorescence quenching

was then calculated for each GO concentration by dividing average sample fluorescence intensity (F) by maximum fluorescence intensity measured in negative control containing no GO (F_0) multiplied by 100%.

$$\% \text{ Quenching efficiency} = 100 - \frac{F}{F_0} \times 100\%$$

To observe if fluorescence restoration occurs in the presence of target DNA and to estimate the sensitivity of this system, another array was conducted with protein and GO concentrations of 10 nM and 2 μg/ml, respectively. Target stx2 236 DNA oligonucleotides (Integrated DNA Technologies, Inc. Coralville, IA) were annealed into dsDNA hairpin targets and final target dsDNA concentrations between 0 and 25 nM were prepared by serial dilution. Experimental sample volumes were brought up to 100 µl with PBS and mixed by pipetting, before allowing 10 minutes of incubation. Fluorescence was measured using the sample excitation and fluorescence method described for the fluorescence quenching array. Fluorescence restoration was then calculated using the following equation where F_o and F_1 were the fluorescence intensity at the maxima in the absence and presence of GO, and F_i was the fluorescence intensity at the maxima in the presence of target DNA.

$$\% \text{ Fluorescence recovery} = 100 - \frac{F_0 - F_i}{F_0 - F_1} \times 100$$

RESULTS AND DISCUSSION

Engineering and Expression of TALE Proteins

TALE stx2 236 and stx2 255 protein expression in BL21 *E.coli* was comparable to the previous yields of TALE proteins using ITPG expression. Bacterial culturing, expression, and isolation required only a week to complete and resulted in about 2 ml of 0.3 µM, or 0.6 nanomoles of final protein as measured by Bradford assay. Nonnative SDS-PAGE gel showed two protein bands at ~76 and ~25 kDa in size. This is not consistent with theoretical predicted protein size which was expected to be 110.3 kDa (see Supplemental Figure 4) due to the presence of MBP used to increase total protein solubility.

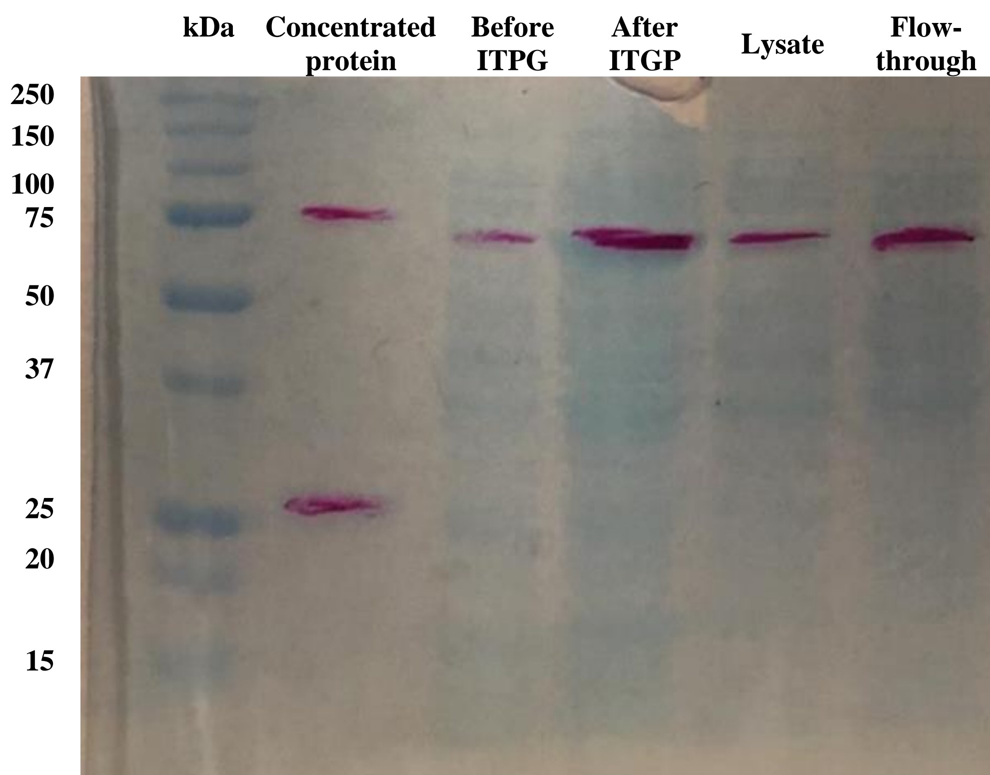


Figure 7. SDS-PAGE gel of TALE stx2 236 protein from bacterial purification. Purple lines mark bands of significant intensity. Lane 1 contains Bio Rad 1,000 Da protein ladder. Lane 2 contains the eluted protein sample after concentration and buffer exchange into TALE Storage Buffer. Lanes 3-6 are quality control samples collected through-out purification including immediately before and after ITPG expression, and immediately before and after Nickel column purification

Theoretical protein size was 67.8 kDa as MBP was replaced with additional amino acids and a double His Tag, as well as other signals from the pEAQ expression system. This method produced 2 ml of approximately 2.5 µM or 6 nanomoles of protein which was approximately 68 kDa in size and in agreement with theoretical predictions (see Supplemental Figure 5). Protein samples were also separated on a nonnative, SDS-PAGE gel and quantified with Bradford assay as shown in Figure 8.

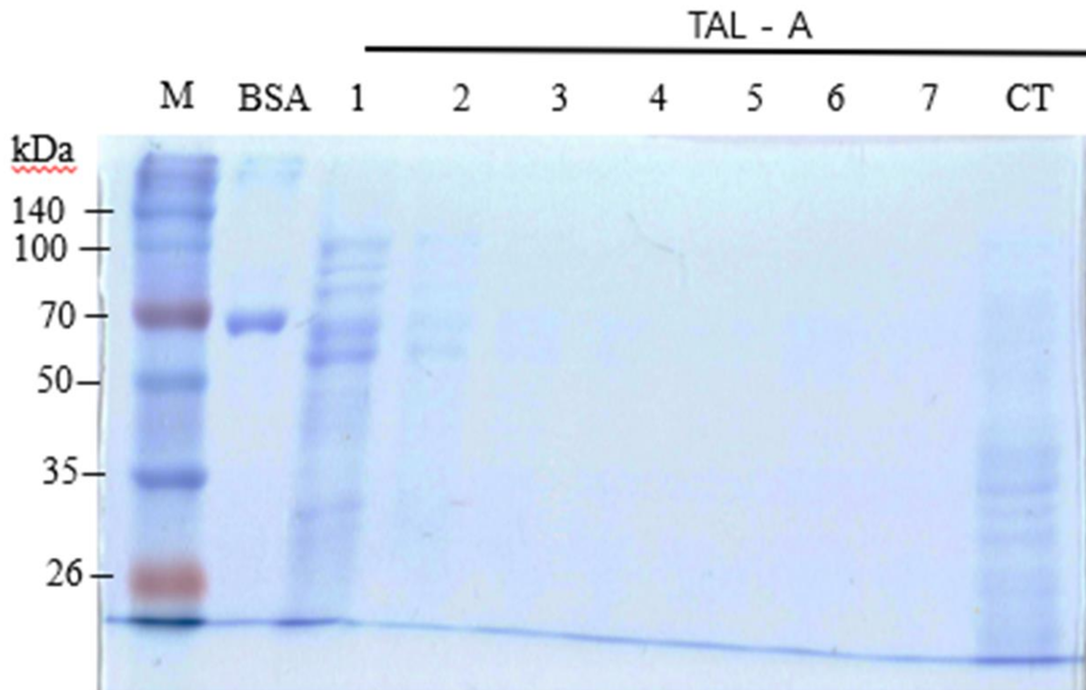


Figure 8. SDS-PAGE gel of TALE stx2 236 protein from plant purification, with protein ladder (M) in Lane 1, Bovine Serum Albumin (BSA) standard in Lane 2, TALE elution samples in lanes 3-9 and Column flow-through (CT) in lane 10.

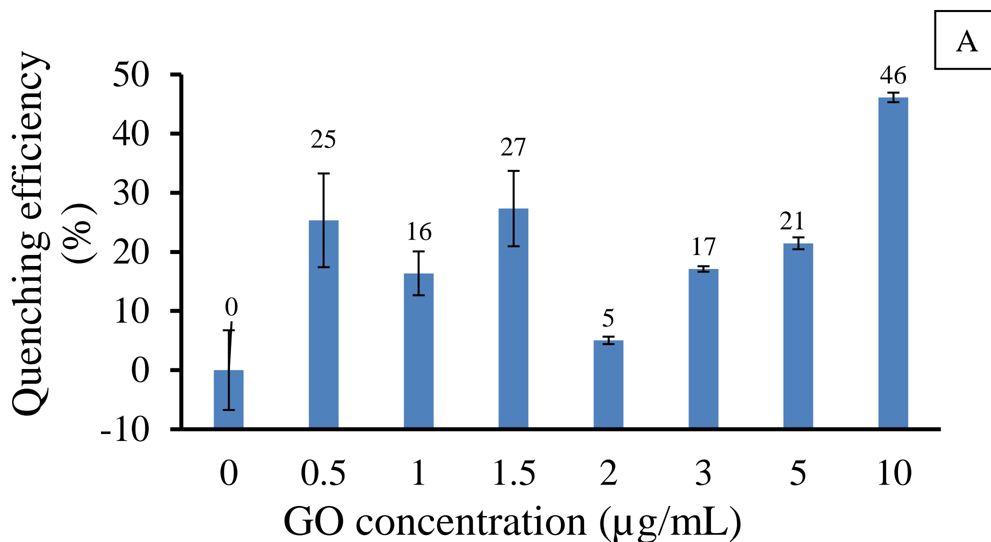
When comparing the BL21 *E. coli* and *N. benthamiana* expression systems, it is also important to consider the systems practicality as a long term method of protein expression. Transformed BL21 *E. coli* can be stored in -80°C glycerol stocks for several years and easily streaked onto a plate. From there, colonies can be used to prepare large

volume cultures in a short period of time. This is very viable for research in which only enough protein needs to be produced for a few experiments before material must be stored. On the other hand, if a lab required a constant and steady supply of TALE proteins, the pEAQ expression system can also be adjusted to create transgenic, rather than transient plants which have the potential to grow constant amounts of the protein if kept in optimal plant growth conditions (Peyret).

Graphene Oxide Array Design and Optimization

Initial protein concentration before labeling was 3.2 µM and the final concentration of labeled protein after restoring sample volume was 2.9 µM, indicating a 90% labeling efficiency for this method. TALE proteins had not yet been labeled with CdSe/ZnS QD in literature, nor in the Kim lab at WKU, but this is significantly better than previous efficiency completed with zinc finger proteins (ZFP) with the same methodology. Labeled proteins were stable for up to two weeks, afterwards total fluorescence rapidly declines and system error increases. Within one week of labeling, QD quenching and sensitivity arrays were completed to minimize system error. First, a quenching array was completed to determine the lowest concentration of GO in a stable region of linearity determined experimentally. After this region, relative fluorescence becomes unpredictable, which could indicate oversaturation of GO. Oversaturation of GO would cause significant error as the density of soluble GO is so high that the dissociation of labeled proteins and GO would not create enough distance between the QD and GO. Therefore significant amounts of background FRET would occur resulting in a false positive dsDNA result. This has also been observed by previous work in the Kim Lab at WKU with ZFPs and GO experiments.

The previous preliminary data demonstrated that GO concentrations ranging from approximately 0 to 10 µg/ml was in the linear range for labeled TALE proteins. This GO concentration range was investigated in duplicate for both endpoint and spectrum quenching data as shown in Figure 9. It was determined that 20 nM protein concentration was sufficient for quenching array, but when used in sensitivity array, QD signal was too intense to be accurately quantified with the synergy plate reader. It was determined through a 10 nM and 20 nM sensitivity trial array that 10 nM QD-labled TALES was sufficient. The DNA sensitivity array was then conducted with the final concentration of 10 nM protein and 2 µg/ml GO, shown in Figure 10. Overall the GO sensitivity array requires only about 15 to 30 minutes to complete depending on the number of samples and skill of laboratory technician. This is significantly faster than PCR and culturing methods used to detect the same pathogens.



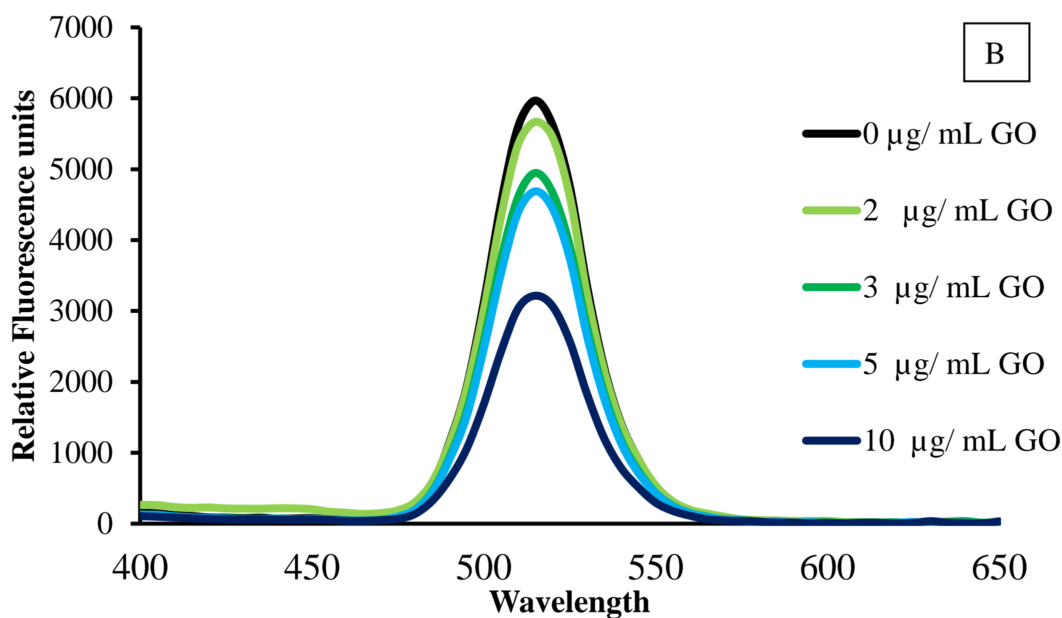
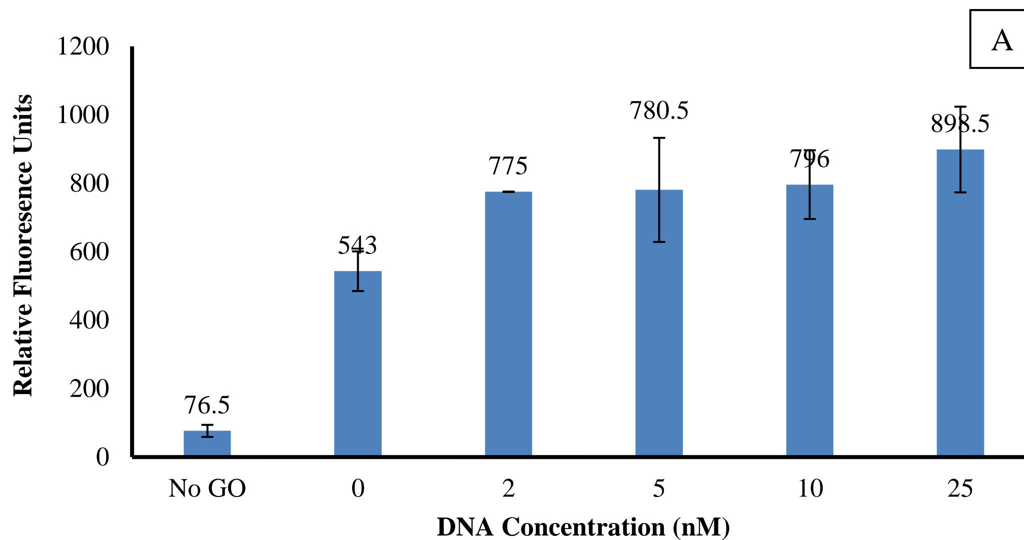


Figure 9 A: Quenching Efficiency for QD quenching array. The region of linearity was determined to be from 2 to 10 µg/ml GO at 20 nM QD-labeled TALE.

Figure 9 B: Spectrum for QD quenching array including concentrations within the determined region of linearity.



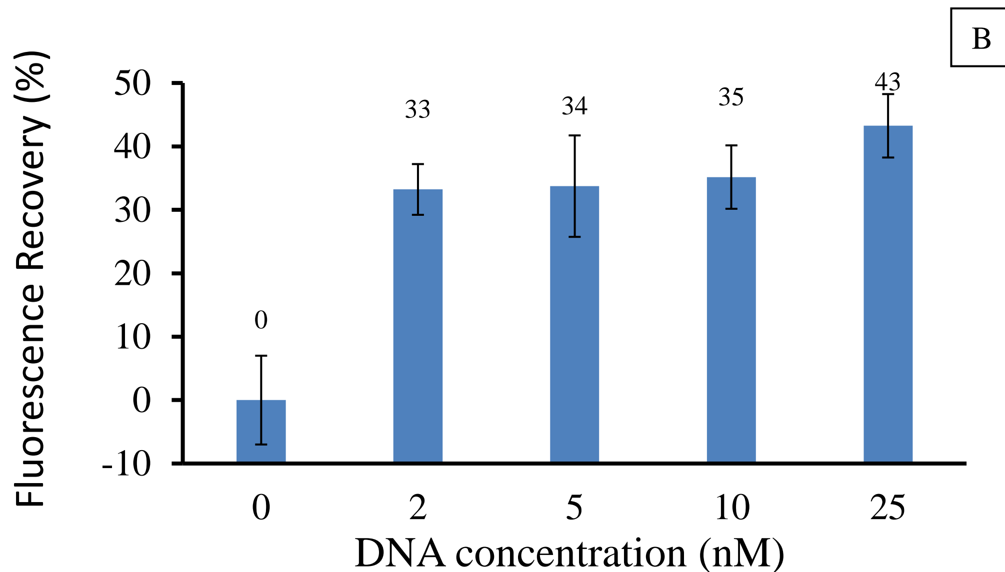


Figure 10 A: Endpoint fluorescence for 10 nM QD-TALE and 2 µg/ml GO with various DNA concentrations.

Figure 10 B: Calculated Fluorescence Restoration for 10 nM QD-TALE and 2 µg/ml over increasing DNA concentrations.

CONCLUSION

For the production of novel TALE proteins, *N. benthamiana* plant expression systems offers a comparable method of protein expression to the current BL21 *E. coli* system, but offers long-term transgenic plant benefits with the trade-off of materials for optimal plant growth. QD labeling EDC and NHS ratios previously optimized in the Kim Lab at WKU resulted in ~90% labeling efficiency for initially low concentration TALE proteins. Such a high efficiency is very promising as TALE protein expression yields are generally low. It was observed that GO quenches QD fluorescence linearly, but the system eventually becomes saturated. Target DNA restores QD signal and as DNA concentrations increase, signal increases. However, observed signal restoration was not as strongly correlated to DNA concentration as expected. Both ZFP and TALE proteins seem to have best DNA sensitivity at 10 nM of labeled protein and 2 μg/ml GO, as the high concentration of QD signal seems too intense for Synergy H1 detectors.

With no previous information or data on QD-labeled TALE proteins in literature, it is crucial that the future of this project focuses on system replicability including high labeling efficiency and GO sensitivity. Furthermore, since only on TALE protein (stx2 236) has been labeled with QD, additional studies involving labeling the recently produced TALE stx2 255 may help determine whether TALE proteins are more sensitive to QD labeling or if the generally low protein concentration results in high labeling efficiency. Since TALE conformational change and thereafter dissociation from GO is dependent on strong protein-DNA binding interactions, additional studies on binding affinity using EMSA for both TALE proteins could reveal more information regarding the relationship between protein binding affinity and GO array sensitivity. Finally, these studies on GO

sensitivity did not include any trial considering target DNA sensitivity in the presence of genomic bacterial DNA, and did not quantify background signal which could occur with nontarget DNA such as single nucleotide mismatches. Further studies on these factors may help determine if this system is sensitive and selective enough to be used as a dsDNA detection system on unprocessed bio-samples such as urine, saliva, or blood. If this system was able to sensitively detect the Shiga Toxin gene in mock bio-samples, it would further validate this system as a rapid and practical laboratory method of detecting pathogenic bacteria.

REFERENCES

Ackerveken, G. V. den; Marois, E.; Bonas, U. Recognition of the Bacterial Avirulence Protein AvrBs3 Occurs inside the Host Plant Cell. Cell 1996, 87 (7), 1307–1316.

Boch, J.; Bonas, U. Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function. Annu. Rev. Phytopathol. 2010, 48 (1), 419–436.

Chou, K. F.; Dennis, A. M. Förster Resonance Energy Transfer between Quantum Dot Donors and Quantum Dot Acceptors. Sensors (Basel) 2015, 15 (6), 13288–13325.

Dong Deng, Chuangye Yan, Xiaojing Pan, Magdy Mahfouz, Jiawei Wang, Jian-Kang Zhu, Yigong Shi, and Nieng Yan. Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. *Science* 2012 335(6069): 720–723.

Dong, H.; Gao, W.; Yan, F.; Ji, H.; Ju, H., Fluorescence Resonance Energy Transfer between Quantum Dots and Graphene Oxide for Sensing Biomolecules. Analytical Chemistry 2010, 82 (13), 5511-5517.

Gao, Haishan et al. “Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region.” Cell research vol. 22,12 (2012): 1716-20.

Georgakilas, V.; Tiwari, J. N.; Kemp, K. C.; Perman, J. A.; Bourlinos, A. B.; Kim, K. S.; Zboril, R. Noncovalent Functionalization of Graphene and Graphene Oxide for Energy Materials, Biosensing, Catalytic, and Biomedical Applications. Chem. Rev. 2016, 116 (9), 5464–5519.

Hopkins CM, White FF, Choi SH, Guo A, Leach JE. 1992. A family of avirulence genes from Xanthomonas oryzae pv. oryzae. Mol. Plant-Microbe Interact. 5:451–59

Joung, J. K.; Sander, J. D. TALENs: A Widely Applicable Technology for Targeted Genome Editing. Nat Rev Mol Cell Biol 2013, 14 (1), 49–55.

Mak, A. N.-S.; Bradley, P.; Bogdanove, A. J.; Stoddard, B. L. TAL Effectors: Function, Structure, Engineering and Applications. *Curr. Opin. Struct. Biol.* 2013, *23* (1), 93–99

Meckler, J. F.; Bhakta, M. S.; Kim, M.-S.; Ovadia, R.; Habrian, C. H.; Zykovich, A.; Yu, A.; Lockwood, S. H.; Morbitzer, R.; Elsäesser, J.; et al. Quantitative Analysis of TALE–DNA Interactions Suggests Polarity Effects. *Nucleic Acids Res* 2013, *41* (7), 4118–4128.

Park, S.-R.; Lim, C.-Y.; Kim, D.-S.; Ko, K. Optimization of Ammonium Sulfate Concentration for Purification of Colorectal Cancer Vaccine Candidate Recombinant Protein GA733-FcK Isolated from Plants. Front Plant Sci 2015, 6, 1040.

Peyret, H.; Lomonossoff, G. P. The PEAQ Vector Series: The Easy and Quick Way to Produce Recombinant Proteins in Plants. *Plant Mol Biol* 2013, *83* (1–2), 51–58.

Ooi, A. T.; Stains, C. I.; Ghosh, I.; Segal, D. J. Sequence-Enabled Reassembly of Beta-Lactamase (SEER-LAC): A Sensitive Method for the Detection of Double-Stranded DNA. *Biochemistry* 2006, *45* (11), 3620–3625.

Soloviev, V. N.; Eichhöfer, A.; Fenske, D.; Banin, U. Molecular Limit of a Bulk Semiconductor: Size Dependence of the "Band Gap" in CdSe Cluster Molecules. J. Am. Chem. Soc. 2000, 122 (11), 2673–2674.

Szurek, B.; Marois, E.; Bonas, U.; Ackerveken, G. V. den. Eukaryotic Features of the Xanthomonas Type III Effector AvrBs3: Protein Domains Involved in Transcriptional Activation and the Interaction with Nuclear Import Receptors from Pepper. The Plant Journal 2001, 26 (5), 523–534

Szurek, B.; Rossier, O.; Hause, G.; Bonas, U. Type III-Dependent Translocation of the Xanthomonas AvrBs3 Protein into the Plant Cell. Molecular Microbiology 2002, 46 (1), 13–23.

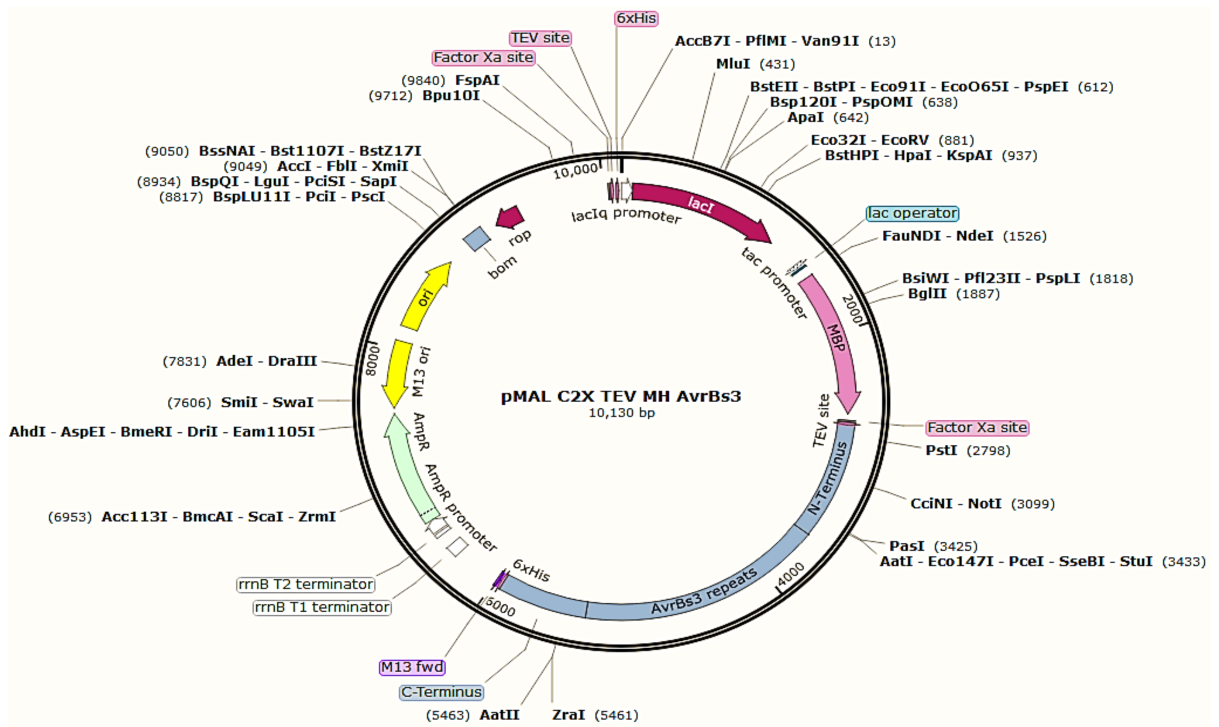
Xing, M., Shen, H., Zhao, W., Liu, Y., Du, Y., Yu, Z., and Chen, X. dsDNA-coated quantum dots. BioTechniques 2011, (50) 259–261.

Zhang, Y.; Chen, X.; Roozbahani, G. M.; Guan, X. Graphene Oxide-Based Biosensing Platform for Rapid and Sensitive Detection of HIV-1 Protease. Anal Bioanal Chem 2018, 410 (24), 6177–6185.

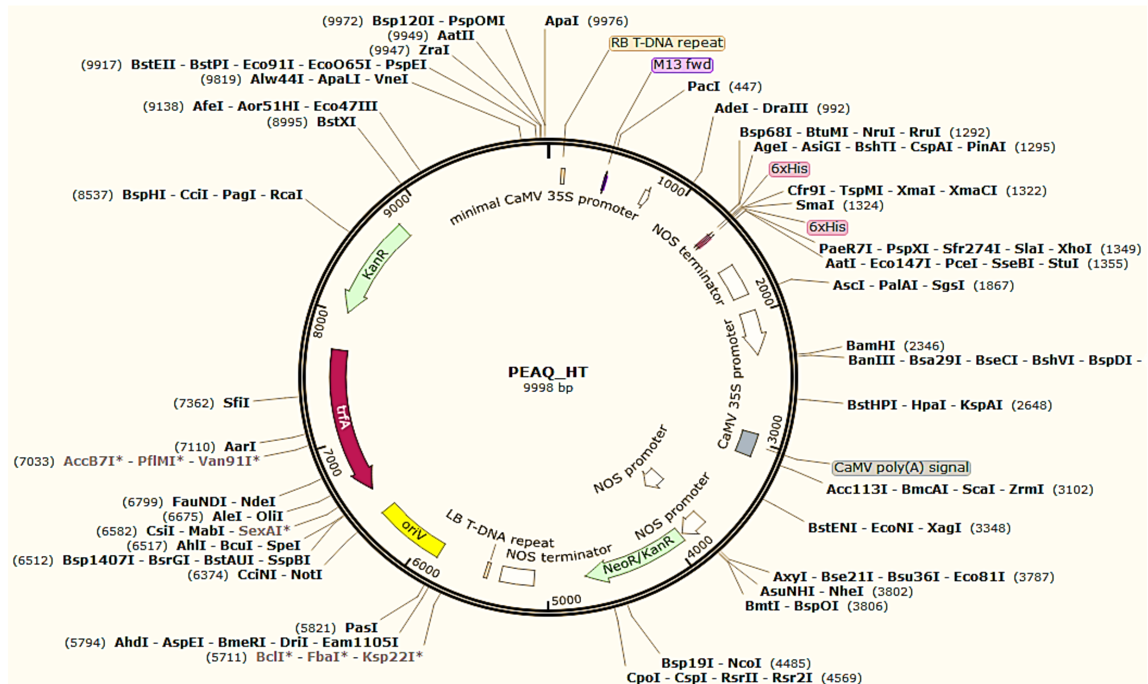
APPENDEX: SUPPLEMANTAL FIGURES

5’ -> 3’ stx2 DNA Sequence (255-274)	T	T	T	G	A	C	C	A	T	C	T	T	C	G	T	C	T
TALE (B) stx2 255 RVDs	NG	NG	NG	NN	NI	HD	HD	NI	NG	HD	NG	NG	HD	NN	NG	HD	NG

Supplemental Figure 1: Design of novel TALE stx2 255 RVD sequence to target Shiga Toxin gene at two adjacent locations. Single letter amino acid abbreviations are used to identify proteins in the 12th and 13th positions of each TALE dsDNA-binding repeats.



Supplemental Figure 2: Schematic of custom pMal c2x vector containing entire natural AvrBs3 TALE protein.



Supplemental Figure 3: Schematic of pEAQ HT vector utilized for in-fusion cloning to insert TALE stx2 236 and 255 into vector at *AgeI* restriction site, for plant expression.

Met S R T R L P S P P A P S P A F S A G S F S D L L R Q F D P S L F N T S L F D S L P P F G A H H T E
A A T G E W D E V Q S G L R A A D A P P P T **Met** R V A V T A A R P P R A K P A P R R R A A Q P S D
A S P A A Q V D L R T L G Y S Q Q Q Q E K I K P K V R S T V A Q H H E A L V G H G F T H A H I V A L
S Q H P A A L G T V A V K Y Q D **Met** I A A L P E A T H E A I V G V G K Q W S G A R A L E A L L T V A
G E L R G P P L Q L D T G Q L L K I A K R G G V T A V E A V H A W R N A L T G A P L N L T P E Q V V
A I A S H D G G K Q A L E T V Q R L L P V L C Q A H G L T P Q Q V V A I A S N G G G K Q A L E T V Q
R L L P V L C Q A H G L T P Q Q V V A I A S N S G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q
V V A I A S N G G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S N I G G K Q A L E T V
Q A L L P V L C Q A H G L T P E Q V V A I A S N I G G K Q A L E T V Q A L L P V L C Q A H G L T P E Q
V V A I A S N I G G K Q A L E T V Q A L L P V L C Q A H G L T P E Q V V A I A S H D G G K Q A L E T V
Q R L L P V L C Q A H G L T P E Q V V A I A S H D G G K Q A L E T V Q R L L P V L C Q A H G L T P Q
Q V V A I A S N G G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S N S G G K Q A L E
T V Q A L L P V L C Q A H G L T P E Q V V A I A S N S G G K Q A L E T V Q R L L P V L C Q A H G L T P
E Q V V A I A S H D G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S H D G G K Q A L
E T V Q R L L P V L C Q A H G L T P E Q V V A I A S H D G G K Q A L E T V Q R L L P V L C Q A H G L
T P Q Q V V A I A S N G G G R P A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S H D G G K Q
A L E T V Q R L L P V L C Q A H G L T P Q Q V V A I A S N G G G R P A L E S I V A Q L S R P D P A L A
A L T N D H L V A L A C L G G R P A L D A V K K G L P H A P A L I K R T N R R I P E R T S H R V A D H
A Q V V R V L G F F Q C H S H P A Q A F D D A **Met** T Q F G **Met** S R H G L L Q L F R R V G V T E L E
A R S G T L P P A S Q R W D R I L Q A S G **Met** K R A K P S P T S T Q T P D Q A S L H A F A D S L E R
D L D A P S P **Met** H E G D Q T R A S S R K R S R S D R A V T G H H H H H H **Stop** K L

Supplemental Figure 4: Single letter amino acid abbreviations for bacterial-expressed TALE stx2 236. Overall smaller large size is due to the presence of MBP and full N and C-terminus flanking regions.

Met P L D T G Q L L K I A K R G G V T A V E A V H A W R N A L T G A P L N L T P E Q V V A I A S N N
G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S N I G G K Q A L E T V Q R L L P V L C
Q A H G L T P E Q V V A I A S N G G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S N
N G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S N G G G K Q A L E T V Q R L L P V
L C Q A H G L T P E Q V V A I A S H D G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S
N G G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S N I G G K Q A L E T V Q R L L P V
L C Q A H G L T P E Q V V A I A S N G G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S
H D G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S N I G G K Q A L E T V Q R L L P V
L C Q A H G L T P E Q V V A I A S N N G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S
N N G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S H D G G K Q A L E T V Q R L L P
V L C Q A H G L T P E Q V V A I A S N N G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A
S H D G G K Q A L E T V Q R L L P V L C Q A H G L T P E Q V V A I A S N N G G R P A L E S I V A Q L
S R P D P A L A A L T N D H L V A L A C L G D V P G **Met** H H H H H H P G H H H H H H **Stop**

Supplemental Figure 5: Single letter amino acid abbreviations for plant-expressed TALE stx2 236. Overall larger large size is due to the absence of MBP and partial N and C-terminus flanking regions.