THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOENGINEERING

PROXIMITY LIGATION FOR PROTEIN DETECTION AND PROTEIN-COMPOUND INTERACTION

JENNIFER L. MCCAFFREY
Spring 2012

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Bioengineering
with honors in Bioengineering

Reviewed and approved* by the following:

William O. Hancock
Associate Professor of Bioengineering
Thesis Supervisor

Peter J. Butler
Associate Professor of Bioengineering
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

Abstract

The ability to generate compounds to treat or cure a disease is central to the pharmaceutical industry. GlaxoSmithKline (GSK) has developed Encoded Library Technology (ELT) for the synthesis and selection of DNA-encoded libraries (DELs) containing millions to billions of potential drug candidates. There is a need for an efficient method to detect the interactions between proteins and compounds in cells and tissues. Proximity ligation is a relatively new method for highly specific and sensitive detection of proteins in solution, cell culture, or localized tissue. Modifying proximity ligation to quantify the interactions between proteins and compounds using DNA-encoded compounds will prove to be beneficial to drug discovery.

Applied Biosystems produces a TaqMan Assay kit that uses the method of proximity ligation with a TaqMan probe to detect protein levels. A modification of the TaqMan Assay was made by replacing an antibody of the probe pair with a DNA-encoded compound in order to quantify protein-compound interactions. These experiments were not accurately characterized with a nonlinear regression, one- site specific binding model. A new model was generated to quantify protein-compound interactions. Although the assay produced a protein dependent signal with both purified protein and cell lysate, there were discrepancies between the experimental data and the theoretical model. The probe concentrations, oligonucleotide lengths, and the connector length need to be optimized in order to quantify the interactions between proteins and potential drug candidates.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would first like to thank my GSK co-op supervisor, Dr. Xin Zeng, and GSK manager, Dr. Xining Wu for allowing me to be a part of this cutting edge research. They taught me so much about how to properly research, conduct experiments, and analyze data. I will remember what they have taught me and be sure to apply it in my future work. From this wonderful co-op experience, I have decided to pursue a career in the pharmaceutical industry. I would also like to thank Julie Cox, also from GSK, for assisting me in the lab. With her help, I was able to learn the equipment and process my data.

I would also like to thank my Thesis Supervisor, Dr. Hancock. He willingly took on the responsibility of helping me to write up my thesis based on work completed during my co-op. I really appreciate his guidance.

I would also like to thank my Honors Adviser, Dr. Butler for his assistance in the thesis writing process and for providing feedback. Lastly, I would like to thank Dr. Slattery for reading over my thesis and offering suggestions

Chapter 1 : Introduction

1.1 Encoded Library Technology

Encoded Library Technology (ELT) is a method for the synthesis and selection of DNA-encoded libraries in which small compounds are attached to an encoding oligonucleotide. The DNA essentially serves as a bar code for the attached compound. The selection process to determine compounds that bind to the protein target of interest is performed by first mixing the DNA- encoded library and the target protein. The mixture is then applied to an affinity resin to capture the target protein. The unbound DNA-encoded compounds are washed away and the bound library molecules are eluted. Sequencing the remaining DNA identifies the DNA-encoded small molecules that were bound to the target protein. This technology represents a significant advancement in the ability to create and screen large numbers of compounds in order to efficiently discover drug candidates [1].

1.2 Proximity Ligation

ELT has been most successful with soluble proteins in purified form. However, it has only been minimally successful with unpurified proteins. There is a need for a method that can detect proteins in their native state that can be modified to incorporate ELT to produce an assay that can quantify protein-compound interactions. A recently developed mechanism, proximity ligation, can sensitively detect proteins in solution, cell culture, or localized tissue. Proximity ligation will be modified to incorporate ELT to quantify protein-compound interactions.

The method of proximity ligation depends on the simultaneous and proximate binding of target proteins by pairs of probes that have a specific affinity for the target [4]. The reactions create an amplifiable DNA molecule, which can be analyzed using Polymerase Chain Reaction (PCR) to identify the detected the protein target [2]. This procedure can be viewed as a reverse translation reaction because the detection of specific proteins can be identified from DNA sequencing. Proximity ligation is an extremely sensitive and specific method to detect proteins at a wide range of concentrations [5].

1.2.1 Proximity Ligation Mechanism of Action

Target proteins are analyzed using two proximity probes, each composed of a component capable of specifically binding to the target protein attached to an oligonucleotide through a biotin-streptavidin linkage [5]. The oligonucleotide presents either a 5’ or 3’ end. When the pair of probes binds to the target, the free ends of the oligonucleotides are brought into close proximity which is shown in step 1 of Figure 1-1 [4]. They are hybridized together when a connector oligonucleotide anneals to the complementary sequences of the oligonucleotides. This process is shown in step 2 of Figure 1-1. Enzymatic DNA ligase ligates the two oligonucleotides together. Probes not in proximity are not ligated and prevented from giving rise to the detectable signal. This results in a low nonspecific signal [5]. The ligation products are then replicated by nucleic acid amplification through PCR which is shown in step 3 of Figure 1-1. PCR allows for the quantification of the ligation products which reflect the number of target protein in the sample. The amplification cycle in which fluorescence in individual

reactions exceeds a threshold value compared to a control is used to calculate the concentration of target protein originally in the sample [2].

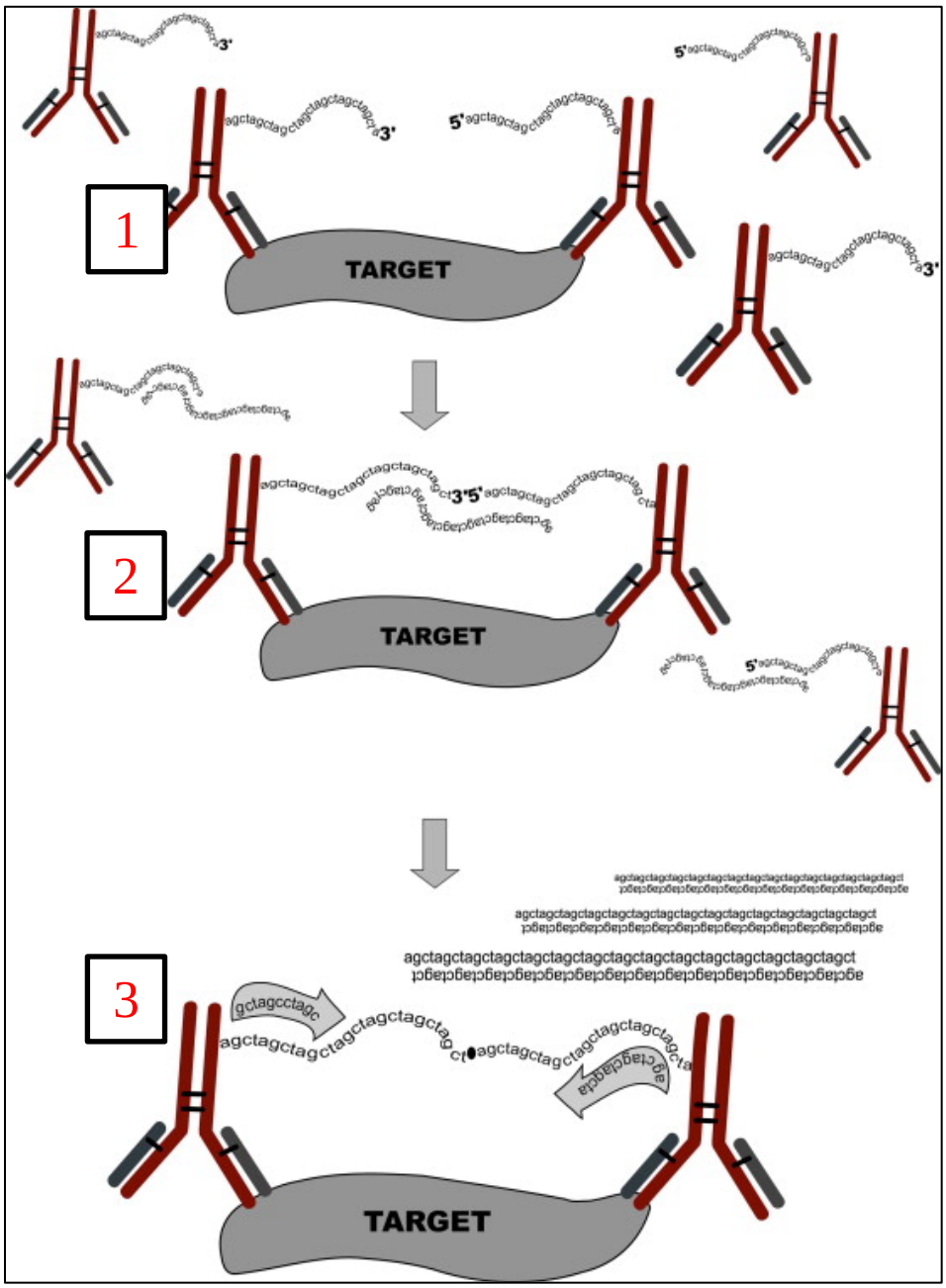

Figure 1-1: Proximity Ligation Mechanism.

(1) Probes bind to the target and the free ends of the oligonucleotides are brought into close proximity. (2) Connector oligonucleotide brings ends together to be joined by enzymatic DNA ligation (3) Ligation products replicated through PCR [2].

1.2.2 Optimization of Proximity Ligation Assays

In the early stages of development, proximity ligation assays were conducted using DNA aptamers, which are short oligonucleotide sequences, and a connector oligonucleotide complementary to the oligonucleotides of the probes. Three adjustments were made to the original assay to increase the signal over background.

Firstly, the aptamers were replaced with antibodies as the target binding part of the probe. Aptamers selected against purified proteins often bound to only one epitope on the target which limited effectiveness of the assay. Oligonucleotide-conjugated monoclonal and polyclonal antibodies have proven to be more successful. Currently, standard sets of streptavidin-oligonucleotide conjugates can be combined with biotinylated monoclonal polyclonal antibodies and used for detection without further purification [2].

Secondly, the DNA extensions of the affinity probes were no longer restricted to 40 nucleotides in length which was previously thought to be optimal. It was determined that the length of the extensions could be varied over a considerable range with negligible effects on ligation efficiency and the nonspecific signal. Therefore, large proteins can be detected using longer extensions of the affinity probes [4].

Lastly, a short hybridizing length at the 5’ and 3’ end sequences of the connector oligonucleotide was added. These additional end sequences do not base-pair to the proximity probes. This prevents the connector oligonucleotide from giving rise to ligation independent amplification [4].

1.2.2 Advantages of Proximity Ligation

Proximity ligation is a simple, quick, and sensitive method to quantify protein concentration in solution, cell culture, or localized tissue. The low background signal due to the infrequent joining of detection probes in the absence of specific target proteins and the efficient PCR amplification of DNA molecules combine to provide high assay sensitivity [2]. Also, coincident recognition by two binding reagents ensures a high degree of specificity. Proximity ligation assays have minimal steps and do not require washing [5]. Proximity ligation can be adapted to different assay formats depending on the focus of study.

Compared to alternative protein detection methods such as Western Blot, proximity ligation assays require a significantly lower amount of sample, are more sensitive, and have a faster workflow. The typical cell input for Western Blot is 100,000 cells/lane compared to proximity ligation assays with only 100-500 cells/well. Proximity ligation assays are highly sensitive with a limit detection of 10-35 cells/well, whereas Western Blot has a limited detection limit of 30,000-100,000 cells/lane. The output for proximity ligation is a numerical value with relative quantification. Western Blot output is an image with only a positive or negative quantification [6]. The results only take 3.5 hours to obtain for proximity ligation, but as long as 1-2 days for Western Blot. Overall, Western blotting is cumbersome, laborious, and much less quantitative than proximity ligation [7].

Proximity ligation assays allow for extremely sensitive and specific detection of proteins in only 1-µL sample volumes, thereby reducing sample amount by a factor of 1,000 compared with regular enzyme- linked immunosorbent assays (ELISAs) and by a

factor of 100 compared to Western blot [8]. The comparison of the limit of detection (LOD) for proximity ligation, ELISA, and Western blot is shown in Figure 1-2. It is evident that proximity ligation is much more sensitive than the other methods and is able to detect a lower concentration of target [2].

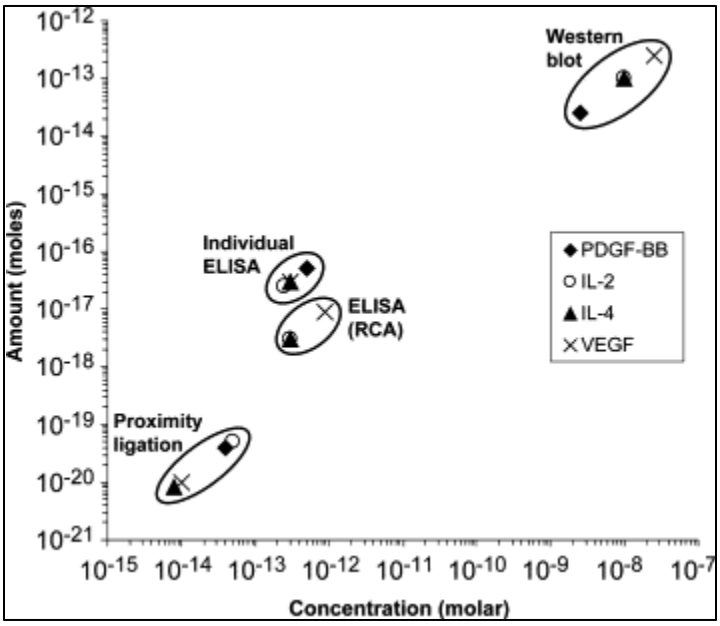

Figure 1-2: Limit of detection (LOD) for proximity ligation.

The x-axis shows the limit of detection measured as concentrations of target protein and the y- axis denotes amounts of target detectable. Proximity ligation is much more sensitive than the other methods and is able to detect a lower concentration of target [2].

1.3 TaqMan Protein Expression Assays

TaqMan Protein Expression Assays are an adapted form of proximity ligation and enable detection and relative quantification of protein targets in samples using proximity ligation in combination with real-time quantitative Polymerase Chain Reaction (qPCR) [7]. Applied Biosystems manufactures a kit for the assay that can be modified for experimental needs. It includes the reagents required for preparing probes, performing ligation, treating with protease, and amplifying with qPCR [9].

The probes are target specific antibodies that are conjugated to oligonucleotides through a biotin-streptavidin linkage. When the pair of probes binds to the target, the free ends of the oligonucleotides are brought into close proximity which is shown in step 1 of Figure 1-3 [7]. They are hybridized when a connector oligonucleotide anneals to the ends and DNA ligase ligates the bond which is shown in step 2 of Figure 1-3. The samples are then amplified by a quantitative real-time Lightcycler shown in step 3 of Figure 1-3.

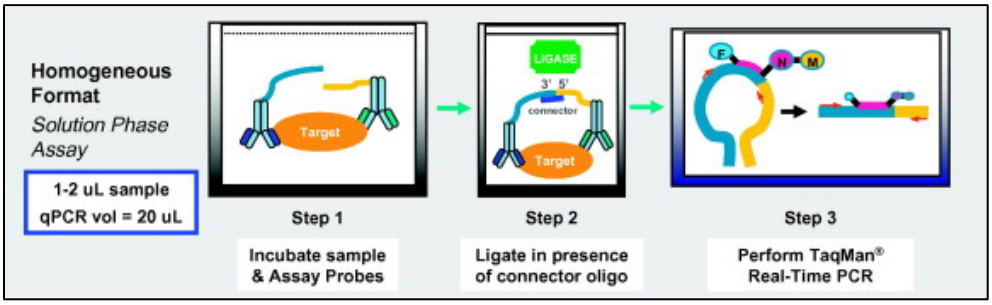

Figure 1-3: Process of TaqMan Protein Assay.

(1) binding of the antibodies to two separate epitopes of the target protein (2) ligation of the oligonucleotides (3) and amplification of the DNA sequence with a TaqMan probe using qPCR [7].

Real-time PCR monitors the progress of amplification of DNA as it occurs. A TaqMan probe recognizes the specific sequence of DNA in which the connector originally annealed to. The TaqMan probe is constructed with a fluorescent reporter dye bound to the 5' end and a quencher on the 3' end. While the probe is intact, the proximity of the quencher to the reporter dye greatly reduces the fluorescence emitted by the reporter dye. If the target sequence is present, the probe anneals between the primer sites and is cleaved by DNA polymerase. This cleavage separates the reporter dye from the quencher and produces a fluorescent signal. This process is shown in Figure 1-4. Real-

time PCR continues to replicate the template strand each cycle and additional reporter dye molecules are cleaved from their respective probes adding to the fluorescence signal.

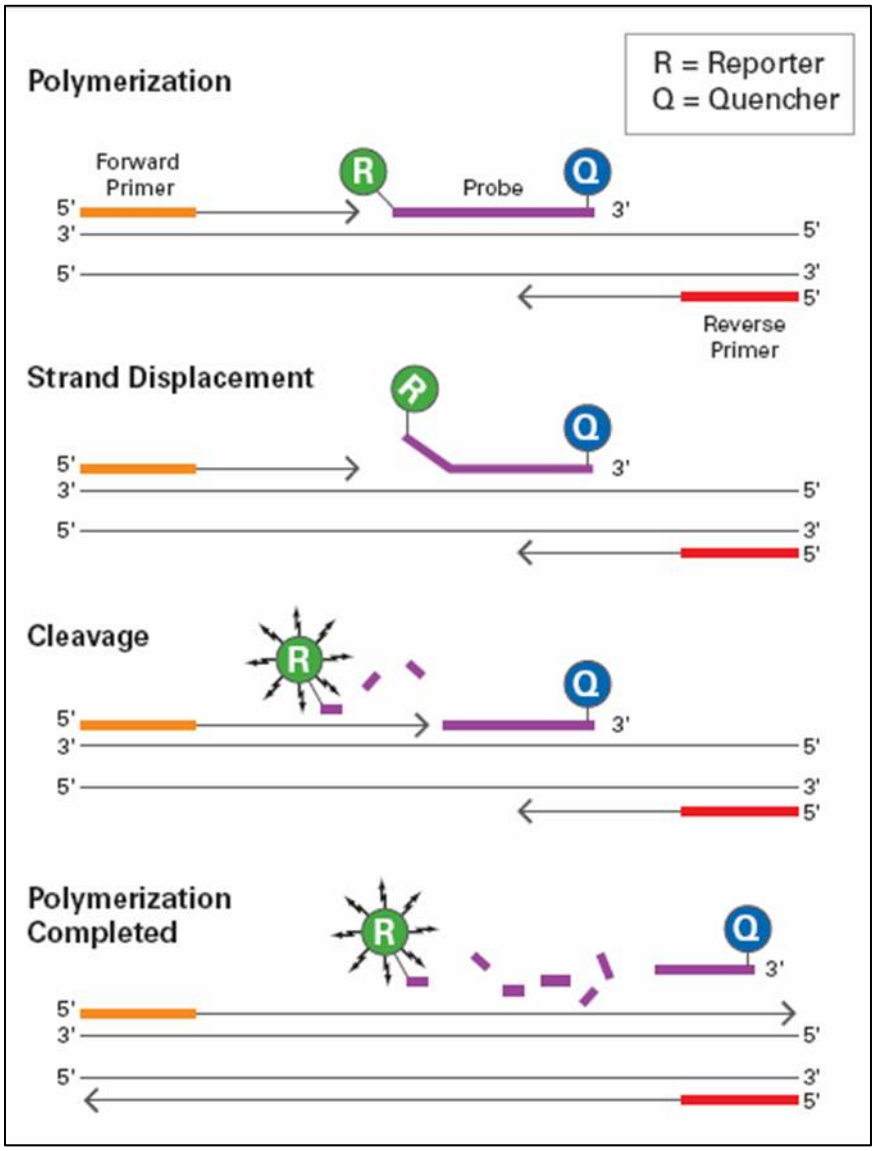

Figure 1-4: TaqMan probe fluorescence.

The TaqMan probe anneals to the DNA. DNA polymerase cleaves the reporter dye from the probe which then emits its characteristic fluorescence [10].

An amplification plot, shown in Figure 1-5, graphically displays the fluorescence detected over 40 cycles of real-time PCR. In the initial cycles of PCR, there is no

significant change in the fluorescence signal. This predefined range of PCR cycles is called the “baseline”. An algorithm identifies the point on the amplification plot at which the ΔRn value crosses a predetermined threshold value. The PCR cycle at which this occurs is defined as the crossing point (C_p or C_T).

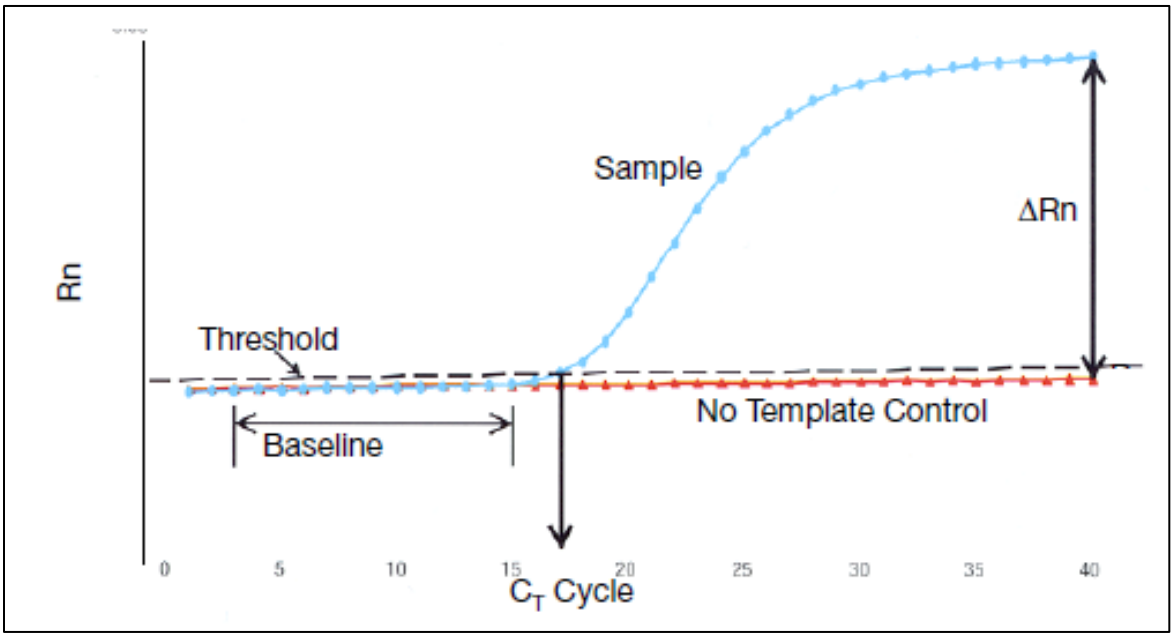

Figure 1-5: qPCR amplification plot.

The reactions are characterized by the PCR cycle, C_T, at which the fluorescence reaches a threshold value [10].

The reactions are characterized by the PCR cycle at C_P. The higher the starting target protein concentration, the earlier in the reaction the target amplification is detected and the lower the C_P value [11]. This is graphically shown in Figure 1-6.

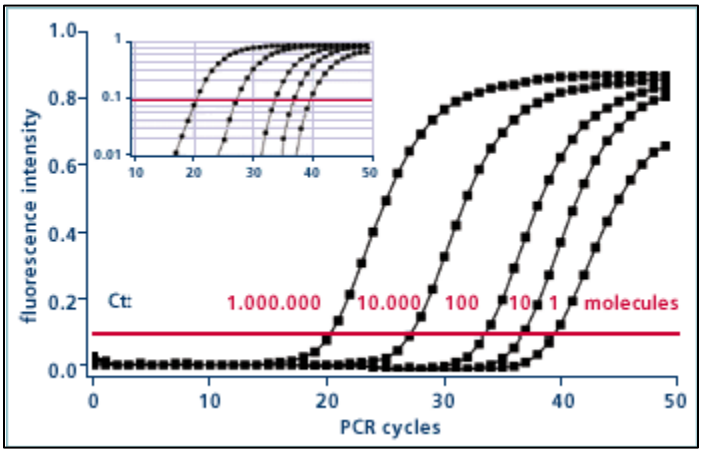

Figure 1-6: Real-time PCR of different target concentrations.

The higher the starting template concentration, the earlier in the reaction the target amplification is detected and the lower the C_p value [12]. The sample with 1,000,000 molecules has a C_p of 20 whereas the sample with 1 molecule has a C_p of 40.

Relative quantification is used to analyze the data and describes the change in expression of the target relative to a control. ΔC_p is the fractional PCR cycle at which the ΔRn value crosses the threshold relative to a no-protein control provides comparison between the initial levels of target in each sample. This is shown mathematically by Equation 1.

$$\Delta C_p(x) = C_p(NPC) - C_p(x) \quad (1)$$

C_p (x) is the PCR cycle that crosses the threshold for a sample containing protein. C_p (NPC) is the PCR cycle that crosses the threshold for a no-protein control sample. The amount of target, normalized to a control is calculated using Equation 2 [10].

$$[\text{Target}] = 2^{\Delta Cp} \quad (2)$$

Using Equations 1 and 2, graphs can be generated to determine the effect of target protein concentration on the detectable signal.

1.4 Curve Fitting

The data from the assay were fit with a nonlinear regression, one -site binding model to test if the model could be used to characterize the experiment. The binding experiments for a ligand and receptor are based on a simple model, called the law of mass action. The law of mass action is based on the assumptions that all receptors are equally accessible to ligands and are either free or bound to ligand. The model ignores any partial binding and any alteration to the ligand or receptor due to binding. The ligand binds to the receptor to form a Ligand-Receptor complex. This is shown in Equation 3.

$$\text{Ligand} + \text{Receptor} \Longleftrightarrow \text{Ligand} \bullet \text{Recepter} \quad (3)$$

In this model, binding occurs when the ligand and receptor collide with the correct orientation and enough energy. The binding of a ligand to a receptor is reversible. The rate of association is the number of binding events per unit of time and is expressed mathematically in Equation 4,

$$\text{Rate Association} = k_{on}\,[\text{Ligand}]\,[\text{Receptor}] \quad (4)$$

where, k_{on} is the association rate constant, [Ligand] is the concentration of ligand, and [Receptor] is the concentration of unbound receptors. Once binding has occurred, the ligand and receptor remain bound together for certain amount of time. This is a function of the affinity of the receptor and ligand for one another. The rate of dissociation is number of dissociation events per unit time and is expressed mathematically in Equation 5.

$$\text{Rate Dissociation} = [\text{Ligand} \bullet \text{Receptor}]\, k_{off} \tag{5}$$

where, k_{off} is the dissociation rate constant and [Ligand • Receptor] is the concentration of ligand bound to receptor. Equilibrium is reached when the rate at which new [Ligand • Receptor] complexes are formed equals the rate at which the [Ligand • Receptor] complexes dissociate. This is shown in Equation 6.

$$[\text{Ligand}][\text{Receptor}]k_{on} = [\text{Ligand} \bullet \text{Receptor}]k_{off} \tag{6}$$

The equilibrium dissociation constant K_d is shown in Equation 7.

$$\frac{[\text{Ligand}][\text{Receptor}]}{[\text{Ligand} \bullet \text{Receptor}]} = \frac{k_{off}}{k_{on}} = K_d \tag{7}$$

The interaction between the ligand and the receptor involves a complex combination of hydrophobic, van der Waals, hydrogen-bond, and electrostatic interactions. Increased affinity generally corresponds to an increased specificity as a high affinity reflects good

complementarity between the ligand and receptor [5]. A small K_d is due to a high concentration of $[\text{Ligand} \bullet \text{Receptor}]$ complexes and is an indication the ligand has a high affinity for the receptor. In contrast, a large K_d is due to a low concentration of $[\text{Ligand} \bullet \text{Receptor}]$ complexes and is an indication the ligand has a low affinity for the receptor.

The law of mass action predicts the fractional receptor occupancy at equilibrium as a function of ligand concentration. Fractional occupancy, Equation 8, is the fraction of all receptors that are bound to ligand.

$$\text{Fractional Occupancy} = \frac{[\text{Ligand} \bullet \text{Receptor}]}{[\text{Receptor}]_{\text{total}}} = \frac{[\text{Ligand} \bullet \text{Receptor}]}{[\text{Receptor}] + [\text{Ligand} \bullet \text{Receptor}]} \tag{8}$$

Multiplying the numerator and denominator by [Ligand], dividing both by$[\text{Ligand} \bullet \text{Receptor}]$, and substituting the definition of K_d yields Equation 9.

$$\text{Fractional Occupancy} = \frac{[\text{Ligand}]}{[\text{Ligand}] + K_d} \tag{9}$$

When [Ligand] =0, the occupancy equals zero. When [Ligand] is very high (many times K_d) the fractional occupancy approaches 1.00. When [Ligand] =K_d the fractional occupancy is 0.50. The fractional occupancy multiplied by the receptor concentration determines the $[\text{Ligand} \bullet \text{Receptor}]$ concentration.

1.5 BacMam Transduction

To test the proximity ligation assay for detection of protein-compound interactions in cell lysate, BacMam virus gene delivery technology was used. The development of modified baculoviruses carrying mammalian cell-active regulatory elements (BacMam viruses) is used for gene delivery and protein expression in mammalian cells. The system is highly versatile and can be used for multicomponent target proteins, different expression levels, and multiple cell line hosts. The target gene sequences are first cloned into a transfer plasmid containing a mammalian cell-active element and are transferred to baculovirus DNA via recombination. The viral DNA is transfected into insect cells where virus production occurs. The stock virus is used to transduce mammalian cells and expression of the recombinant protein is usually validated within 24-48 hours [13].

Chapter 2 : Methods and Materials

2.1 Proximity Ligation Assay Design

The Applied Biosystems TaqMan Protein Assay kit was modified to quantify the interactions between proteins and compounds. The modified design is shown in detail in Figure 2-1. In the design, one antibody probe was removed and replaced with a biotinylated (biotin shown as purple circles in Figure 2-1) tagged DNA-encoded compound that has a high affinity for the target. The experimentally derived K_d was previously determined. Both the antibody and the DNA-encoded compound were biotinylated and linked to streptavidin (denoted SA in Figure 2-1) conjugated oligonucleotides. An anti-Flag antibody was chosen which recognized the Flag portion of the tagged protein. The ligation connector was complementary to the ends of the oligonucleotides. The qPCR primers recognized specific sequences of the oligonucleotides and served as the starting point for DNA amplification.

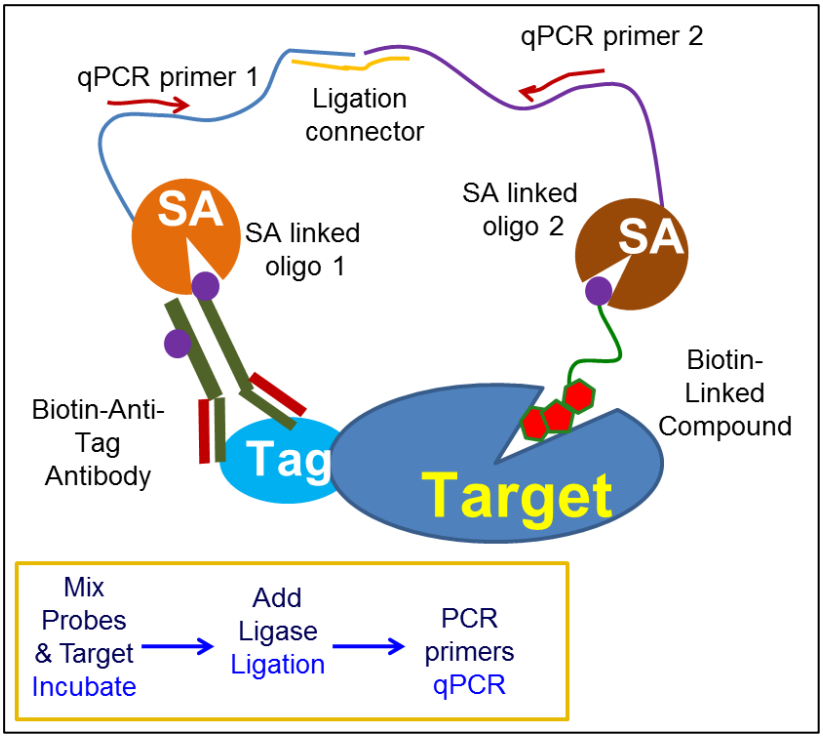

Figure 2-1: Proximity ligation design with biotinylated DNA-encoded compound.

Both the antibody and the DNA-encoded compound were biotinylated and linked to streptavidin (SA) conjugated oligonucleotides. The anti-Flag antibody recognized the Flag portion of the tagged protein. The ligation connector was complementary to the ends of the oligonucleotides. The qPCR primers recognized specific sequences of the oligonucleotides and served as the starting point for DNA amplification.

The modified design was tested in purified Flag-RIPK2 and Flag-TNKS1 proteins and cell lysate of Flag-RIPK2 BacMam transduced HEK MSRII cells. After qPCR was performed, the amplified DNA samples were evaluated with DNA gel electrophoresis to verify the product. The data were fit to a nonlinear regression, one- site specific binding model to determine the K_d of the experiment. These values were compared to the known K_d values. A theoretical model, with an alternative method of analyzing the binding experiment, was generated to compare the results of the assays performed.

2.2 TaqMan Protein Assay.

The methods used were adapted from Applied Biosystems TaqMan Protein Assay method section [11].

2.2.1 Preparing Assay Probes.

The Antibody Dilution Buffer (Applied Biosystems) was placed on ice and thawed. The tube was inverted several times and centrifuged at 1,500 x g for 1 minute. Using the Antibody Dilution Buffer, the biotinylated flag- antibody was diluted to 200nM. The corresponding biotinylated tagged DNA-encoded compound, depending on the protein used (Flag-RIPK2 or Flag-TNKS1), was also diluted to 200nM. Both solutions were mixed by touch-vortex and then spun in a centrifuge at 1,500 x g for 1 minute. The reaction tubes were then returned to ice. The 5' and 3' Prox-oligos (Applied Biosystems) were removed from the 4°C refrigerator, inverted several times, and then centrifuged at 1,500 x g for 1 minute. One 1.5 mL reaction tube was labeled Assay Probe A and another Assay Probe B. To the Assay Probe A tube, 5 µL of diluted biotinylated antibody (200nM) and 5µL of 5' Prox-Oligo (200nM) were added. To the Assay Probe B tube, 5 µL of diluted biotinylated tagged DNA-encoded compound (200nM) and 5 µL of 3' Prox-Oligo (200nM) were added. Both tubes were mixed gently by touch-vortex and then centrifuged at 1,500 x g for 1 minute. The tubes were incubated at room temperature for 60 minutes. The process of preparing the assay probes is shown in Figure 2-2.

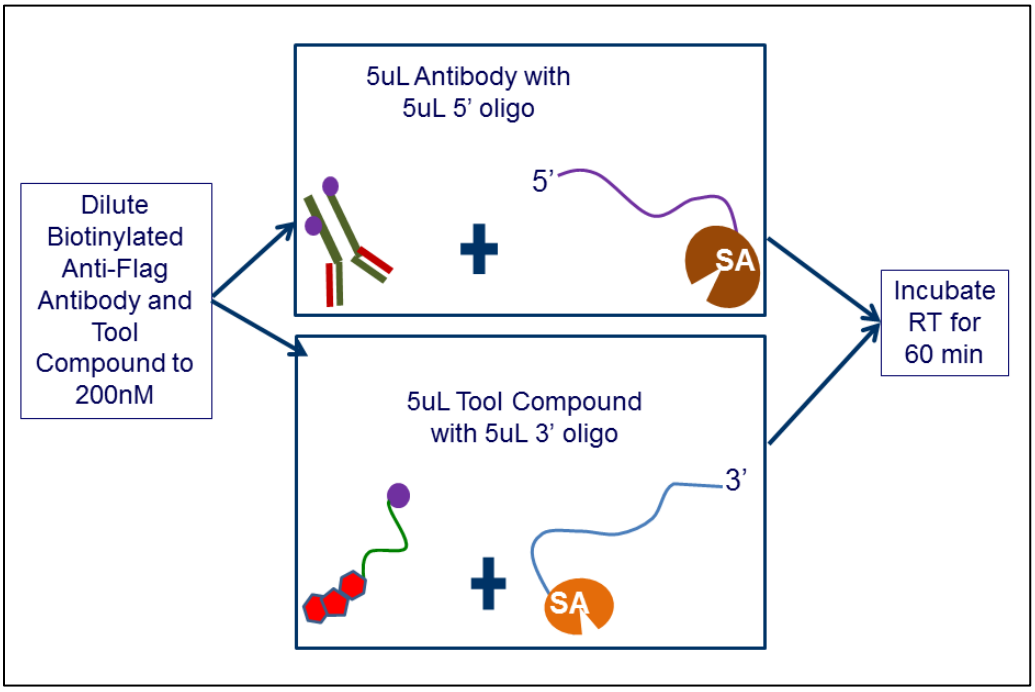

Figure 2-2: Preparation of Assay Probes.

The biotinylated anti-flag antibody and the DNA-encoded compound were diluted to 200 nM. Each was incubated with the corresponding streptavidin linked oligonucleotide. This were then incubated together to form the final product.

The Assay Probe Storage Buffer (Applied Biosystems) was removed from the -20°C freezer and placed on ice for 5 minutes. The tube was inverted several times and then centrifuged at 1,500 x g for 1 minute. To both Assay Probe tubes, 90 µL of the Assay Probe Storage Buffer was added. Both solutions were mixed by touch vortex and then centrifuged at 1,500 x g for 1 minute. The tubes were incubated at room temperature for 20 minutes. The assay Probe Dilution Buffer was removed from the -20°C freezer, placed on ice, allowed to fully thaw, and mixed by touch vortex. The Assay Probe A tube, Assay Probe B tube, and Probe Dilution Buffer bottles were flicked to mix the solutions. All tubes were centrifuged at 1,500 x g for 1 minute. The assay probes were combined in one tube and diluted to 0.25 nM using Assay Probe Dilution Buffer.

The reaction tube was then placed on ice. The Lysate Dilution Buffer (Applied Biosystems) was removed from the -20°C freezer and placed on ice to thaw. The tube was inverted several times and then briefly centrifuged at 1,500 x g for 1 minute. The protein (Flag-RIPK2 or Flag-TNKS1) was diluted using the Lysate Dilution Buffer to a determined concentration. A 1:2 serial dilution of the protein to Lysate Dilution Buffer was then prepared in PCR tubes for seven points. A no protein control was added and served as the eighth data point. Each point had triplicated measurements of dilutions.

To 2 µL of each protein dilution, 2 µL of the Assay Probe A and B mixture was added. The tubes were incubated at 37°C in a thermal cycler for 1 hour. After this, the tubes were centrifuge at 1,500 x g for 1 minute to mitigate the loss of sample when the tubes were opened. The Ligase Dilution Buffer 1X (Applied Biosystems) was briefly placed on a vortex and then centrifuged at 1,500 x g. The DNA Ligase 500X (Applied Biosystems) was removed from the -20°C freezer and inverted to mix the solution, and then centrifuged at 1,500 x g. The ligase was diluted to a 1X solution using the 1X Ligase Dilution Buffer (Applied Biosystems) immediately before use. The Ligation Reaction Buffer (Applied Biosystems) was thawed on ice, briefly mixed using a vortex, and then centrifuged at 1,500xg for 1 minute. In a 1 mL reaction tube, the Diluted Ligase 1X was diluted 100 times using deionized water.

2.2.2 Ligation Reaction

The PCR tubes were incubated in a thermal cycler at 37°C for 10 minutes. The 1 X PBS (Applied Biosystems) was thawed on ice, briefly placed on a vortex, and then centrifuged at 1,500 x g for 1 minute. To a reaction tube, a 1:100 dilution of Protease (Applied Biosystems) to PBS was added. To each PCR tube, 2 µL of diluted protease was

added. The PCR tubes were incubated at 37°C in a thermal cycler for another 10 minutes followed by 5 minutes at 95°C to inactivate the ligase.

2.2.3 qPCR Amplification

To prepare the PCR reaction, the Universal PCR Assay (Applied Biosystems), containing nucleotides, primers, and TaqMan probes, was thawed on ice and the Fast Master Mix (Applied Biosystems) was obtained from the 4°C refrigerator. To a reaction tube, 100 µL of Fast Master Mix and 10 µL of Universal PCR Assay were added. To each well on a PCR plate, 11 µL of PCR mixture and 9 µL of protease treated sample were added. The Lightcycler was programmed to hold for 20 seconds at 95°C for enzyme activation, 40 cycles of 1 second at 95°C for denature and 20 seconds at 60°C for the primers to anneal and polymerase to add the nucleotides. The process flow from protein to titration to qPCR is shown in Figure 2-3.

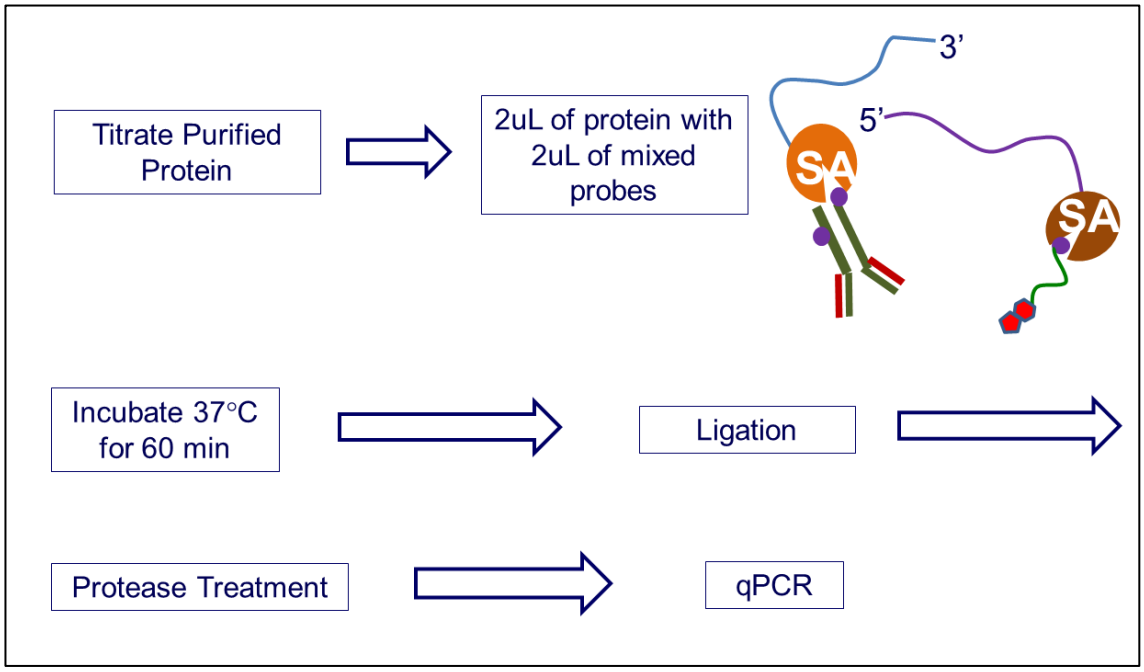

Figure 2-3: Process of Proximity Ligation Assay.

The protein was titrated, mixed with probes, and incubated. Ligation reaction was performed followed by protease treatment to inactivate the ligase. qPCR was performed to amplify the DNA sequence.

2.3 Optimization of Probe Concentration

The optimal probe concentration was determined by performing the TaqMan Protein Assay as described in Section 2.2. However, the DNA-encoded compound probe was serially titrated with a constant concentration of protein. The starting concentration of the DNA-encoded compound probe was 0.250 nM and serially titrated 1:3 in Assay Probe Dilution Buffer (Applied Biosystems) for 6 points with a constant 8.2 nM of Flag-TNKS1 and constant concentration of 0.250 nM of anti-Flag antibody probe. A no protein control was included as the seventh data point. Triplicate measurements were performed for each point to obtain a reliable average.

2.4 TaqMan Assay with BacMam Transduced Samples

HEK MSRII cells were obtained and thawed in a 37°C water bath. The cells were placed in 25mL of DMEM HAMS- F12 +10% Aust FBS (M1) media (Invitrogen). 1 mL was removed and the cell concentration was obtained using a Cedex Cell Counter. The cells were then diluted to $1x10^6$ cells/well in DMEM HAMS- F12 +10% Aust FBS (M1) media. Varied concentrations of a BacMam virus containing the gene for the protein of interest were added to separate tubes with HEK MSRII cells. 3 mL of each mixture of cells and varying concentrations of BacMam virus were added to separate wells of a 6 well plate. The plates were then incubated for 24 hours in a Humidified CO_2 Incubator. After 24 hours, the media was aspirated from the wells. One tablet of protease inhibitor (Roche) was dissolved in10 mL of Lysis buffer (Roche) and kept on ice. 200 µl of the mixture of Lysis buffer (Applied Biosystems) with protease inhibitor was added to each well. The plate was placed on ice for 10 minutes. A 1:2 dilution of cells with the added lysis buffer to lysate buffer was made. The procedure for TaqMan assay was then performed as detailed in Section 2.2 [11].

2.5 Gel Electrophoresis

2% E-Gel pre-cast agarose gels (Invitrogen) were used to perform gel electrophoresis on the PCR samples. The E-Gel agarose gel loading buffer (Invitrogen) was diluted 1:20 in water (Invitrogen). 1µL of the diluted loading buffer was added to each 10 µL sample after PCR. 1 µL of the E-Gel 1 Kb Plus Ladder (Invitrogen) was diluted 1:10 with water and added to lane one of the pre-cast agarose gel. The PCR samples with loading buffer were then added to the lanes. The gel was loaded into the E-

Gel iBase and run for 26 minutes. After completion, the gel was viewed under a UV light [14].

2.6 Nonlinear Regression One-Site Binding

The law of mass action was applied to the proximity ligation experiments. In the assay performed, the target protein was the receptor and the DNA-encoded compound was the ligand. The probe concentration was set at the optimal 0.250 nM and the protein target concentration was varied. Equilibrium specific binding at a particular protein target concentration is equal to the fraction occupancy times the total number of probes shown in Equation 10.

$$\text{Specific Binding} = \text{Fractional Occupancy} \bullet B_{max} = B_{max} \bullet \frac{[T]}{K_d + [T]} \tag{10}$$

This equation describes a rectangular hyperbola. [T] is the concentration of target protein plotted on the x-axis. B_{max} is the maximum number of probes and K_d is the equilibrium dissociation constant. The data were fit with a nonlinear regression, one-site specific binding model. Analysis of the data was performed using GraphPad Prism to determine the experimental K_d [15].

2.7 Theoretical Model

In a previous proximity ligation assay conducted by Gullberg et al [3], a theoretical model was generated to quantify the experiment with two antibody probes. The model was adjusted to produce a new model that to simulates the TaqMan proximity

ligation assay for detection of protein-compound interactions. The concentration of probe bound to the target protein [PT] for both probes was determined by Equation 11. The B_{max} is the total number of probes, K_d is the dissociation constant, and [T] is the target concentration.

$$[PT] = Bmax\left(1 + \frac{K_d}{[T]}\right) \quad (11)$$

This value was divided by the total number of probes to determine the target fraction bound by each probe. By multiplying the fraction of target binding site 1 that is bound by probe 1 with the fraction of site 2 bound by probe two, the fraction of target proteins simultaneously bound by the two proximity probes can be determined. This value was multiplied by the total target concentration to determine the complex concentration with both probes bound.

Assuming the K_d for the biotin-linked antibody was 0.1 nM and the concentration of both probes was the optimal concentration of 0.250 nM, a graph was generated for different dissociation constants of the DNA-encoded compound. The target concentration was plotted on the x- axis and the detectable target concentration on the y- axis.

The relative quantification was determined using Equation 12 and assuming the qPCR detection limit was 1 pM.

$$RQ = \text{Detectable [Target]}/\text{[Detection limit]} = 2^{\Delta Cp} \quad (12)$$

The two graphs were evaluated to determine the effect of K_d on the assay efficiency.

Chapter 3 : Results and Discussion

A probe titration experiment was performed to determine the optimal DNA-encoded compound probe concentration to reduce target-independent ligation. Once this was determined, the proximity ligation assay was performed with purified proteins and cell lysate and proved to be dependent on the target protein concentration. The experimental K_d was determined for purified Flag-TNKS1 and Flag-RIPK2 by fitting the data with a nonlinear regression, one-site specific binding model. These values did not match the known K_d values. This suggested that the one-site binding model does not characterize the experiment and the assumption that the binding is only dependent on the affinity of the DNA-encoded compound probe was incorrect. Gel electrophoresis was performed on the purified Flag-TNKS1 qPCR samples and determined the correct sequence was amplified. The intensities of the bands were analyzed using ImageJ. Statistical analysis proved that the mean intensities of the bands were not statistically similar. This disproved the hypothesis that the fluorescence signal was the same for all concentrations of proteins at the end of qPCR. The probes and qPCR primers need to be redesigned so there is no longer a decrease in fluorescence with a decrease in protein.

A previous study conducted by Gullberg et al [3], compared experimental data to theoretical standard curves. This theoretical model was adapted and generated to quantify the interactions between protein targets and compounds based on the different affinities of the DNA-encoded compound for the target protein. The model characterized the binding experiment as two independent one-site binding models multiplied together. The model proved that the affinity of the probes strongly influences the sensitivity that can be attained in proximity ligation. The smaller the K_d, the lower the amount of protein target

needed to achieve the maximum response. The experimental data were compared to the model, but did not correspond to the maximum response target concentration or the maximum ΔC_p. The assay needs to be optimized in order to obtain the maximum target concentration and ΔC_p.

3.1 Probe Optimization

The DNA-encoded compound probe was titrated with constant concentrations of protein and antibody probes. The optimal probe concentration was determined. 8.2 nM of Flag-TNKS1 was selected as the protein concentration because in previous experiments it corresponded to the largest value of ΔC_p and therefore gave the strongest signal compared to the background. The TaqMan Protein Assay was then performed with these added adjustments. The results are shown in Figure 3-1 and Figure 3-2.

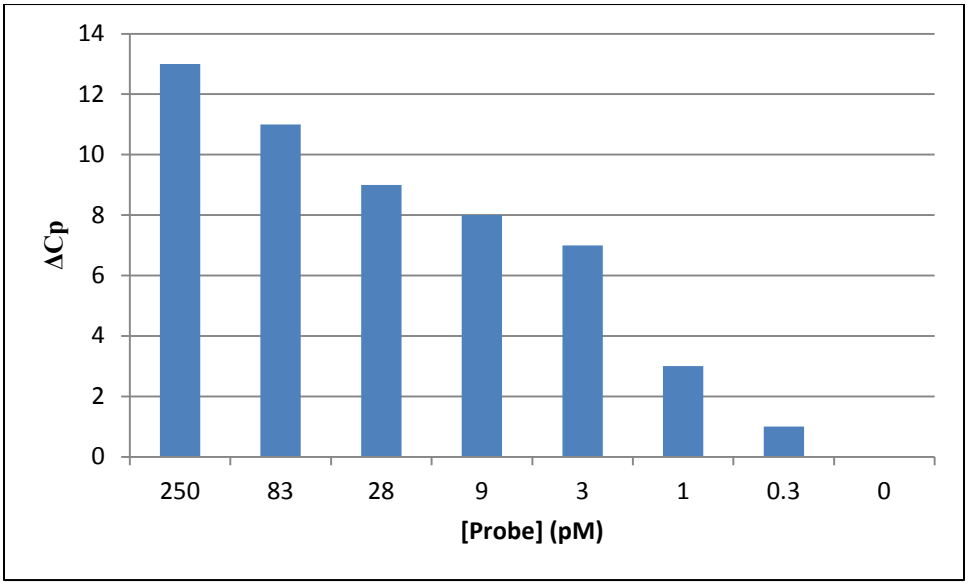

Figure 3-1: DNA-encoded compound probe titration.

The optimal probe titration was determined to be 250 pM due to the largest ΔC_p which indicated the strongest signal over background.

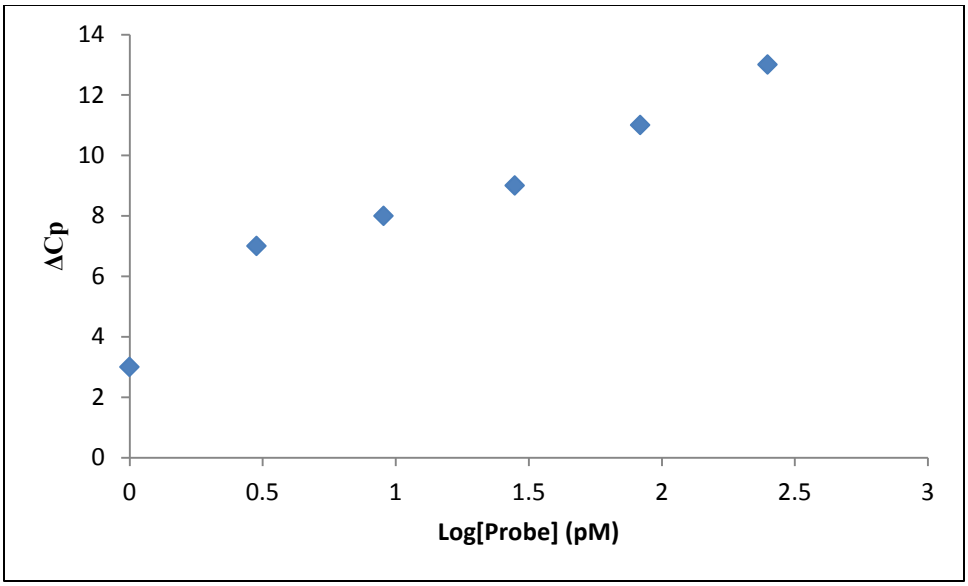

Figure 3-2: Linearization of DNA-encoded compound probe titration.

The optimal concentration of probe was determined to be 0.250 nM.

The optimal probe concentration was determined to be 250 pM or 0.250 nM. At this concentration of probe, the ΔC_p value was the largest indicating the strongest signal over background. A very low proximity probe concentration of 0.250 nM will reduce target-independent ligation. Since the K_d values of Flag-TNKS1 and Flag-RIPK2 are very similar, 15 nM and 10 nM respectively, the results with Flag-TNKS1 were used for proximity ligation assays with Flag-RIPK2 to conserve reagent.

3.2 Purified Flag-TNKS1 Titration

The assay was first peformed with purified Flag-TNKS1 protein. The protein was serially titrated and triplicate measurements of the dilutions were made in order to obtain a reliable average. The anti-Flag antibody probe recognized the Flag portion of the protein and the biotinylated tagged DNA- encoded compound had a high affinity for target protein with a known K_d of 15 nM. The standard deviation was determined for each point and included on the graph.

The data were fit with a nonlinear regression, one-site specific binding model. A one-site binding model was chosen because the antibody has a high affinity for the Flag tagged portion of the target protein (K_d of 0.1nM) and is assumed to not dissociate once associated. The K_d of the DNA-encoded compound probe however, is two magnitudes greater. The reaction is theorized to be limited by the affinity of the DNA-encoded compound for the target protein. Therefore, it was assumed that the reaction is only dependent on the binding of the compound to the target protein at only one site. The one-site specific binding model is specifically used because the nonspecific binding was subtracted out by determining ΔC_p.

The binding for a one-site specific binding model follows the law of mass action and assumes the reaction has equilibrated. It also assumes there is only one population of receptors, only a small fraction of the ligand binds so that the free concentration is essentially identical to the concentration added, and binding of a ligand to one binding site does not alter the affinity of another binding site.

The results of the titration of purified Flag-TNSK1 with standard error of the mean are shown in Figure 3-3. The data were linearized by graphing log[protein] vs. ΔC_p and is shown in Figure 3-4. From these two graphs, it is evident the reaction was dependent on the protein concentration. As the protein concentration increased, the earlier the qPCR cycle the detection was observed. Subtracting the cycle number for background noise, the ΔC_p was determined. Therefore, a higher concentration of target protein sample corresponded to a greater ΔC_p value. The assay was able to detect very low concentrations of protein on the level of 30 picomolar.

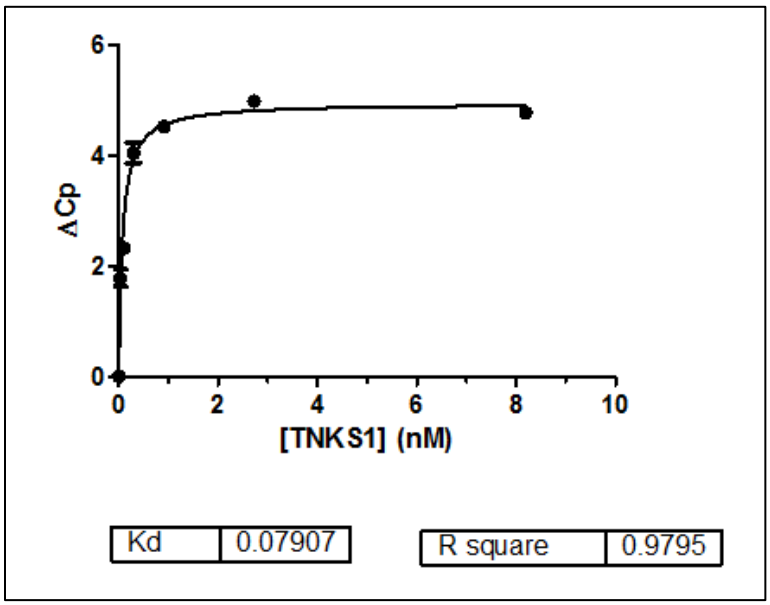

Figure 3-3: Flag-TNKS1 Protein Titration.

The K_d was determined to be 0.0791 nM with a 95% confidence interval of 0.038 to 0.120 nM.

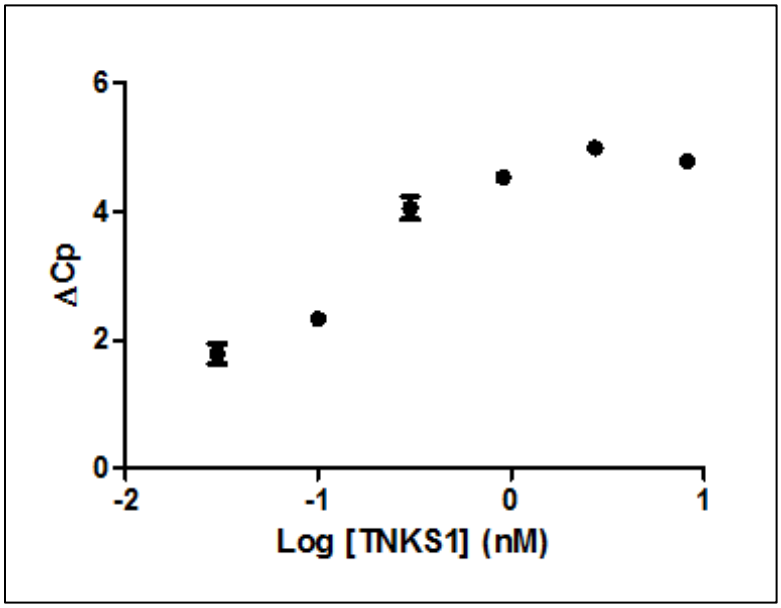

Figure 3-4: Linearization of Flag-TNKS1 Protein Titration.

The reaction is dependent on protein concentration. As the protein concentration increased, there was a corresponding increase in signal

The data were fit to the nonlinear regression, one-site specific binding model and the experimental K_d was determined to be 0.079 nM. The R square value, 0.9795, indicates a good fit. The 95% confidence interval for K_d was 0.038 to 0.120. The experimentally determined K_d was much lower than the actual value for the DNA-encoded compound for Flag-TNKS1 which is 15 nM. This is approximately 190 times greater of affinity than the actual value. Using the nonlinear regression, one-site specific binding model can not be used to accurately measure the K_d of the reaction. Therefore, another model needs to be generated to quantify the protein-compound interaction.

The qPCR samples were analyzed by gel electrophoresis which is shown in Figure 3-5. Lane 1 contained 8.2 nM of Flag-TNKS1 with a decreasing protein concentration through Lane 6. Lane 7 served as the negative control and contained only buffer. The amplified oligonucleotide sequence of interest was 100 base pairs. The bands from the amplified PCR product occurred at the expected 100 base pair marker. This confirmed our previous results that the correct sequence was amplified. The faint band below the 100 base pairs marker was due to the excess primers. The primers were much smaller than 100 base pairs. They were found at the expected lower location on the gel compared to the amplified DNA sequence. The primers were smaller and therefore could travel faster through the gel and reach a further distance from the loading lane.

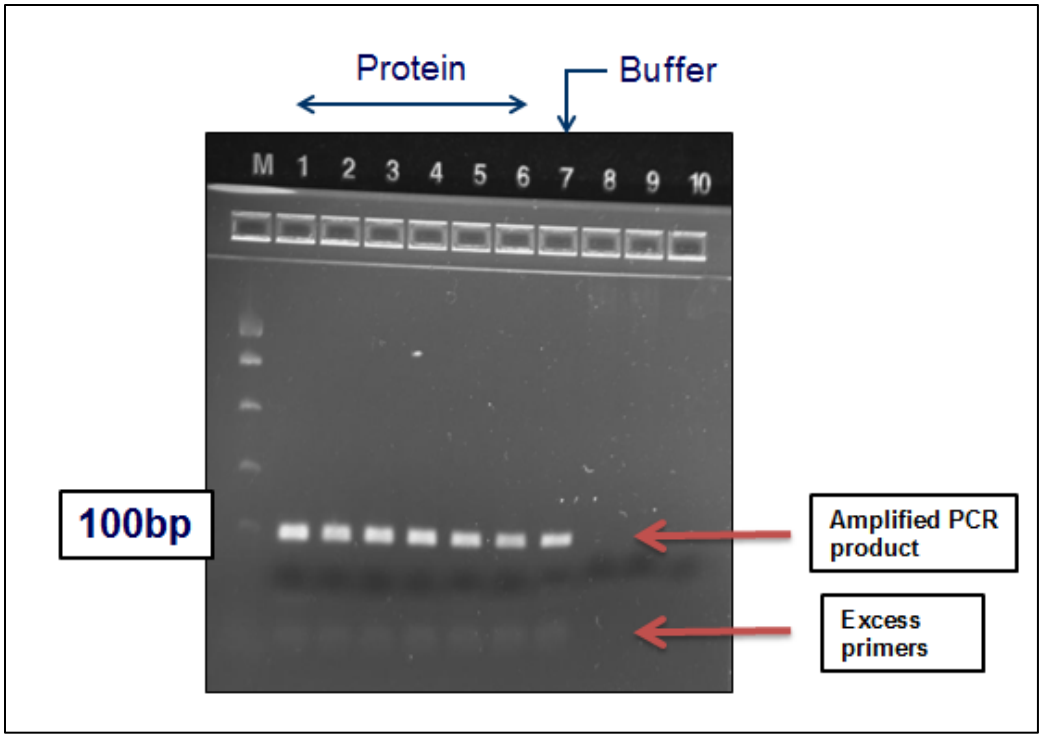

Figure 3-5: DNA Gel of qPCR Product.

The bands from the amplified PCR product occurred at the expected 100 base pair marker.

ImageJ was used to determine the intensities of the product bands. The intensities were expected to be relatively the same due to the fluorescence values achieving the same value at the end of qPCR as shown in Figure 1-6.

Table 3-1: ImageJ Results

Lane	Pixels	Mean	SD	Min	Max
1	152	207	28	116	235
2	187	181	29	112	221
3	144	197	25	121	231
4	190	186	35	99	230
5	133	192	24	120	223
6	119	179	16	128	203
7	144	185	30	91	225

Each band was evaluated separately. ImageJ recorded the intensity of the pixels. The sample size was the number of pixels evaluated. The mean, standard deviation, minimum, and maximum intensities were recorded in Table 3-1. The Leven's test was performed to assess the equality of variances of the different bands. The equation used is shown in Equation 13 [16].

$$W = \frac{(N-k)}{(k-1)} \frac{\sum_{i=1}^{k} N_i (Z_{i\cdot} - Z_{\cdot\cdot})^2}{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (Z_{ij} - Z_{i\cdot})^2}, \tag{13}$$

where, W is the result of the test, k is the number of different groups to which the samples belong, N is the total number of samples, N_i is the number of samples in the *i*th group, Y_{ij} is the value of the *j*th sample from the *i*th group,

$$Z_{ij} = \begin{cases} |Y_{ij} - \bar{Y}_{i\cdot}|, & \bar{Y}_{i\cdot} \text{ is a mean of i-th group} \\ |Y_{ij} - \tilde{Y}_{i\cdot}|, & \tilde{Y}_{i\cdot} \text{ is a median of i-th group} \end{cases}$$

$$Z_{\cdot\cdot} = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{N_i} Z_{ij} \text{ and}$$

$$Z_{i\cdot} = \frac{1}{N_i} \sum_{j=1}^{N_i} Z_{ij}$$

The calculation of W was modified because individual observations were unattainable from ImageJ. Therefore, it was assumed $(Z_{ij}\text{-}Z_i)^2 = \sigma (n\text{-}1)$ where, σ is the variance and n is the sample size. The modified W value was 0.000125 which was then tested against F(alpha, N-k, k-1). Alpha was 0.05, N-k was 1062, and k-1 was 6. The F value was 3.67. Since the W value was less than F, the null hypothesis was not rejected and all variances were assumed to be equal. Two sided t tests were performed against all

values. The null hypothesis was $\mu_1=\mu_2$ and the alternative hypothesis was $\mu_1\neq\mu_2$. The p values were all less than 0.05. The null hypothesis was rejected in favor of the alternative hypothesis and concluded there is a significant difference between the sample means. The intensities of each band were not the same. The fluorescence of each sample did not reach the same value at the completion of qPCR. There was a decrease in fluorescence with decreasing protein concentration. According to Applied Biosystems, a decrease in fluorescence is due to the interactions between the primers and the probes that are not compensated for by the baseline subtraction. The primer and probe sequences need to be evaluated and redesigned to ensure they are not complementary to each other [9].

3.3 Purified Flag-RIPK2 Titration

The assay was also peformed with purified Flag-RIPK2. The results of the titration are shown in Figure 3-6. The data were linearized by graphing ΔC_p vs. log[protein] and are shown in Figure 3-7. The reaction was also determined to be dependent on the protein concentration. A higher concentration of target protein sample corresponded to a greater ΔC_p. The assay detected a very low concentration of protein on the level of 20 picomolar.

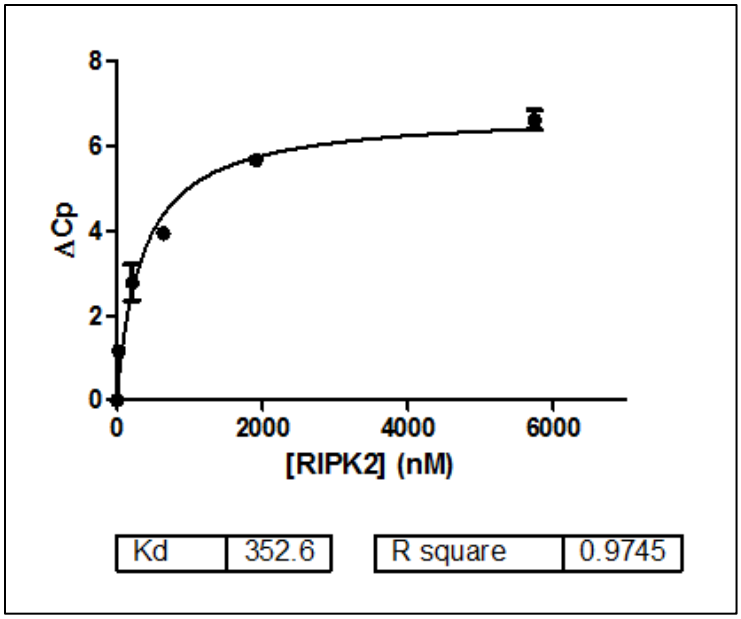

Figure 3-6: Flag- RIPK2 Protein Titration.

The K_d was determined to be 352.6 nM with a 95% confidence interval of 75.65 to 629.6 nM.

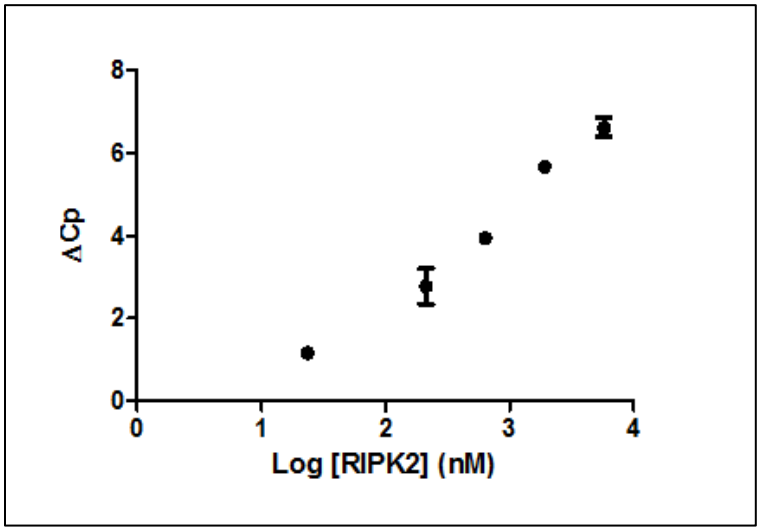

Figure 3-7: Linearization of Flag-RIPK2 Protein Titration.

The reaction is dependent on protein concentration. As the protein concentration increased , there was a corresponding increase in signal.

The data were fit with a nonlinear regression, one- site specific binding model. The standard error of the mean was determined for each of the triplicate measurements of dilutions and added to the graph. The data were fit to the model and the experimental K_d was determined to be 352.6 nM. The R square value, 0.9745, indicates a good fit. The 95% confidence interval for K_d was 75.65 to 629.6 indicating that the line was not a good fit. The experimentally determined K_d was much higher than the known value for the DNA- encoded compound for Flag-RIPK2 which is 10 nM. This is approximately 35 times less affinity than the true value. This confirmed the previous conclusion that the nonlinear regression one-site binding model is not applicable to this assay. Another model needs to be generated to fit the data.

3.4 HEK MSRII Transduced Cells with Flag-RIPK2

The assay was then tested with cell lysate. HEK MSRII cells were transduced with Flag-RIPK2 BacMam virus. The data were graphed with standard error of the mean in Figure 3-8.

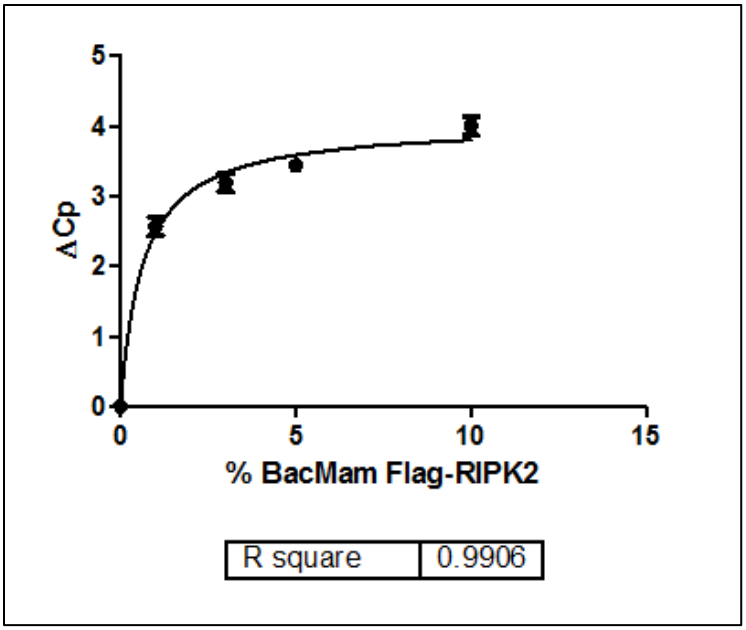

Figure 3-8: Flag-RIPK2 Protein Titration with BacMam transduced HEK MSRII cells.

Protein dependent signal with a R square value of 0.9906.

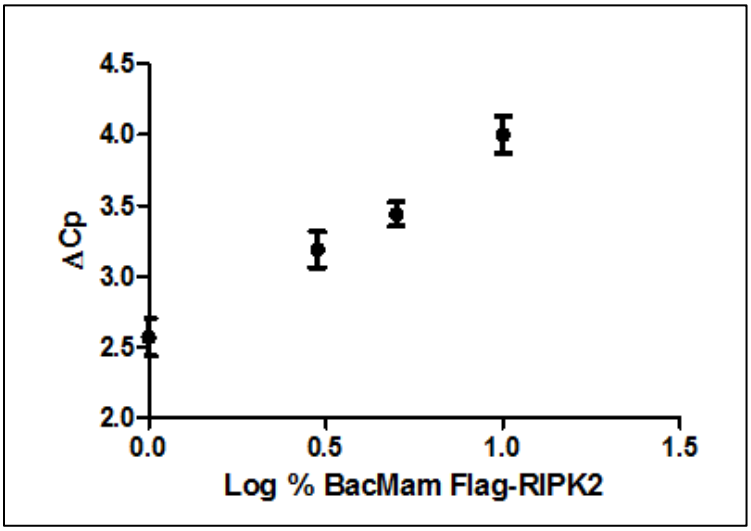

Figure 3-9: Linearization of Flag-RIPK2 Protein Titration data.

The reaction is dependent on protein concentration. As the protein concentration increased, there was a corresponding increase in signal.

The assay performed with Flag-RIPK2 cell lysate also proved to be dependent on the protein concentration. As the percent BacMam increased, there was a corresponding increase in signal. This proves that in addition to purified proteins, this assay can be used to quantify protein-compound interactions in cell lysate.

3.5 Theoretical Model

Theoretical graphs were generated to compare the signal of assays performed with different DNA-encoded compound probe affinities. The antibody was assumed to have a K_d of 0.1 nM and the probe concentrations 0.25nM. The graphs estimated standard curves assuming DNA-encoded compound probe-target interactions with the indicated dissociation constants. The data used to generate the graphs are shown in Appendix A.

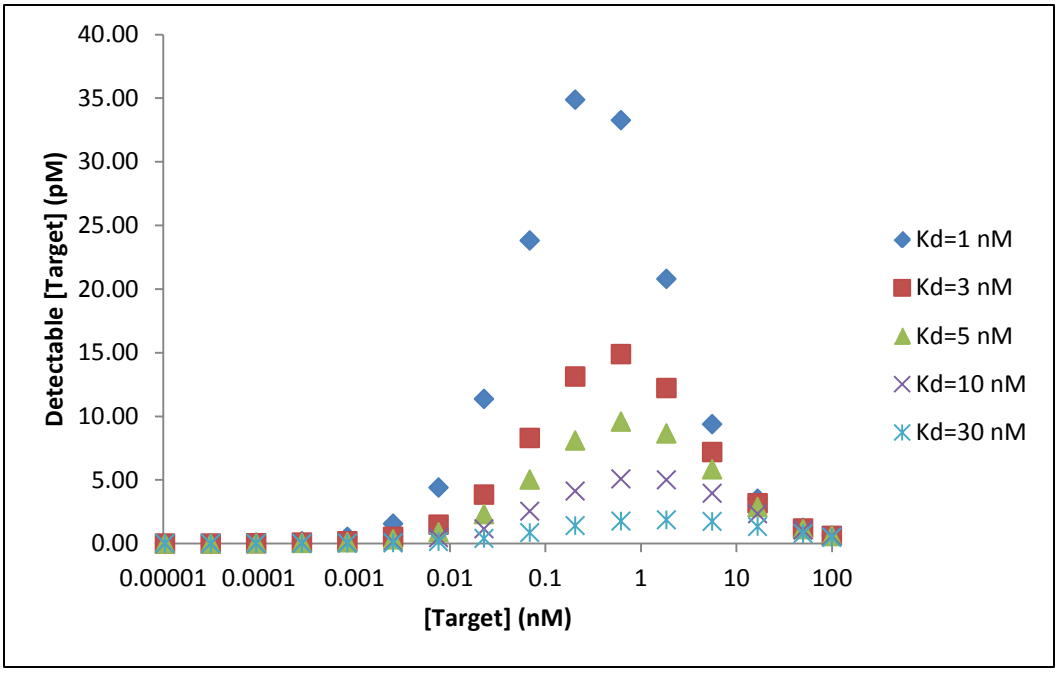

Figure 3-10: Theoretical model for DNA-encoded compounds with different affinities.

Model to determine detectable target concentration for proximity ligation with DNA encoded compounds. Less target protein is needed for assays with a small K_d for the DNA-encoded compound

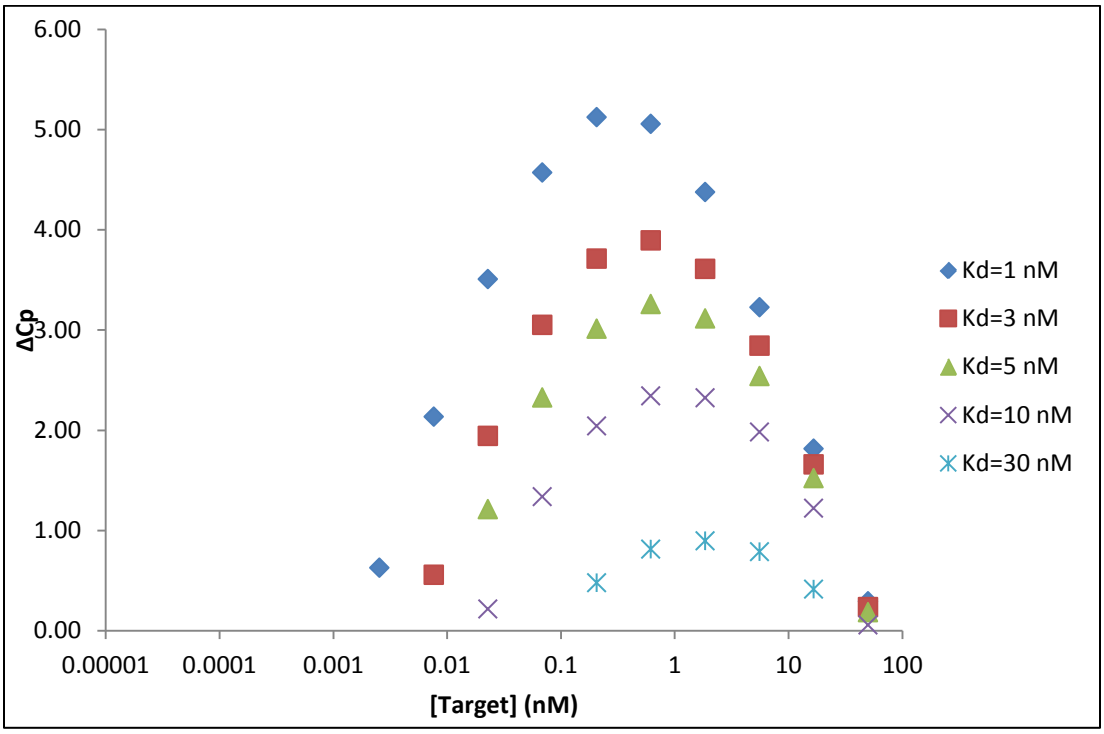

Figure 3-11: Model to determine qPCR signal with DNA encoded compound.

Less target protein is needed for assays with a small K_d for the DNA-encoded compound probe in order to achieve strongest qPCR signal.

A model was generated to quantify the interactions between protein targets and compounds. The model characterized the binding experiment as two independent one-site binding models multiplied together. From these graphs it is apparent that the affinity of the probes strongly influences the sensitivity that can be attained in proximity ligation [2]. The model proved that the affinity of the probes strongly influences the sensitivity that can be attained in proximity ligation. A smaller K_d of the compound corresponds to a stronger affinity for the target protein. Less target protein is needed to achieve the maximum response. According to the theoretical model, proximity ligation signals increase linearly with increasing target up to a point where the probability of each target

molecule being bound by two probes decreases. This point depends on the affinity of the particular probes used [3].

These estimates were compared with experimental results from detection of Flag-RIPK2 and Flag-TNKS1.

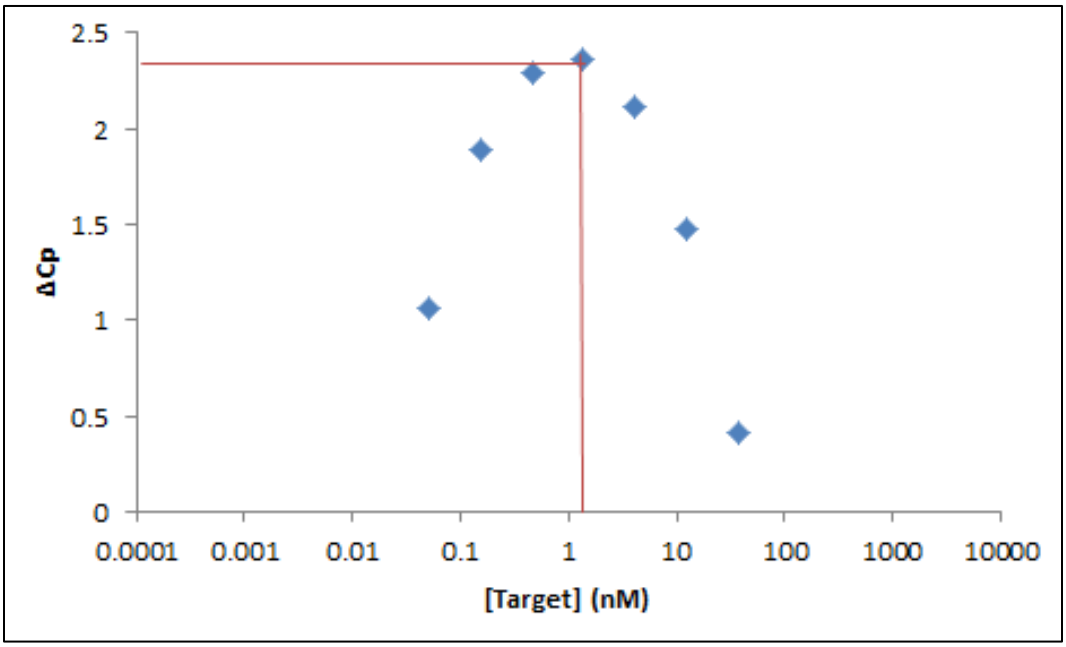

Figure 3-12: qPCR Signal vs. [Flag-RIPK2]

The maximum qPCR signal, for a DNA-encoded compound with a K_d of 10 nM, is 3.8 ΔC_p corresponds to a target protein concentration of 3,130 pM.

The theoretical model predicts the Flag-RIPK2 proximity ligation assay performed with a DNA-encoded compound with a K_d of 10 nM to reach a maximum ΔC_p of 2.36 at 1.37 nM of protein. From the experimental assay the ΔC_p of 5 was obtained at 2.8 nM of protein. There is discrepancy in the theoretical model and the experimental data. In the theoretical model, there is a decrease in signal after the highest signal is obtained with increasing target concentration. In the experimental data, this did not occur. The assay needs to be optimized.

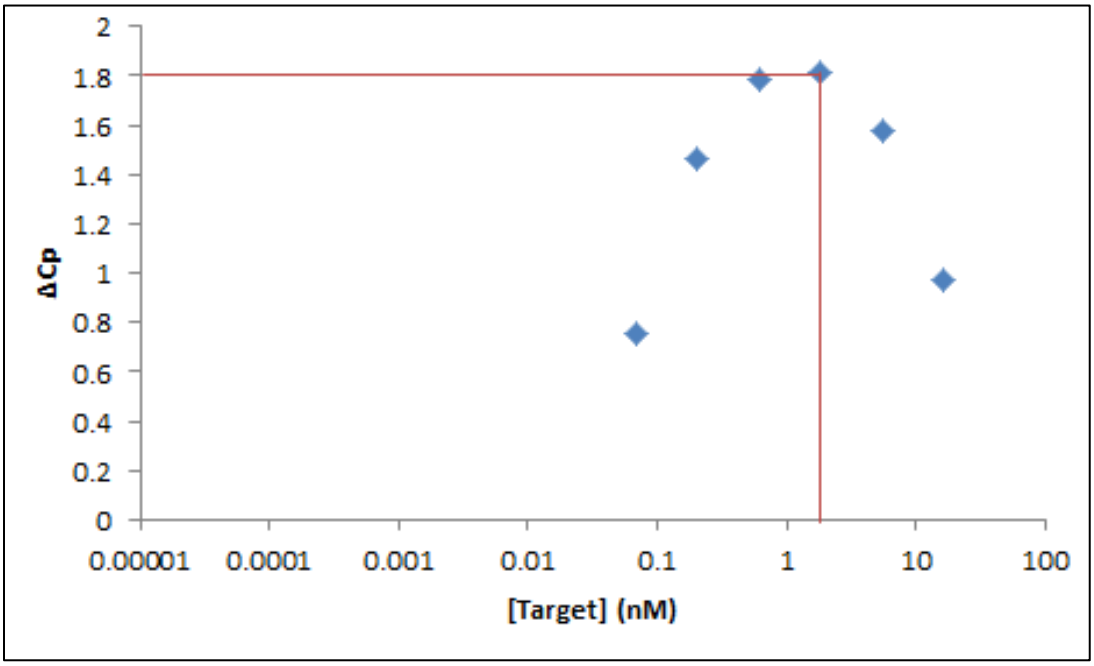

Figure 3-13: qPCR Signal vs. [Flag-TNKS1]

The maximum qPCR signal, for a DNA-encoded compound with a K_d of 15 nM, is 3.6 ΔC_p corresponds to a target protein concentration of 4,000 pM.

The theoretical model predicts Flag-TNKS1 with a DNA-encoded compound of K_d of 15 nM to reach a maximum ΔC_p of 1.82 at 1.85 nM of protein. From the experimental assay the maximum ΔC_p of 6.2 was obtained at 5.5 nM of protein. There is discrepancy in the theoretical model and the experimental data. The experimental maximum ΔC_p and corresponding protein target concentration did not match the theoretical model. The assay needs to be optimized by determining the lowest optimal probe concentrations, confirming the sequences of the DNA, adjusting the oligonucleotide length, and decreasing the connector length.

3.6 Summary of Results

A probe titration experiment was performed to determine the optimal DNA-encoded compound probe concentration to reduce target-independent ligation to be 0.250 nM. Once this was determined, the proximity ligation assay was performed and proved to be dependent on the protein target concentration. The experimental K_d values for purified Flag-TNKS1 and Flag-RIPK2 did not match the actual K_d values. The one-site binding model previously thought to characterize the experiment does not fit the data. Binding is not only dependent on the affinity of the DNA-encoded compound probe, but rather both probes. A new model was generated based on the work published by Gullberg et al [3] to model the assay. The purified Flag-TNKS1 qPCR samples were run on a DNA gel and determined the correct sequence was amplified. Statistical analysis proved that the mean intensities of the bands were not statistically similar. The probes and qPCR primers need to be redesigned so there isn't a decrease in fluorescence with a decrease in protein. The assay was also performed using cell lysate of Flag-RIPK2 and proved to also be protein dependent.

The theoretical model proposed by Gullberg et al [3] was adapted to quantify the interactions between target proteins and compounds based on the different affinities of the DNA-encoded compound for the target protein. The model showed that the affinity of the probes strongly influences the sensitivity that can be attained in proximity ligation. The smaller the K_d, the lower the amount of protein target needed to achieve the maximum response. The experimental data did not correspond to the maximum response target concentration or the maximum ΔC_p. The assay needs to be optimized in order to obtain the theoretical values.

Chapter 4 : Conclusion

There is a need for a method to detect the interactions between proteins and compounds in cells and tissues [2]. Proximity ligation is a relatively new method for highly specific and sensitive detection of proteins in solution, cell culture, or localized tissue [3]. Modifying proximity ligation to quantify the interactions between proteins and compounds using DNA- encoded compounds will prove to be beneficial to drug discovery.

A modification of the TaqMan Assay was made by replacing an antibody of the probe pair with a DNA-encoded compound that has a high affinity for the target. This developed assay was then performed in purified protein and cell lysate. The specific signal from proximity ligation was detected which proved that proximity ligation can be used to detect protein compound interactions in purified protein and cell lysate. Proximity ligation assays can quantitate protein concentrations over three orders of magnitude, and they can reach picomolar detection sensitivities [2]. Theoretical graphs were generated to model the proximity ligation reaction which is dependent on the affinities of both probes for the target protein. The experimental data showed to be protein dependent, but did not match the theoretical graphs. The specific aspects of the assay that need to be optimized are the probe concentrations, the DNA sequences of the oligonucleotides, the oligonucleotide length, and the connector length.

4.1 Future Work

Since the nonlinear regression, one-site specific binding model did not fit the experimental data to determine the known K_d values of the DNA-encoded compound, a theoretical model was generated. Based on the theoretical model, the proximity ligation

assay with a DNA-encoded compound probe and an antibody probe is dependent of the affinities of both probes for the target protein. Experimental error in the assay can be due to the probe concentrations, the DNA sequences, the oligonucleotide lengths, and connector lengths. These aspects of the assay need to be further studied in order to match the experimental data with the theoretical models.

According to results published by Gullberg et al [3], probes with K_d values between 0.1 and 10 nM, only 5-25 pM of proximity probes in a 5 µL incubation volume should be used. The concentration of probes used in the assay should generate a stable protein-independent background. The experimentally determined optimal probe concentration for the DNA-encoded compound probe was determined to be 0.250 nM. This is 10 times greater than the suggested value. The probe titration should be repeated to determine the lowest concentration of probe that produces a signal just above the background. This concentration will minimize the background of proximity probes in the assay because proximity probes with inactive protein binders reduce assay performance [3].

A probe titration should be performed to determine the optimal anti-Flag antibody probe concentration. This should be conducted with a constant concentration of DNA-encoded compound probe. The concentration of the anti-Flag antibody should be varied to determine the lowest concentration of probe. This will also help to decrease the background signal.

The oligonucleotides used in the proximity probes and the DNA of the encoded compound should be sequenced to avoid repeating sequences. If there are repeating sequences, the connector can anneal to all locations which will make it difficult to

determine the positive signal. Eliminating repeating sequences will prevent hybridization of the connector oligonucleotide and formation of inter-probe hybrids.

A previous experiment conducted by Gullberg et al [3] determined for standard proximity ligation assays that the length of the extensions could be varied over a considerable range with negligible effects on ligation efficiency and nonspecific signal. However, this may not be the case in proximity ligation assays with DNA-encoded compounds. The oligonucleotide extensions should be varied to determine the optimal length. The connector length should also be adjusted to form the shortest length possible which will make the assay more specific.

In order to obtain more accurate data, the primer concentrations should be limited. This will cease amplification after the C_p has been established. A matrix to define the limiting primer concentration is shown in Table 4-1.

Table 4-1: Matrix of Varying Concentrations of Forward and Reverse Primers

Forward: Reverse:	100 nM 100 nM	100 nM 80 nM	100 nM 60 nM	100 nM 40 nM	100 nM 20 nM
Forward: Reverse:	80 nM 100 nM	80 nM 80 nM	80 nM 60 nM	80 nM 40 nM	80 nM 20 nM
Forward: Reverse:	60 nM 100 nM	60 nM 80 nM	60 nM 60 nM	60 nM 40 nM	60 nM 20 nM
Forward: Reverse:	40 nM 100 nM	40 nM 80 nM	40 nM 60 nM	40 nM 40 nM	40 nM 20 nM
Forward: Reverse:	20 nM 100 nM	20 nM 80 nM	20 nM 60 nM	20 nM 40 nM	20 nM 20 nM

Once the probe concentrations, the DNA sequences of the oligonucleotides, the oligonucleotide lengths, and the connector length are optimized, the assay can be tested in a DNA-encoded library containing millions of compounds.

References

[1] M. A. Clark, R. A. Acharya, C. C. Arico-Muendel, S. L. Belyanskaya, D. R. Benjamin, N. R. Carlson, P. A. Centrella, C. H. Chiu, S. P. Creaser, J. W. Cuozzo, C. P. Davie, Y. Ding, G. J. Franklin, K. D. Franzen, M. L. Gefter, S. P. Hale, N. J. V. Hansen, D. I. Israel, J. Jiang, M. J. Kavarana, M. S. Kelley, C. S. Kollmann, F. Li, K. Lind, S. Mataruse, P. F. Medeiros, J. A. Messer, P. Myers, H. O'Keefe, M. C. Oliff, C. E. Rise, A. L. Satz, S. R. Skinner, J. L. Svendsen, L. Tang, K. van Vloten, R. W. Wagner, G. Yao, B. Zhao and B. A. Morgan, "Design, synthesis and selection of DNA-encoded small-molecule libraries," *Nature Chemical Biology,* no. 5, pp. 647-654, 2009.

[2] S. M. Gustafsdottir, E. Schallmeiner, S. Fredriksson, M. Gullberg, O. Soderberg, M. Jarvius, J. Jarvius, M. Howell and U. Landegren, "Proximity ligation assays for sensitive and specific protein analyses," *Analytical Biochemistry,* no. 345, pp. 2-9, 2005.

[3] M. Gullberg, S. M. Gustafsdottir, E. Schallmeiner, J. Jarvius, M. Bjarnegard, C. Betsholtz, U. Landegren and S. Fredriksson, "Cytokine detection by antibody-based proximity ligation," *PNAS,* vol. 101, no. 22, pp. 8420-8424, 2004.

[4] S. Fredriksson, M. Gullberg, J. Jarvius, C. Olsson, K. Pietras, S. M. Gustafsdottir, A. Ostman and U. Landegren, "Protein detection using proximity-dependent DNA ligation assays," *Nature Biotechnology,* vol. 20, pp. 473-477, 2002.

[5] M. Gullberg, S. Fredriksson, M. Taussig, J. Jarvius, S. Gustafsdottir and U. Landegren, "A sense of closeness: protein detection by proximity ligation," *Current Opinion in Biotechnology,* no. 14, pp. 82-86, 2003.

[6] Y. Liu, Y. Gu, A. Agner-McWhirter, P. Sathiyanarayanan, M. Gullberg, O. Soderberg, J. Johansson, M. Hammond, D. Ivansson and U. Landegren, "Western blotting via proximity ligation for high performance protein analysis," *Molecular Cell Proteomics,* pp. 1-26, 2011.

[7] E. Swartzman, M. Shannon, P. Lieu, S.-M. Chen, C. Mooney, E. Wei, J. Kuykendall, R. Tan, T. Settineri, L. Egry and D. Ruff, "Expanding applications of protein analysis using proximity ligation and qPCR," *Methods,* no. 50, pp. S23-S26, 2010.

[8] S. M. Gustafsdottir, S. Wennstrom, S. Fredriksson, E. Schallmeiner, A. D. Hamilton, S. M. Sebti and U. Landegren, "Use of Proximity Ligation to Screen for Inhibitors of Interactions between Vascular Endothelial Growth Factor A and Its Receptors," *Clinical Chemistry,* vol. 54, no. 7, pp. 1218-1225, 2008.

[9] Applied Biosystems, *TaqMan Protein Assays Core Reagents Base Kit,* Foster City: Life Technologies, 2010.

[10] "Real-Time PCR Systems," Applied Biosystems, 2005. [Online]. Available: http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_041440.pdf. [Accessed 23 September 2011].

[11] "Real-Time PCR Systems TaqMan Protein Expression Assay," Invitrogen, 2011. [Online]. Available: http://tools.invitrogen.com/content/sfs/manuals/cms_071273.pdf. [Accessed 20

September 2011].

[12] Bio-Equip, "Quantitative Real-time PCR," 24 April 2011. [Online]. Available: http://www.bio-equip.cn/enshow1equip.asp?equipid=1180&division=2801. [Accessed 11 February 2012].

[13] T. A. Kost, J. P. Condreay, R. S. Ames, S. Rees and M. A. Romanos, "Implementation of BacMam virus gene delivery technology in a drug discovery setting," *Drug Discovery Today,* vol. 12, no. 9, pp. 396-403, 2007.

[14] "E-Gel Agarose Gel Electrophoresis," Life Technologies, 2011. [Online]. Available: http://www.invitrogen.com. [Accessed 9 September 2011].

[15] D. H. Motulsky, "The GraphPad Guide to Analyzing Radioligand Binding Data," GraphPad Software, Inc., 1995. [Online]. Available: http://www2.uah.es/farmamol/Public/GraphPad/radiolig.htm. [Accessed 2 March 2012].

[16] N. SEMATECH, "Levene Test for Equality of Variances," 2012. [Online]. Available: http://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm. [Accessed 3 March 2012].

Appendix A: Theoretical Graph Tables

Table A-1: Theoretical Graph for [Complex] vs. [Target] Kd= 1nM

[Probe 2] (nM) (Bmax)	Target (nM) [T]	[Complex] with probe 1 (Y)	[Complex] with probe 2 (Y)	Target fraction bound by Probe 1	Target fraction bound by Probe 2	Target fraction bound by Both	[complex] with Both Probes Bound (Signal)	Signal (pM)
0.25	100.00	0.250	0.248	0.00	0.00	6.18E-06	0.001	0.62
0.25	50.00	0.250	0.245	0.00	0.00	2.45E-05	0.001	1.22
0.25	16.67	0.249	0.236	0.01	0.01	2.11E-04	0.004	3.52
0.25	5.56	0.246	0.212	0.04	0.04	1.69E-03	0.009	9.37
0.25	1.85	0.237	0.162	0.13	0.09	1.12E-02	0.021	20.79
0.25	0.62	0.215	0.095	0.35	0.15	5.39E-02	0.033	33.26
0.25	0.21	0.168	0.043	0.82	0.21	1.70E-01	0.035	34.88
0.25	0.07	0.102	0.016	1.48	0.23	3.47E-01	0.024	23.80
0.25	0.02	0.047	0.006	2.03	0.24	4.97E-01	0.011	11.37
0.25	0.01	0.018	0.002	2.32	0.25	5.76E-01	0.004	4.39
0.25	0.00	0.006	0.001	2.44	0.25	6.08E-01	0.002	1.54
0.25	0.00	0.002	0.000	2.48	0.25	6.19E-01	0.001	0.52
0.25	0.00	0.001	0.000	2.49	0.25	6.23E-01	0.000	0.18
0.25	0.00	0.000	0.000	2.50	0.25	6.24E-01	0.000	0.06
0.25	0.00	0.000	0.000	2.50	0.25	6.25E-01	0.000	0.02
0.25	0.00	0.000	0.000	2.50	0.25	6.25E-01	0.000	0.01

Table A-1: Theoretical Graph for [Complex] vs. [Target] Kd= 3nM

[Probe 2] (nM) (Bmax)	Target (nM) [T]	[Complex] with probe 1 (Y)	[Complex] with probe 2 (Y)	Target fraction bound by Probe 1	Target fraction bound by Probe 2	Target fraction bound by Both	[complex] with Both Probes Bound (Signal)	Signal (pM)
0.25	100	0.250	0.243	0.002	0.0024	6.06E-06	0.0006	0.61
0.25	50	0.250	0.236	0.005	0.0047	2.35E-05	0.0012	1.18
0.25	16.67	0.249	0.212	0.015	0.0127	1.90E-04	0.0032	3.16
0.25	5.56	0.246	0.162	0.044	0.0292	1.29E-03	0.0072	7.18
0.25	1.85	0.237	0.095	0.128	0.0515	6.60E-03	0.0122	12.22
0.25	0.62	0.215	0.043	0.349	0.0691	2.41E-02	0.0149	14.87
0.25	0.21	0.168	0.016	0.818	0.0780	6.38E-02	0.0131	13.12
0.25	0.07	0.102	0.006	1.483	0.0815	1.21E-01	0.0083	8.29
0.25	0.02	0.047	0.002	2.035	0.0827	1.68E-01	0.0038	3.85
0.25	0.01	0.018	0.001	2.323	0.0831	1.93E-01	0.0015	1.47
0.25	0.00	0.006	0.000	2.438	0.0833	2.03E-01	0.0005	0.52
0.25	0.00	0.002	0.000	2.479	0.0833	2.07E-01	0.0002	0.17
0.25	0.00	0.001	0.000	2.493	0.0833	2.08E-01	0.0001	0.06
0.25	0.00	0.000	0.000	2.498	0.0833	2.08E-01	0.0000	0.02
0.25	0.00	0.000	0.000	2.499	0.0833	2.08E-01	0.0000	0.01
0.25	0.00	0.000	0.000	2.500	0.0833	2.08E-01	0.0000	0.00

Table A-2: Theoretical Graph for [Complex] vs. [Target] Kd= 5nM

[Probe 2] (nM) (Bmax)	Target (nM) [T]	[Complex] with probe 1 (Y)	[Complex] with probe 2 (Y)	Target fraction bound by Probe 1	Target fraction bound by Probe 2	Target fraction bound by Both	[complex] with Both Probes Bound (Signal)	Signal (pM)
0.25	100	0.250	0.238	0.00	0.0024	5.95E-06	0.00059	0.595
0.25	50	0.250	0.227	0.00	0.0045	2.27E-05	0.00113	1.134
0.25	16.67	0.249	0.192	0.01	0.0115	1.72E-04	0.00287	2.867
0.25	5.56	0.246	0.132	0.04	0.0237	1.05E-03	0.00582	5.816
0.25	1.85	0.237	0.068	0.13	0.0365	4.67E-03	0.00865	8.654
0.25	0.62	0.215	0.027	0.35	0.0445	1.55E-02	0.00958	9.575
0.25	0.21	0.168	0.010	0.82	0.0480	3.93E-02	0.00808	8.079
0.25	0.07	0.102	0.003	1.48	0.0493	7.31E-02	0.00502	5.017
0.25	0.02	0.047	0.001	2.03	0.0498	1.01E-01	0.00232	2.315
0.25	0.01	0.018	0.000	2.32	0.0499	1.16E-01	0.00088	0.884
0.25	0.00	0.006	0.000	2.44	0.0500	1.22E-01	0.00031	0.310
0.25	0.00	0.002	0.000	2.48	0.0500	1.24E-01	0.00010	0.105
0.25	0.00	0.001	0.000	2.49	0.0500	1.25E-01	0.00004	0.035
0.25	0.00	0.000	0.000	2.50	0.0500	1.25E-01	0.00001	0.012
0.25	0.00	0.000	0.000	2.50	0.0500	1.25E-01	0.00000	0.004
0.25	0.00	0.000	0.000	2.50	0.0500	1.25E-01	0.00000	0.001

Table A-3: Theoretical Graph for [Complex] vs. [Target] Kd= 10nM

[Probe 2] (nM) (Bmax)	Target (nM) [T]	[Complex] with probe 1 (Y)	[Complex] with probe 2 (Y)	Target fraction bound by Probe 1	Target fraction bound by Probe 2	Target fraction bound by Both	[complex] with Both Probes Bound (Signal)	Signal (pM)
0.25	100	0.250	0.227	0.00	0.0023	5.68E-06	0.00057	0.568
0.25	50	0.250	0.208	0.00	0.0042	2.08E-05	0.00104	1.040
0.25	16.67	0.249	0.156	0.01	0.0094	1.40E-04	0.00233	2.330
0.25	5.56	0.246	0.089	0.04	0.0161	7.10E-04	0.00395	3.947
0.25	1.85	0.237	0.039	0.13	0.0211	2.70E-03	0.00500	5.003
0.25	0.62	0.215	0.015	0.35	0.0235	8.21E-03	0.00507	5.066
0.25	0.21	0.168	0.005	0.82	0.0245	2.00E-02	0.00412	4.121
0.25	0.07	0.102	0.002	1.48	0.0248	3.68E-02	0.00253	2.525
0.25	0.02	0.047	0.001	2.03	0.0249	5.08E-02	0.00116	1.160
0.25	0.01	0.018	0.000	2.32	0.0250	5.80E-02	0.00044	0.442
0.25	0.00	0.006	0.000	2.44	0.0250	6.09E-02	0.00015	0.155
0.25	0.00	0.002	0.000	2.48	0.0250	6.20E-02	0.00005	0.052
0.25	0.00	0.001	0.000	2.49	0.0250	6.23E-02	0.00002	0.018
0.25	0.00	0.000	0.000	2.50	0.0250	6.24E-02	0.00001	0.006
0.25	0.00	0.000	0.000	2.50	0.0250	6.25E-02	0.00000	0.002
0.25	0.00	0.000	0.000	2.50	0.0250	6.25E-02	0.00000	0.001

Table A-4: Theoretical Graph for [Complex] vs. [Target] Kd= 30nM

[Probe 2] (nM) (Bmax)	Target (nM) [T]	[Complex] with probe 1 (Y)	[Complex] with probe 2 (Y)	Target fraction bound by Probe 1	Target fraction bound by Probe 2	Target fraction bound by Both	[complex] with Both Probes Bound (Signal)	Signal (pM)
0.25	100	0.250	0.192	0.00	0.0019	4.80E-06	0.00048	0.480
0.25	50	0.250	0.156	0.00	0.0031	1.56E-05	0.00078	0.780
0.25	16.67	0.249	0.089	0.01	0.0054	7.99E-05	0.00133	1.331
0.25	5.56	0.246	0.039	0.04	0.0070	3.11E-04	0.00173	1.727
0.25	1.85	0.237	0.015	0.13	0.0078	1.01E-03	0.00186	1.862
0.25	0.62	0.215	0.005	0.35	0.0082	2.85E-03	0.00176	1.757
0.25	0.21	0.168	0.002	0.82	0.0083	6.77E-03	0.00139	1.392
0.25	0.07	0.102	0.001	1.48	0.0083	1.23E-02	0.00085	0.846
0.25	0.02	0.047	0.000	2.03	0.0083	1.69E-02	0.00039	0.387
0.25	0.01	0.018	0.000	2.32	0.0083	1.94E-02	0.00015	0.147
0.25	0.00	0.006	0.000	2.44	0.0083	2.03E-02	0.00005	0.052
0.25	0.00	0.002	0.000	2.48	0.0083	2.07E-02	0.00002	0.017
0.25	0.00	0.001	0.000	2.49	0.0083	2.08E-02	0.00001	0.006
0.25	0.00	0.000	0.000	2.50	0.0083	2.08E-02	0.00000	0.002
0.25	0.00	0.000	0.000	2.50	0.0083	2.08E-02	0.00000	0.001
0.25	0.00	0.000	0.000	2.50	0.0083	2.08E-02	0.00000	0.000

Table A-5: ΔCp vs. [Target] for corresponding Kd values

	Detectable [Target] (pM)					***ΔCp***				
Target (nM)	**Kd=1**	**Kd=3**	**Kd=5**	**Kd=10**	**Kd=30**	**Kd=1**	**Kd=3**	**Kd=5**	**Kd=10**	**Kd=30**
100	0.62	0.61	0.59	0.57	0.48	-0.69	-0.72	-0.75	-0.82	-1.06
50	1.22	1.18	1.13	1.04	0.78	0.29	0.23	0.18	0.06	-0.36
16.67	3.52	3.16	2.87	2.33	1.33	1.81	1.66	1.52	1.22	0.41
5.56	9.37	7.18	5.82	3.95	1.73	3.23	2.84	2.54	1.98	0.79
1.85	20.79	12.22	8.65	5.00	1.86	4.38	3.61	3.11	2.32	0.90
0.62	33.26	14.87	9.58	5.07	1.76	5.06	3.89	3.26	2.34	0.81
0.21	34.88	13.12	8.08	4.12	1.39	5.12	3.71	3.01	2.04	0.48
0.07	23.80	8.29	5.02	2.53	0.85	4.57	3.05	2.33	1.34	-0.24
0.02	11.37	3.85	2.32	1.16	0.39	3.51	1.94	1.21	0.21	-1.37
0.01	4.39	1.47	0.88	0.44	0.15	2.13	0.56	-0.18	-1.18	-2.76
0.00	1.54	0.52	0.31	0.15	0.05	0.63	-0.96	-1.69	-2.69	-4.28
0.00	0.52	0.17	0.10	0.05	0.02	-0.93	-2.52	-3.25	-4.25	-5.84
0.00	0.18	0.06	0.04	0.02	0.01	-2.51	-4.09	-4.83	-5.83	-7.41
0.00	0.06	0.02	0.01	0.01	0.00	-4.09	-5.67	-6.41	-7.41	-9.00
0.00	0.02	0.01	0.00	0.00	0.00	-5.67	-7.26	-8.00	-9.00	-10.58
0.00	0.01	0.00	0.00	0.00	0.00	-7.26	-8.84	-9.58	-10.58	-12.16

Jennifer L. McCaffrey
Address: 1200 Slough Drive ♦ Collegeville, PA 19426 ♦ **Email:** jlm5455@psu.edu

Education

The Pennsylvania State University
University Park, PA
Schreyer Honors College

Bachelor of Science in Bioengineering
Chemical Engineering Option
Expected Graduation: May 2012

Experience

Research and Development Co-op
GlaxoSmithKline (GSK), Collegeville PA (January-June 2011)

- ♦ Adapted Proximity Mediated Ligation (PML) to detect compound-target interaction within purified proteins as well as cell lysate transduced with BacMam viruses
- ♦ Utilized quantitative real time- Polymerase Chain Reaction (qPCR) to quantify compound-target interaction
- ♦ Used Agarose Gel Electrophoresis to validate the PML product after qPCR
- ♦ Applied PML theory to GSK's Encoded Library Technology to detect difficult to purify proteins

BioVue Process Improvement Specialist Co-op
Ortho Clinical Diagnostics, a Johnson & Johnson Company, Raritan NJ (June -December 2010)

- ♦ Reduced downtime on BioVue production lines by implementing a Changeover Process
- ♦ Performed root cause investigations by evaluating equipment logs and inspecting retain samples
- ♦ Attended 5S trainings and sustained the program throughout the production lines to create a lean manufacturing area
- ♦ Assisted Team Leader with Quarterly Meeting Preparation and assumed responsibility for the department in absence of Team Leader
- ♦ Utilized Six Sigma methodology to design experiments performed on the production lines

Leadership/Activities

Member, Penn State THON Operations Committee (October-February 2012)

- ♦ Mail Chair responsible for communication with Pen Pal

Member, Penn State Women in Engineering Program (August 2007-May 2010)

- ♦ Participant in Women in Engineering Program Orientation (WEPO)
- ♦ Mentor for Girl Scout Saturdays

Jennifer L. McCaffrey
Address: 1200 Slough Drive ♦ Collegeville, PA 19426 ♦ **Email:** jlm5455@psu.edu

---

Leadership/Activities

Leader, Intern and Co-op Association (ICA) Community Service Chair (July- December 2010)
- ♦ Organizer of community service events
- ♦ Leader of the United Way "Week of Caring" for Ortho Clinical Diagnostics

Leader, Solar Panel Inauguration Media Team (June-September 2010)
Leader, Facilitated Study Group (August 2010- May 2011)