UNIVERSITY OF CALIFORNIA
RIVERSIDE

Development of *In vitro* Quantitative Förster Resonance Energy Transfer (qFRET) Reporter Assays for the Characterization of Non-covalent Interaction of Membrane Proteins and Covalent SUMOylation Modification of Viral Proteins Coupled with Mass Spectrometry Identification of SUMO1 Modified Lysine.

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Bioengineering

by

Vipul Madahar

September 2021

Dissertation Committee:
Dr. Jiayu Liao, Chairperson
Dr. Victor GJ Rodgers
Dr. John Jefferson Perry

The Dissertation of Vipul Madahar is approved:

Committee Chairperson

University of California, Riverside

Acknowledgements

I would like to acknowledge and thank Dr. Jiayu Liao for the opportunity to join his lab as a graduate student. His mentorship and drive to continually push every project to the highest level of quality is instilled in my academic work. I have gotten to know Dr. Liao through many years as his student, and his dedication, investment, and pursuit of excellence are unbound.

I am thankful to my committee members, Dr. Victor GJ Rodgers and Dr. John Jefferson Perry for helpful discussions, and for their support and helpful advice. I am grateful to Dr. Joshua Morgan for recommending me for the Dean of Graduate Division for the Dissertation Year Fellowship. I would like to sincerely thank the Department of Bioengineering and Department of Anthropology for providing teaching positions for financial support. The team in GradScucess, Dr. Hillary Jenks, Patrick Thomas, and Dr. Phillip Brisk, for the Gradaute Student Mentorship position and for the financial support. I am very thankful towards all the members in Liao Lab for very close collaborative work, the graduate students, Michael Xiong and George Way for putting together the viral plaque assays, to Chenghao Shen and Runrui Dan for helping with protein expression. I would like to thank the alumni of Liao lab, Harbani Malik, Yan Liu, Hilda Chan. I also want to extend a thanks to the Yang Song for teaching me how to do traditional cloning using T4 ligase and for help in cloning the original plasmids for the SUMO cascade.

I would like to thank my friends and colleagues in the BIG program, my friend Lorenzo Carlos IV and his family for the mental support, and my friends in Shanghai,

especially Guang Yao Lin for all the support and friendship while overseas. Also, the undergraduate mentees, Mary Nwangwu, Amanda Xaypraseuth, and Garrett Scott for the chance to mentor.

I would like to thank and dedicate this dissertation to my family, my Nani ji Ravinder Kaur Badhan, Varun, Shama, Maya, and Arjun Madahar for inspiring me to constantly innovate, to my family in India and the States, Smriti, Suruchi, Anju Bhua and Satya Prakash Kazal Fufar Ji for being my drive, and my parents Surinder and Narinder Madahar for being the tsunami in my life of overwhelming love and support.

ABSTRACT OF THE DISSERTATION

Development of *In vitro* Quantitative Förster Resonance Energy Transfer (qFRET) Reporter Assays for the Characterization of Non-covalent Interaction of Solubilized Membrane Proteins and Covalent SUMOylation Modification of Viral Proteins Coupled with Mass Spectrometry Identification of SUMO1 Modified Lysine

by

Vipul Madahar

Doctor of Philosophy, Graduate Program in Bioengineering
University of California, Riverside, September 2021
Dr. Jiayu Liao, Chairperson

Disease progression is often facilitated by protein-protein interaction. Reports of commandeering host protein mechanisms by cancer and viral proteins has armed scientist with novel drug targets. The challenge arises when attempting to quantitatively evaluate the variety of covalent and non-covalent interactions that occur during disease progression. This thesis covers the development of a quantitative Förster resonance energy transfer (qFRET) assay for the evaluation of covalent SUMOylation of viral proteins and non-covalent interaction of programmed cell death 1 (PD1) with programmed cell death ligand 1 (PDL1).

The interaction of PD1 and PDL1 results in the negative regulation of T cell immune response and is reported to be used by cancer cells to circumvent the immune checkpoint mechanism. Immune checkpoints inhibitors, that target PD1-PDL1 interaction, have been the focus for cancer therapies in the past decade, but recent

development of novel treatments has plateaued. Reported here is a FRET based reporter for the *in vitro* characterization of the PD1 and PDL1 interaction. The results of the work produced a qFRET based K_D of full-length custom codon optimized PD1 and PDL1 of 0.82 µM with standard error of 0.13 µM and determined a qFRET based K_i of Permbrolizumab^{TM} to be 1.14 nM with a 95 % confidence interval of 0.94 to 2.14 nM.

The remainder of the work focuses on the development of an *in vitro* SUMOylation of viral proteins with qFRET as a reporter, in combination with mass spectrometry to identify SUMO1 modified lysine sites. The qFRET assay for the *in vitro* SUMOylation of IAV-Matrix Protein 1 in tandem with MS, resolved 5 SUMO1 modification. The evaluation of the novel K21 modification proved fatal to the viral pathogenesis, and M1 K242R mutant reported diminished IAV infectivity. The qFRET based *in vitro* SUMOylation of severe acute respiratory syndrome coronavirus 2 Nucleocapsid protein (SARS-CoV- 2 N) coupled with MS analysis identified 4 novel SUMO modified lysine residues. The evaluation of identified lysine *in vitro* and in cell mutant studies found lysine 61 and 65 to dimmish SUMOylation activity and modulate cellular translocation.

Table of Contents

Chapter 5

Chapter 6

List of Figures

the FRET wavelength shown. C.) Em_{FRET} equations is used to extract the Em_{FRET} signal, and D.) Alpha α ratio for the donor is calculated by dividing the emission at 530 nm by emission at 475 nm when excited at 414 nm. The β parameter is determined by diving the emission of YPet at 530 nm when excited at 414 nm over the emission at 530 nm when excited by 475 nm

List of Tables

List of Abbreviations

qFRET: Quantitative Förster Resonance Energy Transfer

SUMO: Small Ubiquitin like/related Modifier

PTM: Post Translation Modification

PD1: Programmed Death 1

PDL1: Programmed Death Ligand 1

SPR: Surface Plasmon Resonance

ITC: Isothermal Titration Calorimetry

MS: Mass Spectrometry

IAV: Influenza A Virus

HA: Hemagglutinin

NA: Neuraminidase

PB1: RNA polymerase subunit 1

PB2: RNA polymerase subunit 2

NS1: Non-Structural Protein 1

M1: Matrix 1

NP: Nucleocapsid Protein

SARS-CoV-1: severe acute respiratory syndrome coronavirus 1

SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

S: Spike Protein

M: Membrane Protein

E: Envelope Protein

N: Nucleocapsid Protein

ATG4: Autophagy Related Protein 4

Atg8: Autophagy Related Protein 8

HTS: High Throughput Screening

Chapter 1: Introduction

1.1 Quantitative Protein-Protein Interaction Detection

We have reached another historic milestone for human genome sequencing, the historic 2003 report of human genome sequence which excluded regions of repeating sequences was updated May 2021.[1] Began with a small group of scientist part of the "Telomere-to-Telomere consortium" utilized a multifaceted approach to overcome the technological hurdles that came with sequencing region of human genome with repeating base pairs (bp) sequences longer than the read fragment.[1] Repeating DNA sequences are an inherent vulnerability in current DNA sequencing technology. Each technology fragments the genome into shorter DNA fragments, if the repeating region is longer then the fragment then the genome alignment has difficulty constructing the whole genome.

Only through the combination of multiple sequencing technologies, specifically the longer read lengths, HiFi PacBio (read length 20K bp), Nanopore (read length 1K bp), and Illumina sequencing (read length 251-500 bp), the group was able to sequence and align the final 8 percent of the Human genome in the X chromosome. This feat provides evidence of the technological challenge faced when approaching the magnitude of problem as sequencing the human genome. Similarly, the proteomics community has reached a major milestone in identification of up to 90 percent of the human proteome.[2]

The human proteome project (HPP) established recently in comparison to Human genome project, HPP began in 2010, set out to construct a blueprint database of the human proteome network. In 2020 the project reached a milestone in the blueprint covering more than 90 percent of the human proteome. This set of data consisted of the

combination of mass spectrometry (MS) and anti-body pull down accumulated by labs across the globe.[2,3] The overall impact of sequencing the human genome and the near completion of identifying all the proteome are observed throughout personalized treatment of disease such as cancer and is still an uphill battle. The near completion of these two major projects is the starting line on the path to understanding the relationship of genotypes to phenotypes.

The breakthroughs in sequencing technology and MS have produced a greater understanding in genomic and proteomic identification of our biological network. Lessoned learned from the investigations of prediction of genotypic mutation to phenotype leads us to the natural transition to the interactome.[4] In cancer models evidence of genomic mutations that are a marker for cancer subtypes are an amazing feat and are found to be the starting line in understanding the mechanism of this disease's progression.[5] At this starting line, our progress now moves towards understanding the mechanistic outcomes of genotypic mutations.

The bottle neck in the transition from the proteome to interactome is the technological and biological challenges with characterization of protein-protein interactions. This work consists of the development of a protein-protein interaction technology platform based on Förster Resonance Energy Transfer (FRET). Quantitative Förster Resonance Energy Transfer (qFRET) a singular platform that provides kinetic constants such as equilibrium dissociation constant, the enzymatic constants like k_{cat}/K_m, and can also be adapted for characterization of enzymatic inhibition and determine K_i. The following set of work showcases the development of a qFRET approach to

characterize non-covalent receptor and ligand interaction and their response to inhibitors. The qFRET platform also adapted for evaluating the dissociation constant and covalent binding events of the SUMO modification in heterologous proteins from viral proteome.

1.2 Technological Advancements in Detection of Protein-Protein Interaction

A brief overview of PPI characterization technology highlights the advancements made in the characterization of this type of interaction across platforms and presents evidence of advantages of qFRET technology. The two-alternative set of quantitative protein-protein interaction platforms are Surface plasmon resonance (SPR) and Isothermal titration calorimetry (ITC). The summaries below provide an overview of the two alternative platforms and highlight the advantages and disadvantages of the platforms.

Surface Plasmon Resonance (SPR) Biosensor

Surface plasmon resonance (SPR) biosensor is a quantitative spectroscopic technique that is widely used for measuring binding events and specializes in characterization of non-covalent interactions. The method detects the binding events at a functionalized surface, by monitoring modulation of SPR angle caused by changes in the refractive index at the functional surface.[6] The sensing surface is a metallic (gold, silver, aluminum) surface interfaced with a prism.[7,8] The metal surface is functionalized for attachment of a binding partner of interest. A monochrome light is reflected from the functionalized metal surface at an angle at which the excited metal surface electrons begin to oscillate and generate plasmon, SPR angle. The binding partner of interest is bound to the functionalized metal surface and potential binding molecules are solubilized

in binding solution and flown over the functionalized surface. On the event of biomolecules that bind to the bound proteins then the refractive index would change and the SPR angle will shift. This modulation is recorded in the change in the angle and or the intensity of the light reflected.[9]

The advantages of this method are that you can detect both rate of binding partner on and rate of off at high sensitivity. Advancements of this technique include waveguide SPR sensors, that demonstrate an increase in sensitivity by replacing the prism metal interface with a core-cladding interface within a mated optical fiber instead of a prism. The technological advances in wave guide SPR sensors has paved the way to miniaturizing the SPR sensors with higher sensitivity.[10] The major disadvantage of the SPR technique is the required surface attachment of one of the binding molecules to the surface. The attachment procedure can include covalent modification of biotin to the protein, or amine coupling to the sensor surface. SPR technique has been applied to characterize binding of small molecules to a protein surface, as well as detection of an organism by attaching antigen on the functionalized surface. The future development of this technology includes its adaptation to high-throughput studies and applications of functional hydrogels for small molecule diffusion studies. This technique is versatile, SPR has been adapted for non-covalent PPI K_D, small molecule and protein interaction K_i, and antigen-antibody interaction.

Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is a method for characterization of affinity and stoichiometry of interacting molecules. The ITC method records minutiae changes in

temperature caused by the endothermic or exothermic reaction between two interacting molecules.[11,12] General instrument setup contains two metallic chambers with a thermopile attached to them that measures, records, and helps modulate temperature. One of the chambers acts as a reference and the other is the sample chamber and the temperature is kept at a constant between the two chambers. The temperature difference between the sample chamber and the reference is continuously adjusted and the change in temperature is recorded in real time.

The sample chamber is connected to a motorized syringe system that injects the binding molecule (substrate of interest), solubilized in the sample buffer in series of injections. The change in enthalpy is determined with each cycle of injection of binding molecule into the sample chamber, and the correction of temperature increase, exothermic, or decrease, endothermic reaction is recorded as a raw feedback signal. The feedback signal to power the heating element for adjustment of temperature in an endothermic or exothermic reaction, provides the raw report for the interaction of the two molecules along with the magnitude of feedback signal. Series of injections provide a feedback response till the samples in the chamber are saturated and results in no change in temperature detected. Extraction and analysis of the raw feedback power signal gives differences in free and bound states and ultimately can be used to determine dissociation constant.[12]

This method has been adapted for enzyme kinetics by having a low concentration of enzyme in the sample chamber and introducing titration of substrate till the enzyme is saturated.[13] This technique does not require modification of the molecules when taking

the measurement, immobilization or tag are not needed for this technique, measured K_D, K_i, k_{cat}, and K_m have been reported.[14] The disadvantage in this technique is the need to use high concentration of molecules for response. The technique is limited by the thermopile, and temperature regulators to observe sub nano-molar range of interaction.

1.3 qFRET Platform

The applications of the Förster resonance energy transfer (FRET) phenomenon have revolutionized protein studies and provided numerous dimensions of freedom to protein scientists.[15] The FRET signal reports the non-radiative energy transfer from a fluorescent donor molecule to an acceptor fluorescent molecule and the efficiency of that transfer depends on a known set of parameters that include the spectral overlap of the two molecules and their proximity. The technique allows scientists to study the interaction of solubilized or immobilized molecules with only a bench top fluorometer to detect the FRET signal.[16]

Historically the FRET phenomenon was first reported by a father and son pair, Jean-Baptiste Perrin, and Francis Perrin. This pair were the first to note the distance dependency of the dipole-dipole interaction between the donor and accepter pairs.[17] Theodor Förster later characterized the phenomenon and found that the dipole-dipole coupling occurred only within the range of 1 – 10 nm.[18] Förster described the fundamental energy transfer efficiency equation E (Eq1) between the donor molecule and the acceptor molecule, shown below.[17] The FRET requirements extended to not only the distance between the two molecules (R), but also their spectral overlap (J) and the orientation kappa (κ) of the transition dipoles of both molecules that defines R_0. Each

FRET pair of molecules will have a Förster distance, R_0, value which is the distance at which half of the excited energy of the donor is transferred to the acceptor.

$$\text{Equation 1.1:} \quad E = \frac{R_0^6}{R_0^6 + R^6}$$

$$\text{Equation 1.2:} \quad R_0^6 = 0.021 * \kappa^2 * n^{-4} * \Phi_D * J$$

$$\text{Equation 1.3:} \quad J = \int \bar{I}_D * \epsilon_A * \lambda^4 * d\lambda$$

The R_0 distance is determined using the relationship above, the Φ_D is the quantum yield of the donor, and the n is the refractive index. The kappa squared (κ^2) parameter is assumed to be 2/3 as dipole orientation is very difficult to predict due to the uncertainty in molecular motion. The spectral overlap J is dependent on the normalized emission of the donor spectrum, $\bar{I}_D$, and the molar extinction coefficient, ϵ_A , of the acceptor at its peak absorption wavelength, and the wavelength at the spectral overlap, λ^4. The larger the overlap the larger the J component of R_0. Thus, with a pair of donor and acceptor molecules with a known R_0, we can relate the efficiency of FRET to the physical distance between the donor and acceptor pair. Equation 1.1 highlights the R^6 magnitude of sensitivity to the distance between acceptor and donor on FRET efficiency. The applications of this principle have sparked a series of biomolecular assays that can interpret the FRET efficiency between two molecules coming together or apart down to nano meter resolution.[16] The applications of this phenomenon is dependent on the development of sensitive donor and acceptor pairs. In this work we utilize an engineered FRET fluorescent protein pair to detect the interaction of proteins.

1.4 CyPet and YPet a FRET Pair for Reporting Protein-Protein Interaction

The green fluorescent protein (GFP) mutants have paved the way for fluorescent protein applications such as reporters for gene expression, protein intercellular trafficking, and protease activity.[19,20] Fluorescent proteins can be expressed in cells at sufficient levels to be detected for in-cell analysis and purified in significant quantities for high-throughput screening assays. The GFP mutants have provided a variety of fluorescent proteins that cover much of the visible spectrum. The group Nguyen Annalee W. and Daugherty Patrick S., from U.C. Santa Barbara, report a mutation study on the cyan-fluorescent protein (CFP) and yellow-fluorescent protein (YFP) for improvement in FRET efficiency.[21] The study resulted in the two proteins CyPet and YPet which demonstrated a 20-fold improvement in FRET efficiency in comparison to CFP-YFP. The optimized FRET pairs reported R_0 is 5.15 nm and reported range of FRET signal falls within 1 – 10 nm. The predicted dynamic range of FRET efficiency of the optimized FRET pair is approximately 3 – 10 nm (Figure 1A).[22] The optimized pair have high efficiency between 0 – 3 nm, which can make is very difficult to detect changes in R between 0 – 3 nm and distances between 8 – 10 nm will be more difficult to detect due to lower 5 % efficiency range.

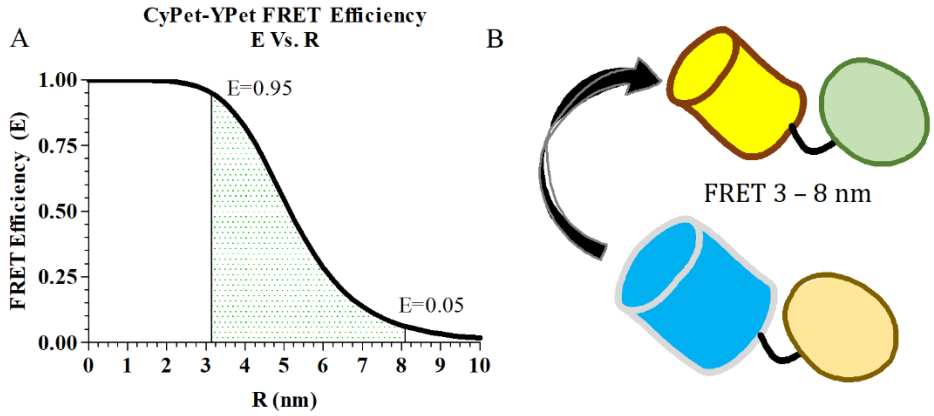


Figure 1:The predicted FRET Efficiency Vs R plot with green pattern highlighting the range in R in nanometer of detectable FRET response. Illustration of FRET pair CyPet (blue barrel) and YPet (yellow barrel) fused to an interaction pair of interest.

The protein-protein interaction of two separate proteins recombinantly fused to CyPet or YPet provide effective range in FRET efficiency and fluorescent signal to be detected from a bench top fluorometer (Figure 1B). The nature of the FRET phenomenon requires evaluation of the observed fluorescence for false positives, as the interpretation of the FRET signal reports the distance between the donor and acceptor and not the direct binding event. Additionally, the FRET signal can have crosstalk from donor and acceptor emission in the FRET wavelength emission. The magnitude of the FRET fluorescent signal could be heavily impacted by the amount of crosstalk that occurs when the donor fluoresces into the FRET wavelength, and if the acceptor is excited by the donor excitation wavelength.

1.5 Quantitative FRET Signal Processing by Cross-Channel Analysis

In Liao lab we have developed a qFRET technique for the investigation of protein-protein interactions in-vitro.[23] The qFRET technique leverages the principles of the "three cube FRET" (three filter cube FRET) which account for the channel cross talk into the FRET wavelength.[15,24] The raw total FRET emission, EmTotal, signal for CyPet-YPet pair is measured at emission wavelength 530 nm with excitation at 414 nm (Figure 2A). Within the "EmTotal" signal, at 530 nm, is the embedded cross talk from the free (unbound) donor and acceptor emission (Figure 2A). To extract the crosstalk signal originating from the donor pair, a crosstalk coefficient, α, and acceptor pair cross talk coefficient, β, are first resolved.

Table 1: Three channel fluorescent measurements, EmTotal, donor fluorescence FL_D, and the acceptor fluorescence FL_A. These excitation and emission wavelengths are used to extract Em_{FRET} for the FRET pair CyPet-YPet

	Excitation (λ)	Emission (λ)
Em_{Total}	414 nm	530 nm
Fl_D	414 nm	475 nm
FL_A	475 nm	530 nm

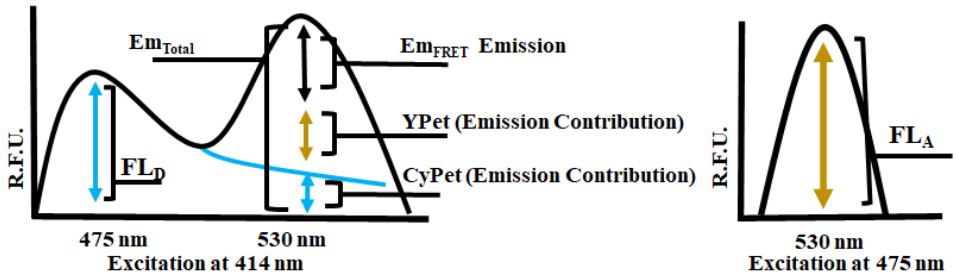


Figure 2: Principles of "three filter cube FRET". Illustrated the embedded crosstalk from donor and acceptor pairs within the FRET signal.

The Em_{FRET} signal is resolved by taking three measurements of a qFRET reaction (Table 1). The EmTotal taken with excitation at 414 nm and emission at 530 nm. The CyPet contribution is resolved by, FL_D, the emission of the donor taken at 414 nm excitation and 475 nm emission, multiplied by the crosstalk coefficient α. The YPet contribution is resolved by, FL_A, the emission of the acceptor, at 475 nm excitation and 530 nm emission, multiplied by the crosstalk coefficient β. The equation Em_{FRET} derived from the principles of "three cube FRET".

$$Em_{FRET} = (Em_{Total}) - ((FL_D * \alpha) + (FL_A * \beta)) \quad \text{Equation 1.4}$$

$$\alpha = \frac{\text{Donor emisson at 530 nm with excitation at 414 nm}}{\text{Donor emisson at 475 nm with excitation at 414 nm}} \quad \text{Equation 1.5}$$

$$\beta = \frac{\text{Acceptor emission at 530 nm with excitation at 414 nm}}{\text{Acceptor emission at 530 nm with excitation at 474 nm}} \quad \text{Equation 1.6}$$

The crosstalk parameters, alpha (α), and beta (β) are part of the qFRET signal extraction equation resolving Em_{FRET} (Equation 4). Alpha and beta are unitless ratiometric constants that are unique to the instrument's response to the donor and acceptor fluorescence. The donor protein, CyPet, is excited at 414 nm and has peak emission is at 475 nm. The acceptor protein, YPet, is excited at 475 nm with peak emission at 530 nm. The alpha (α) parameter is determined for the fluorometer by exciting the donor fusion protein CyPet at 414 nm and then a ratio of emission at FRET wavelength of 530 nm divided by the emission at 475 nm. The beta (β) parameter for the acceptor protein, YPet, is derived by dividing the emission at 530 nm when excited at 414 nm by the emission at 530 nm when excited at 475 nm. We observed the crosstalk parameter of donor CyPet to be at 0.34 (α) standard deviation of 0.003 and a CV of 0.1 %. The accepter YPet to be at 0.03 (β) standard deviation of 0.001 and a CV of 3 % n = 9

for both parameters. Table 2 lists the alpha and beta parameter determined for the instrument Molecular Devices Spectra Max3^{TM} used in this work.

Table 2: Listed mean, standard deviation, Coefficient of variation over 9 measurements at constant concentration of both acceptor and donor.

Parameter	**Alpha (α)**	**Beta (β)**
Mean	0.34	0.03
Standard Deviation	0.003	0.001
Coefficient Of Variation	0.001	0.03
n	9	9

The qFRET platform has been developed into a toolbox of assays for characterization of protein-protein interaction. The dissociation constant K_D of SUMO1 and UBC9 is reported in the study conducted by Song *et al.*.[23] The work completed provides a foundation for qFRET based K_D where two fusion proteins, CyPet-SUMO1, and YPet-UBC9 can be accurately measured for their K_D. The two fusion proteins are expressed in *E. coli* and purified for in-vitro reactions. The K_D assay is setup with CyPet-SUMO1 fusion protein kept at a singular concentration. Followed by a titration of the partner fusion protein YPet-UBC9 across a series of qFRET reactions.[23] Each singular reaction can be measured on a benchtop fluorometer for each of the Em_{FRET} signals. Compared to the ITC and the SPR techniques this method does not require any specialized equipment and the applicability of the assay onto a fluorescence plate reader is seamless.

The qFRET platform is versatile in its ability to also be applied to qFRET based biosensors. The SUMO1 and UBC9 interactions are a non-covalent transient protein-

protein interaction the work reported by Liu *et al.* provide an example of qFRET reporter for enzymatic activity. The SUMO endopeptidase, Sentrin/SUMO specific protease (SENP) cleaves pre-SUMO to activate SUMO and is responsible for removing SUMO from a target. The qFRET method required no modification of the purified proteins, required no specialized equipment such as a fluidics module to run the assay. Additionally, only the substrate is tagged with the fluorescent protein, the enzyme is completely untagged and solubilized in solution, to mimic biological activity as closely as possible. By observing the Em_{FRET} signal decrease we can interpret the enzymatic activity and determine k_{cat}/K_m across multiple substrates. This application showcases the flexibility of the qFRET platform and its accessible application to a variety of protein-protein interaction assays.

Tabulated below provides a summary of the advantages and disadvantages of the three techniques outlined. The SPR technique has reported a k_{cat} or K_m by coupling mass spectrometry analysis following binding kinetics. Each technique is advantageous in its unique method, however the qFRET technique is the most accessible in comparison. The range of qFRET applications and robust assays make it a go-to technique for rapid assessment of protein-protein interaction studies in-vitro.

Table 3: List of protein-protein interaction detection techniques, with their advantages and disadvantages.

Technique	Surface Plasmon Resonance (SPR)	Isothermal Titration Calorimetry (ITC)	Quantitative Förster Resonance Energy Transfer (qFRET)
Advantages	- Sub nanomolar detection - Resolve K_D/K_i	- Solubilized - No tag required - Resolve $K_D/K_i/k_{cat}/K_m$	- Both soluble and immobilized assay - In-cell PPI - Resolve $K_D/Ki/k_{cat}/K_m$
Disadvantages	- Immobilization - Protein modification - Specialized instrument	- High concentration of protein required - Specialized instrument	-Fluorescent tag required

1.6 qFRET Assay for the Assessment of Programmed Death 1 (PD1) and Programmed Death Ligand 1 (PDL1) Interaction

The non-covalent interaction between cell surface proteins is a crucial part of cellular function and has implications on the cell life cycle. The second chapter in this thesis applies the qFRET platform onto the interaction between the Programmed Death 1 (PD1) and Programmed Death Ligand 1 (PDL1) cell surface proteins. These two proteins are part of an immune checkpoint mechanism between the immune cell presenting the PD1 protein and an antigen presenting cell presenting the PDL1 protein.[25] The interaction is based on the affinity of PD1-PDL1 and the successful interaction of the two proteins sets off a cascade of events that can allow diseases to circumvent our immune system.

Due to the implications of this non-covalent protein-protein interaction, protein scientists have studied these two proteins extensively.[26–28]

Table 4: Dissociation constant K_D measured across instruments of PD1-PDL1 external domain. FlowC.(Flow Cytometry),Exp Host(Expression Host), Post Mod.(Modification)

Authors	**Flow C. µM**	**SPR µM**	**ITC µM**	**BLI µM**	**Year**	**Exp. Host**	**Domain**	**Post Mod.**
Youngnak *et al.*	0.526	0.114	-	-	2003	*Human* Jurkat Cells	External Domain	Amino Coupling
Butte *et al.*	-	0.77	-	-	2008/2013	COS Cells	External Domain	Amine Coupling
Cheng *et al.*	-	8.2	2.2	-	2013	*E. coli* BL21(DE3)	External Domain	Biotinylated
Tang *et al.*	-	-	-	4.1	2019	*Human* Expi293F	External Domain	Amine Coupling

The table above lists the previously measured interaction of the external domains of PD1 and PDL1. The historical study completed in 2003 that first demonstrated the difference in affinity of PDL1 and PDL2 to PD1. The group utilized two sets of technologies, flow cytometry and SPR on part of the protein, and found affinity for PD1, and found the measured K_D to be sub micromolar. This observation is unique as more recent measurements of this interaction have a much higher K_D. The highest observed is 8.2 µM and the lowest being 0.77 µM measured on SPR. The Biolayer interferometry techniques measured a K_D of 4.1 µM. Biolayer interferometry (BLI), a technique that applies similar principles as SPR however implemented with throughput in mind and use an array of "tip" sensors without a fluidics module. However, it has been criticized for the reproducibility of results, and is still in the developmental phase of applications.[29]

All three SPR measurements require immobilization of one of the interacting pairs onto the sensor surface, as the method used is amino coupling which non-

specifically attaches the amino backbone to the sensor surface. The alternative method used is biotinylating of one of the interacting proteins, which allows a streptavidin coated sensor surface to immobilize the biotinylated protein. Both amino coupling and biotinylating modifications are non-specific and add an unknown variable in the assessment of the PD1-PDL1 interaction with SPR.

The evaluation of PD1-PDL1 using ITC the group observed a K_D of 2.2 µM, an affinity of PD1-PDL1 four-fold higher than the one measured using SPR. This supports the claim that the immobilization of one of the proteins can be a factor in the difference in observed K_D as ITC does not immobilize either protein to any surface. However, the resolution of ITC is hindered by the amount of protein used, where SPR can be sub-nanomolar, ITC cannot and requires a high concentration of bound product to make accurate readings. In this work full length PD1 and PDL1 proteins are expressed and qFRET based K_D is applied and resolved dissociation constant, K_D of 816 nM with standard error of 130 nM is reported. Additionally, a qFRET based competition assay is developed to observe interaction inhibitors of PD1-PDL1. The impact of PD1-PDL1 inhibitors on cancer treatments and the demand for novel or improved inhibitors is very high. The qFRET platform provides an elegant benchtop assay to approach challenges like the characterization of PD1-PDL1 interaction.

1.7 SUMOylation Post Translation Modification

The discovery of the post-translation modification (small ubiquitin like modifier) SUMOylation is relatively new, in the early 1990s several groups discovered a 12 kDa small protein modification that regulated protein function. A group from the Howard Huges Medical Institute, reported SUMOylation of GTPase-activating protein for Ran, RanGAP1[30]. The group discovered that SUMOylation is a reversible covalent attachment to the C-terminal of RanGAP1 and observed to modulate the translocation of RanGAP1 to the nucleus, and RanGAP1 association with nuclear pore complex. Similarly, another group, out of Los Lamos New Mexico, identified that RAD51/RAD52 proteins are modified by what they described as Ubiquitin-like 1 protein, and found several proteins to have the same modification.

Furthermore, the group out of New Mexico identified that this modification plays a role in mitosis and DNA repair and recombination.[31] The role of SUMOylation PTM is then investigated across the scientific community and found to be as ubiquitous as Ubiquitin. The SUMO proteins occur in four paralogs throughout the human proteome, the SUMO1, 2 and 3 variants can be found throughout the body, but the SUMO4 variant is found predominantly in the kidney.[32] SUMOylation plays an essential role in cellular process and homeostasis, that include DNA damage repair, transcription regulation, translocation, and the modified protein interaction with the proteome.[33–35]

SUMOylation Enzymatic Cascade

The SUMOylation enzymatic cascade is the SUMO activation, conjugation, and ligation onto a target protein. The mechanism of SUMOylation is well understood through rigorous investigations in the enzymes involved in the covalent attachment of this PTM. The PTM SUMO covalently attaches to lysine residue of a target protein with help of a cascade of enzymes E1 activating enzyme, E2 conjugating enzyme, and E3 ligase. The E1 activating enzyme a heterodimer consists of the proteins AOS1 and UBA2. The E1 activating enzyme is first adenylated by ATP and will recruit SUMO onto a conserved cysteine-173 creating a temporary thioester bond with C-terminal glycine-98 of SUMO.[36] The E1-SUMO complex shuttles the SUMO onto the catalytic cysteine-93 of Ubiquitin Conjugating enzyme 9 (UBC9) the SUMO E2 conjugating enzyme. The E2 conjugating enzyme and the E3 ligase interact and shuttle SUMO onto the target protein.[37] This enzymatic cascade is illustrated in figure 3.

In comparison to Ubiquitin, not as many proteins have been identified to have SUMO E3 ligase activity. Like the ubiquitin E3 ligase mechanism, the complex formation of SUMO bound UBC9 and the E3 ligase collected target initiates the SUMO interaction with the really interesting new gene (RING) domain of the E3 ligase. The RING domain shuttles the SUMO to the lysine of the target protein in a mechanism that has been observed to transfer SUMO from the E2 directly or indirectly onto the target protein. The most notable E3 ligase is the protein inhibitor of activated STAT 1 (PIAS1), discovered by a team out of UCLA, Lui Bin and Liao Jiayu *et al.*.[38] PIAS1 E3 ligase of SUMO is observed to express in high levels in breast cancer patients and is regarded as a

bio-marker for breast cancer.[38,39] Recent proteomics on PIAS1 identified 62 substrates for PIAS1 mediated SUMOylation.[40] The SUMOylation cascade results in the target protein lysine terminal amine group forming a isopeptide bond to the carboxyl group on the C-terminal glycine.

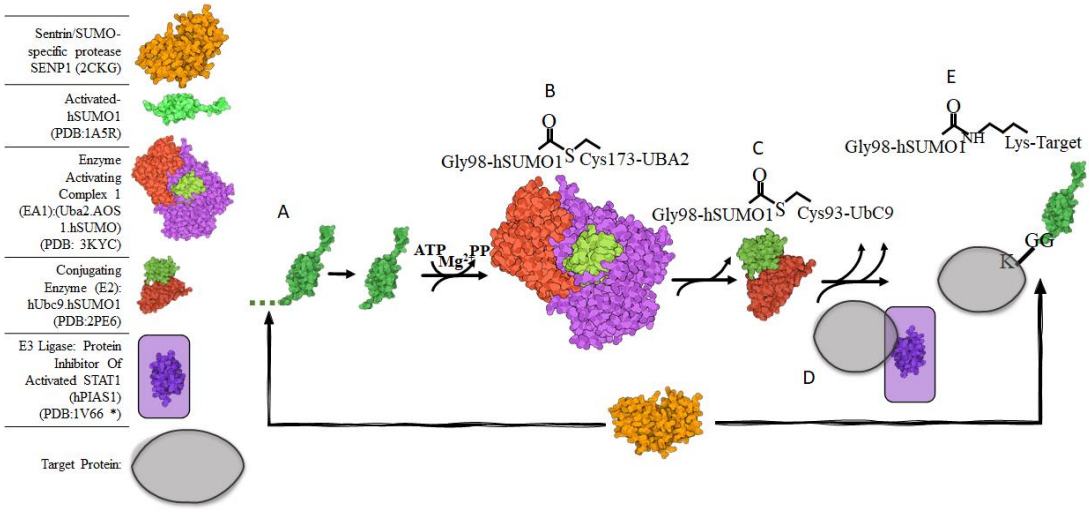


Figure 3:SUMOylation cascade illustrated here with each step in the cascade. A.) The Pre-SUMO is processed by the SENP enzyme, and the di-glycine motif is exposed. B.) The E1 Enzyme activating Complex 1, a heterodimer, UBA2 and AOS1, are adenylated and SUMO attaches to UBA2 with a temporary thioester bond between SUMO-Gly98 and UBA2-Cys173. C.) E1 shuttles SUMO onto E2 conjugating enzyme, onto the Cys93. D.) The E3 ligase shown to have affinity to the target protein, and the RING domain, shown in purple, is a key component in the mechanism of SUMO attachment to the target protein. E.) Target protein is SUMOylated, by covalent attachment to a lysine residue at the di-glycine motif. PDB images were converted to illustration using the online server
illustrate.[45]

The lysine residue location has been noted to follow a sequence motif of an upstream neighboring hydrophobic residue (ψ), and downstream of any amino acid, (x) next to an aspartic acid (D) or glutamic acid (E) residue. The current motif follows ψKxD/E and has been observed in numerous in-cell SUMOylation events.[41–43] The

reversibility, or de-SUMOylation function falls on the protease Sentrin/SUMO-specific protease (SENP) which cleave SUMO at the C-terminal of the SUMO again exposing the diglycine motif. SENP functions to also process pre-SUMO to mature SUMO protein for activation and attachment.[44]

1.8 Current Methods for Investigating SUMOylation of Viral Proteins

The overall cellular modulation of SUMOylation associated with the Influenza A virus (IAV) infection has been observed by two groups.[46,47] The observation at the proteome level by these groups provide a broad overview of the intersection of viral proteome and the host human proteome. Furthermore, as the modulation of SUMOylation occurred only during infection sparked a series of investigations in the viral proteome's deliberate modulation of the SUMOylation mechanism. For example, proteomic studies of endogenous proteins such as TRIM28, a known SUMO E3 ligase, and is found to be SUMOylated under native conditions, however during IAV infection the protein modification levels are significantly lower. The loss of SUMOylated TRIM28 results in the inhibition of human antiviral response by inhibiting the IFN-mediated autoimmune response. The IFN-mediated autoimmune response is part of the host defense against viral RNA.[48]

The identification of SUMOylated proteins during viral infection became the center of focus for numerous investigations of host and virus proteome interaction. The current methods for isolation of SUMOylated products from cell lysate have been very useful in the primary discovery of the phenomenon. However, a need for mechanistic

understanding of the SUMOylation site on the protein and its outcome is the natural next step in the understanding of SUMOylation and viral proteome interaction.

Immunoblotting for Identification of Covalent Modification

The first of the broader proteome SUMOylation study was the work done by Pal S. *et al.* group, in 2011. The investigation involved extensive immunoprecipitation with western blots and the group confirmed instances of the Influenza Viral proteome heavily SUMOylated during infection. The study had also completed in-vitro SUMOylation reactions of all the Influenza A viral proteins reported by immunoblots. The in-vitro reactions are ran on a sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) and analyzed by size. The SDS-PAGE process separates the proteins within the acrylamide gel by size in the vertical direction, as top of the gel are larger proteins and the bottom are smaller proteins. The band shift in the vertical direction in the immunoblots are assumed to be mass increase in the target protein due to SUMOylation. For the in-vitro SUMOylation events, RNA-directed RNA polymerase catalytic subunit 1 (PB1), Nucleoprotein (NP), Non-Structured protein 1 (NS1), Non-Structured Protein 2 (NS2), and Matrix 1 protein (M1) were observed to be SUMOylated.[47] However, the in-cell methods used to elucidate these SUMOylation events, are under scrutiny as SUMO is overexpressed along with the viral genes within the human cell lines and no information of lysine site of SUMOylation was resolved.

"Slice-By-Slice" Method for Identifying Modified Proteins

The work published by Domingues *et al.*, observed a profound difference in the SUMOylation activity on endogenous proteins involved in, DNA repair, transcription, and RNA quality control mechanism were found to be SUMOylated during infection. Additionally, the study observed the SUMOylation of IAV, Non-Structured protein 1 (NS1), Non-Structured Protein 2 (NS2), and Matrix 1 protein (M1). The method used by Domingues *et al.* was to identify SUMOylated proteins directly by mass spectrometry coupled with 1D SDS-PAGE gel, called "slice-by-slice". The extraction and purification of SUMOylated proteins was through Tandem-affinity-purified (TAP) proteins.

Tandem-affinity-purification is a two-part affinity purification to select proteins from cell lysate used in many studies to identify modified proteins.[46,49] The technique is based on the principles of immunoprecipitation of target proteins. The Domingues *et al.* group expressed an exogenous SUMO protein fused with the TAP tag, that contains two epitopes, the calmodulin-binding peptide (CBP) and Staphylococcus aureus (Protein A) with a Tobacco Etch Virus (TEV) proteasomal cut site between the two. The exogenous SUMOylated proteins are first captured by an Ig bound beads with affinity towards the Protein A tag. After the first round of purification, the protein A tag is cleaved through protease TEV cut and the complex is released, followed by binding to calmodulin beads, which have an affinity to CBP. The two rounds of purification yield a complex of proteins that are SUMOylated or are non-specific proteins bound to the complex. This method of purification is arduous and the yield of SUMOylated proteins is often low,

however it is a widely used method to retrieve a complex of proteins from within cell lysate.

The two-step purification is necessary due to the nature of in-cell SUMOylation and purification process. The initial in-cell SUMOylation occurs with the over expression of an exogenous TAP-SUMO in the cell. Due to the high concentration of exogenous TAP-SUMO, the probability of binding to an exogenous-over expressed-SUMO is much higher than the native SUMO. Furthermore, once the cells are lysed the SUMOylated products tend to form a complex with non-specific proteins. The non-specific proteins result in noise within the downstream mass-spectrometry analysis. Thus, a secondary purification step, CBP-Calmodulin, is applied to purify the SUMOylated complex further and isolate the SUMO bound target proteins.[46,50] Applications of this method have also found that the SUMOylated proteins are at a much lower concentration compared to the unmodified proteins, this major drawback is in part due to the high de-SUMOylation activity by SENP within cell lysate.[51]

The next step in the process is the identification of SUMOylated protein using mass-spectrometry. The method used in these studies is called “slice-by-slice”, and it is a process where the TAP purified proteins are ran on an SDS-PAGE and Coomassie stained. The SDS-PAGE process separates the proteins within the acrylamide matrix by size in the vertical direction, each stain band is assumed to be a target protein shifted in sized due to SUMO modification. The gel is then cut into many slices along the vertical axes and each section of the gel is sent for mass-spectrometry for protein identification.

The indirect method of identification of SUMOylation occurs at this point. The slices of the gel represent changes in mass of the protein due to the covalent attachment of SUMO. Thus, if a protein is detected above the expected mass of the protein, at a higher slice, it is assumed to be SUMOylated. The slice-by-slice method of identifying SUMOylated proteins is applicable for a general overview of identifying protein modification. However, the technique lacks the resolution to determine which lysine is modified, and due to the in-cell SUMOylation event, not every possible SUMOylation site can be detected as other possible lysine modifications can replace SUMO, such as Ubiquitination, NEDDylation, or acetylation.

Mutant SUMO for identification of SUMOylation site

The in-cell SUMOylated purified target protein yields are very low, and the downstream processing of the modified proteins results in very limited possibilities for analysis. Post purification, the proteins can be directly digested by a protease to create peptides for MS fragmentation and identification. However, trypsin digestion of the SUMO protein for MS analysis results in a very large peptide of 19-32 amino acids that is not ideal for identification with MS.[43] Trypsin cleaves the C-terminal of the lysine K and arginine R residues. The SUMO glycine isopeptide bond formed to the lysine of the target protein protects the lysine from tryptic digestion and cleaves the covalently bound SUMO. The C-terminal cleaved peptide can be found very low concentrations and MS fragmentation of that peptide makes it very difficult to identify. The figure 4 below outlines the potential products formed with trypsin cleavage.

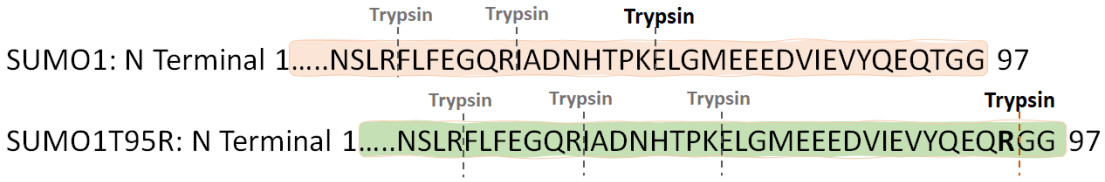


Figure 4:Predicted Trypsin digestion of C-terminal wildtype SUMO1 K88 and mutant SUMO1 R95. Other potential cleavage sites noted with gray font. Prediction to made by peptide cutter ExPASy. [52]

The substitution of arginine at the C-terminal of the SUMO protein, T95R, provides a method to overcome the pitfalls of trypsin digestion. The mutant SUMO provides the certainty of the size of peptide formed with trypsin digestion and allows for the identification of modified lysine site from in-cell SUMOylation. This method in combination with enrichment of product using immunoprecipitation make this a viable option for identifying SUMOylation sites. However, the combination of a mutant SUMO overexpressed in a cellular environment can add unknown variables that can lead to false positive and false negative lysine identifications. This method aims to overcome the digestion pitfall of wildtype SUMO proteins by the introduction of a mutant SUMO. The wildtype SUMO trypsin digestion can produce a fragment, but the large size of the peptide in combination of the low concentration of modified protein in the sample is the main disadvantage of most SUMO pull down techniques.

Issues with SUMO Overexpression

The current methods to improve identification of SUMOylation in cell can tend to produce false positives by over expression of exogenous SUMO, false negative by deSUMOylation during purification, and eventual cost of time and money on product enrichment and multiple MS analysis. Each method outlined required SUMO to be over

expressed within the cell during the experiment. The overexpression of exogenous SUMO improved the chances of isolating a SUMOylated product, as the native SUMO levels are not always enough for downstream analysis with MS. Over expression of SUMO within a cellular environment can have unintended outcomes which can compromise an investigation. For example, the over expression of SUMO paralogs results in overall change in the proteomic profile of the cell due to cross talk of SUMO with ISG.[53]

A separate study found that overexpression of SUMO3 within the cell modulate the interferon production. The over expression of SUMO3 decreased interferon alpha induced STAT1 phosphorylation, a mechanism critical in innate immunity response.[54] The innate immunity of a cell in response to a viral infection is an important aspect of anti-viral activity. The modulation of these cellular mechanisms can ultimately burden investigations of viral infection and SUMOylation. Alternatively with the bulk identification of possible SUMOylated targets, *in vitro* techniques provide a clear assessment of SUMOylated lysines on proteins.

1.9 *In vitro* qFRET SUMOylation assay for the Identification of Modified Lysines Using Mass Spectrometry

The third and fourth chapter of this thesis outline the systematic assessment of the interaction of SUMOylation cascade on heterologous viral proteins with qFRET as a reporter. The SUMO covalent modification has been a target of viral proteome and has been "hijacked" by disease to progress in its pathogenesis.[34,55] Though slice-by-slice method of identifying SUMOylated proteins is applicable for a broad overview of

SUMOylated proteins. The follow up studies require a robust method for identifying all possible SUMOylation sites without hindrance of cellular mechanisms. In addition, yield high concentration of SUMOylated product that can be directly transferred to mass spectrometry analysis.

Demonstration in this work the qFRET platform is applied as a reporter for in vitro SUMOylation of the viral proteins matrix 1 (M1) from the influenza A virus and nucleocapsid (N) protein from the severe acute respiratory syndrome coronavirus 2 (SARS-CoV 2).[34,35,55] The result of the *in vitro* assay are the qFRET report of SUMO covalent attachment, and the SUMO modified product for downstream MS analysis. The qFRET assay is nondestructive and can be readily applied to secondary analysis techniques such as MS, immunoblots, or other secondary assays for broader investigations.

In the work presented we applied the qFRET plus MS assay to both the IAV M1 protein and SARS-CoV-2 N protein. Influenza A virus M1 proteins was observed to have 5 lysine residues modified. The SARS-CoV-2 N protein was found to have 4 lysine residues modified. For both protein's lysine mutants were evaluated and found to have an impact in the viral pathogenesis for M1 and found to have modulation of in cell activity with N protein mutants. The assays outlined in chapters 3 and 4 provide a framework for rapid assessment of modified lysine sites from *in vitro* SUMO modification.

1.10 References

1. Nurk, S. *et al.*. The complete sequence of a human genome. bioRxiv 2021.05.26.445798 (2021) doi:10.1101/2021.05.26.445798.

2. Adhikari, S. *et al.*. A high-stringency blueprint of the human proteome. Nat Commun **11**, 5301 (2020).

3. Rolland, T. *et al.*. A proteome-scale map of the human interactome network. Cell **159**, 1212–1226 (2014).

4. Orgogozo, V., Morizot, B. & Martin, A. The differential view of genotype–phenotype relationships. Front. Genet. **0**, (2015).

5. Kim, Y.-A., Cho, D.-Y. & Przytycka, T. M. Understanding Genotype-Phenotype Effects in Cancer via Network Approaches. PLoS Comput Biol **12**, e1004747 (2016).

6. Yesudasu, V., Pradhan, H. S. & Pandya, R. J. Recent progress in surface plasmon resonance based sensors: A comprehensive review. Heliyon **7**, (2021).

7. Pattnaik, P. Surface Plasmon Resonance: Applications in Understanding Receptor–Ligand Interaction. ABAB **126**, 079–092 (2005).

8. Vachali, P. P., Li, B., Bartschi, A. & Bernstein, P. S. Surface plasmon resonance (SPR)-based biosensor technology for the quantitative characterization of protein–carotenoid interactions. Archives of Biochemistry and Biophysics **572**, 66–72 (2015).

9. Nguyen, H. H., Park, J., Kang, S. & Kim, M. Surface Plasmon Resonance: A Versatile Technique for Biosensor Applications. Sensors (Basel) **15**, 10481–10510 (2015).

10. Gupta, B. D. & Sharma, A. K. Sensitivity evaluation of a multi-layered surface plasmon resonance-based fiber optic sensor: a theoretical study. Sensors and Actuators B: Chemical **107**, 40–46 (2005).

11. Saboury, A. A. A review on the ligand binding studies by isothermal titration calorimetry. JICS **3**, 1–21 (2006).

12. Johnson, C. M. Isothermal TitrationTitrations Calorimetry. in Protein-Ligand Interactions: Methods and Applications (eds. Daviter, T., Johnson, C. M., McLaughlin, S. H. & Williams, M. A.) 135–159 (Springer US, 2021). doi:10.1007/978-1-0716-1197-5_5.

13. Falconer, R. J., Schuur, B. & Mittermaier, A. K. Applications of isothermal titration calorimetry in pure and applied research from 2016 to 2020. Journal of Molecular Recognition **n/a**, e2901.

14. Abis, G., Pacheco-Gómez, R., Bui, T. T. T. & Conte, M. R. Isothermal Titration Calorimetry Enables Rapid Characterization of Enzyme Kinetics and Inhibition for the Human Soluble Epoxide Hydrolase. Anal. Chem. **91**, 14865–14872 (2019).

15. Liao, J., Song, Y. & Liu, Y. A new trend to determine biochemical parameters by quantitative FRET assays. Acta Pharmacol Sin **36**, 1408–1415 (2015).

16. Algar, W. R., Hildebrandt, N., Vogel, S. S. & Medintz, I. L. FRET as a biomolecular research tool — understanding its potential while avoiding pitfalls. Nat Methods **16**, 815–829 (2019).

17. Sun, Y., Wallrabe, H., Seo, S.-A. & Periasamy, A. FRET microscopy in 2010: The legacy of Theodor Förster on the 100th anniversary of his birth. Chemphyschem **12**, 462–474 (2011).

18. Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. Annalen der Physik **437**, 55–75 (1948).

19. Heim, R. & Tsien, R. Y. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. Current Biology **6**, 178–182 (1996).

20. Jones, J., Heim, R., Hare, E., Stack, J. & Pollok, B. A. Development and Application of a GFP-FRET Intracellular Caspase Assay for Drug Screening. J Biomol Screen **5**, 307–317 (2000).

21. Evolutionary optimization of fluorescent proteins for intracellular FRET | Nature Biotechnology. https://www.nature.com/articles/nbt1066?platform=hootsuite.

22. Erickson, H. P. Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy. Biol Proced Online **11**, 32–51 (2009).

23. Song, Y., Madahar, V. & Liao, J. Development of FRET assay into quantitative and high-throughput screening technology platforms for protein-protein interactions. Ann Biomed Eng **39**, 1224–1234 (2011).

24. Chen, H., Puhl, H. L., Koushik, S. V., Vogel, S. S. & Ikeda, S. R. Measurement of FRET Efficiency and Ratio of Donor to Acceptor Concentration in Living Cells. Biophys J **91**, L39–L41 (2006).

25. Zak, K. M. *et al.*. Structural Biology of the Immune Checkpoint Receptor PD-1 and Its Ligands PD-L1/PD-L2. Structure **25**, 1163–1174 (2017).

26. Patel, S. P. & Kurzrock, R. PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy. Mol Cancer Ther **14**, 847–856 (2015).

27. Brown, M. E. *et al.*. Assessing the binding properties of the anti-PD-1 antibody landscape using label-free biosensors. PLOS ONE **15**, e0229206 (2020).

28. Ganesan, A. *et al.*. Comprehensive *in vitro* characterization of PD-L1 small molecule inhibitors. Sci Rep **9**, 1–19 (2019).

29. Weeramange, C. J., Fairlamb, M. S., Singh, D., Fenton, A. W. & Swint-Kruse, L. The strengths and limitations of using biolayer interferometry to monitor equilibrium titrations of biomolecules. Protein Science **29**, 1004–1020 (2020).

30. Matunis, M. J., Coutavas, E. & Blobel, G. A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. Journal of Cell Biology **135**, 1457–1470 (1996).

31. Shen, Z., Pardington-Purtymun, P. E., Comeaux, J. C., Moyzis, R. K. & Chen, D. J. UBL1, a Human Ubiquitin-like Protein Associating with Human RAD51/RAD52 Proteins. Genomics **36**, 271–279 (1996).

32. Sumoylation on its 25th anniversary: mechanisms, pathology, and emerging concepts. https://febs.onlinelibrary.wiley.com/doi/epdf/10.1111/febs.15319.

33. Müller, S., Ledl, A. & Schmidt, D. SUMO: a regulator of gene expression and genome integrity. Oncogene **23**, 1998–2008 (2004).

34. Gareau, J. R. & Lima, C. D. The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. Nat Rev Mol Cell Biol **11**, 861–871 (2010).

35. Eifler, K. & Vertegaal, A. C. O. Mapping the SUMOylated landscape. The FEBS Journal **282**, 3669–3680 (2015).

36. Johnson, E. S., Schwienhorst, I., Dohmen, R. J. & Blobel, G. The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. EMBO J **16**, 5509–5519 (1997).

37. Desterro, J. M. P., Thomson, J. & Hay, R. T. Ubch9 conjugates SUMO but not ubiquitin. FEBS Letters **417**, 297–300 (1997).

38. Chanda, A. *et al.*. Identification of the SUMO E3 ligase PIAS1 as a potential survival biomarker in breast cancer. PLOS ONE **12**, e0177639 (2017).

39. Galanty, Y. *et al.*. Mammalian SUMO E3-ligases PIAS1 and PIAS4 promote responses to DNA double-strand breaks. Nature **462**, 935–939 (2009).

40. Li, C. *et al.*. Quantitative SUMO proteomics identifies PIAS1 substrates involved in cell migration and motility. Nat Commun **11**, 834 (2020).

41. Lamoliatte, F. *et al.*. Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling. Nature Communications **5**, 5409 (2014).

42. Yang, X.-J. & Seto, E. Lysine acetylation: codified crosstalk with other posttranslational modifications. Mol Cell **31**, 449–461 (2008).

43. Matic, I. *et al.*. Site-Specific Identification of SUMO-2 Targets in Cells Reveals an Inverted SUMOylation Motif and a Hydrophobic Cluster SUMOylation Motif. Molecular Cell **39**, 641–652 (2010).

44. Kunz, K., Piller, T. & Müller, S. SUMO-specific proteases and isopeptidases of the SENP family at a glance. J Cell Sci **131**, jcs211904 (2018).

45. Goodsell, D. S., Autin, L. & Olson, A. J. Illustrate: Software for Biomolecular Illustration. Structure **27**, 1716-1720.e1 (2019).

46. Domingues, P. *et al.*. Global Reprogramming of Host SUMOylation during Influenza Virus Infection. Cell Rep **13**, 1467–1480 (2015).

47. Pal, S., Santos, A., Rosas, J. M., Ortiz-Guzman, J. & Rosas-Acosta, G. Influenza A virus interacts extensively with the cellular SUMOylation system during infection. Virus Research **158**, 12–27 (2011).

48. Schmidt, N. *et al.*. An influenza virus-triggered SUMO switch orchestrates co-opted endogenous retroviruses to stimulate host antiviral immunity. PNAS **116**, 17399–17408 (2019).

49. Peng, J. & Cheng, D. Proteomic Analysis of Ubiquitin Conjugates in Yeast. in Methods in Enzymology vol. 399 367–381 (Academic Press, 2005).

50. Aitken, C. E., Marshall, R. A. & Puglisi, J. D. An Oxygen Scavenging System for Improvement of Dye Stability in Single-Molecule Fluorescence Experiments. Biophys J **94**, 1826–1835 (2008).

51. Esteras, M., Liu, I.-C., Snijders, A. P., Jarmuz, A. & Aragon, L. Identification of SUMO conjugation sites in the budding yeast proteome. Microb Cell **4**, 331–341 (2017).

52. Gasteiger, E. *et al.*. Protein Identification and Analysis Tools on the ExPASy Server. in The Proteomics Protocols Handbook (ed. Walker, J. M.) 571–607 (Humana Press, 2005). doi:10.1385/1-59259-890-0:571.

53. Cross-talk between SUMOylation and ISGylation in response to interferon | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S1043466620300417?token=BD6F094A4B82980D98E0077CF8C646639BBA0C05B8A7A65EC4B73FC458C01324A01A56BE9F5AB46EAB5E7789AA394E0F&originRegion=us-east-1&originCreation=20210816201830 doi:10.1016/j.cyto.2020.155025.

54. Maroui, M. A. *et al.*. Promyelocytic Leukemia Protein (PML) Requirement for Interferon-induced Global Cellular SUMOylation. Mol Cell Proteomics **17**, 1196–1208 (2018).

55. Boggio, R. & Chiocca, S. Viruses and sumoylation: recent highlights. Current Opinion in Microbiology **9**, 430–436 (2006).

Chapter 2: The Development of QFRET Based Assay for the Interaction of PD1 and PDL1

2.1 The Biological Impact of PD1 interaction with PDL1

The immune checkpoint inhibitors that target the regulation mechanism of our immune response to cancer cells have revolutionized cancer therapy.[1–3] The inhibition of activated T cell immune response is initiated by a coinhibitory receptor Programmed death 1 (PD1) expressed on the T cell surface. Programmed death 1 (PD1) is a transmembrane glycoprotein receptor with an immunoglobulin (IgV) like extracellular domain and a cytoplasmic domain with an immune inhibitor switch motif that plays a part in the downstream immune regulation once bound. The programmed death ligand 1 (PDL1), is a type 1 transmembrane glycoprotein that is composed of IgC type extracellular domains and is expressed on the surface of antigen presenting cells (APC) that bind to PD1 on the T cell.[3–5] The interaction of PD1-PDL1 initiates the downstream inhibition of T cell receptor (TCR) signaling, and inhibits the release of cytokines such as interleukin 2 (IL-2) to exhaust the T cell response.(Figure 5)

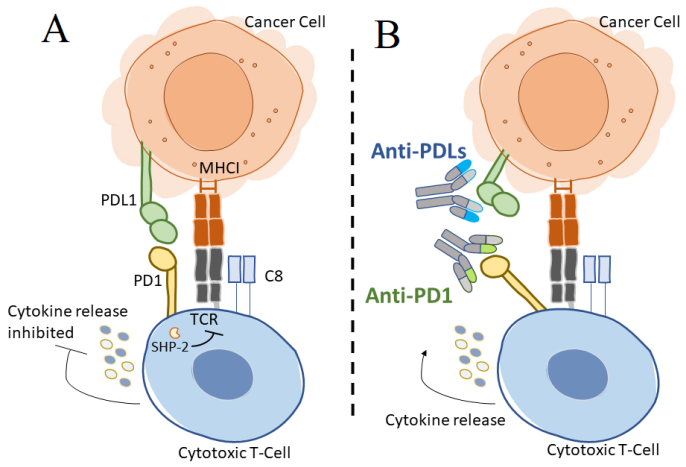


Figure 5: A.) The interaction of PD1 with PDL1 or 2 exhausts the T-Cell receptor (TCR) response and allows the cancer cells to survive. The T cell surface receptor contacts the major histocompatibility complex I surface protein. The PD1 also interacts with the PDL1 protein, and downstream signaling by TCR is inhibited by the recruitment of SHP-2. In contrast to when PD1 does not interact with PDL1 or 2 the SHP-2 is not recruited and does not inhibit the TCR response. B.) The anti-PD1 or anti-PDL mAbs block PD1 and PDL from interacting. PD1 can no longer recruit SHP-2 and the T-Cell activates to destroy the cancer cell.

Adaptive immune resistance and treatment toxicity has been observed in tumor cells with increased induction of PDL1 expression to exhaust the TCR response to cancer cells.[6–8] Furthermore, trends in clinical trials over time for monoclonal antibody treatment leading up to 2020 show an increase in combinatory therapies, a decrease in mono-therapies, and ultimately very few reaching FDA approval.[9,10] The efficacy of the treatment varies with numerous factors such as age of the patient and type of cancers. A clinical study reports failure to treat brain cancer in pediatric patients using FDA approved anti-PD1 antibodies, as type of cancer and patient age were noted as factors.[11] Highlighting the need for continued development on check point inhibitors and of novel and viable protein-protein interaction inhibitors of PD1-PDL1 for use in cancer therapy. Alternative methods for assessing protein-protein interactions are surface plasmon

resonance (SPR) and isothermal titration calorimetry (ITC). Both SPR and ITC require extensive specialized equipment for implementation, in comparison to the qFRET platform which requires a bench top fluorometer.[12,13] Additionally, SPR is a low throughput assay that requires the immobilization of one of the interacting proteins onto a functional surface, which is not ideal for investigating protein-protein interaction. In comparison the qFRET assay is implemented with both fusion proteins solubilized in solution in a 384 well format that can be readily scaled up for large scale inhibitor screening.[14] We demonstrate a robust qFRET assay for the assessment of interacting proteins PD1 and PDL1 to resolve their dissociation constant, K_D, and the characterization of PD1-PDL1 interaction inhibitors.

2.2 PD1-PDL1 qFRET Assay Principle and Design

The qFRET technology leverages the Förster Resonance Energy Transfer (FRET) phenomenon for the assessment of two biomolecules interaction within the range of 1 – 10 nm.[15] FRET is a non-radiative energy transfer that occurs by dipole-dipole coupling interaction of an excited fluorescent molecule, donor, to a ground state fluorescent molecule, acceptor, only when both are within 1 to 10 nanometers of each other. [16,17] The FRET efficiency relationship depends on the inverse-sixth power distance between the acceptor and donor, making this phenomenon sensitive to the distance between the donor and acceptor. The reported Förster distance R_0 of 5.15 nm of CyPet-YPet falls within range of 1-10 nm protein-protein interaction.[18,19] The proposed quantitative application is designed to report the interaction of fusion protein, CyPet-PD1 (acceptor fusion protein), with YPet-PDL1 (donor fusion protein). The fusion protein design is illustrated in figure

6A, with the fluorescent protein on the N-terminal and the interacting proteins on the C-terminal. The Em_{FRET} signal is extracted by post processing the measured EmTotal signal following the Em_{FRET} equation in figure 6C. The EmTotal is the unprocessed fluorescent signal at the FRET emission wavelength that has the donor, CyPet, and acceptor, YPet, fluorescent crosstalk embedded (figure 6B). In the implementation of this assay three measurements are taken at each reaction, EmTotal, the unbound CyPet-PD1 (FL_D), and the unbound YPet-PDL1 (FL_A). Tabulated in figure 6B the EmFRET wavelength is the excitation of the donor and the emission of the acceptor and FL_D and FL_A are specific to the donor and acceptor. The Em_{FRET} equation figure 6C, provides the post processing equation for the accurate extraction of Em_{FRET} signal and removal of donor and acceptor cross talk contributions. The resulting Em_{FRET} signal is the extracted FRET response from the interaction of our two proteins of interest without crosstalk from CyPet and YPet.

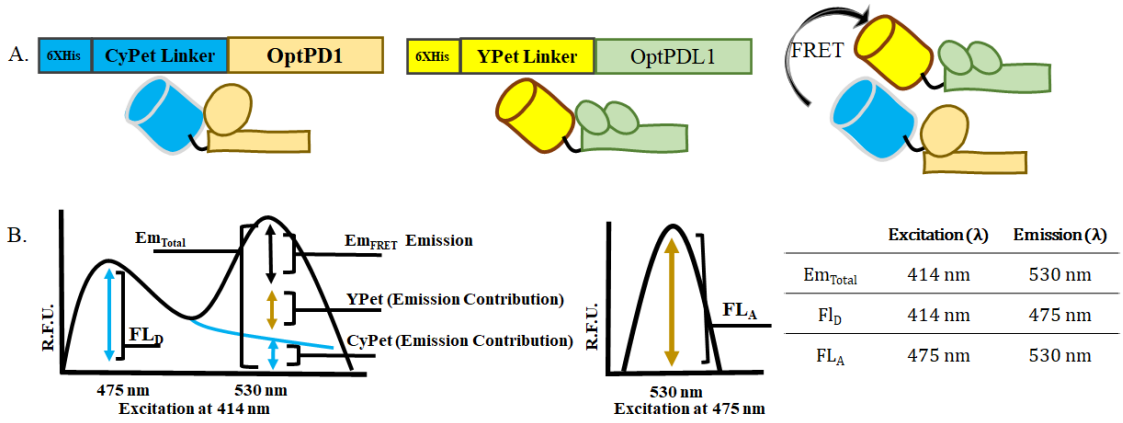


Figure 6:A.) qFRET design is the donor fluorescent protein CyPet fused to OptPD1, and acceptor fluorescent protein YPet fused to OptPDL1. B.) Data is collected at three wavelengths, used to extract the Em_{FRET} signal at 530 nm, directly related to the interaction of PD1 to PDL1. Em_{FRET} is calculated by subtracting the fluorescent protein emission crosstalk at the FRET wavelength shown. C.) Em_{FRET} equations is used to extract the Em_{FRET} signal, and D.) Alpha α ratio for the donor is calculated by dividing the emission at 530 nm by emission at 475 nm when excited at 414 nm. The β parameter is determined by diving the emission of YPet at 530 nm when excited at 414 nm over the emission at 530 nm when excited by 475 nm.

PD1-PDL1 qFRET Alpha (α) and Beta (β) cross talk coefficients

The primary assessment in the setup of the qFRET assay is evaluation of the donor and acceptor pair's fluorescence cross talk coefficients parameters alpha (α) and beta (β). The alpha (α) and beta (β) parameters are utilized in this method to extract the Em_{FRET} signal from the measured EmTotal fluorescence and evaluate the fluorescence from each protein to the fluorometer. The equations shown in figure 6D, we define the alpha (α) and beta (β) values for both the donor and acceptor. We determined the two constants by measuring the fluorescence of the two molecules individually, at constant concentration, across the FRET wavelength and dividing it by the fluorescent protein's

own emission wavelength. We observed the crosstalk parameter of donor CyPet to be at 0.34 (α), standard deviation of 0.003, (n=9). The accepter YPet to be at 0.03 (β) standard deviation of 0.001, (n=9)

2.3 qFRET dissociation constant (K_D)

The qFRET based dissociation constant is determined by analysis of qFRET signal across a titration of concentration of the non-covalent interaction of proteins CyPet-OptPD1 and YPet-OptPDL1. The dissociation constant, K_D, is a widely utilized metric for evaluating affinity of two interacting molecules at equilibrium. The schematic of the receptor [R] and ligand [L] binding to form the product [RL] at equilibrium we define the dissociation constant K_D, Equation 5. We design the qFRET assay to observe the change in Em_{FRET} signal with the change in ligand concentration while keeping the receptor concentration constant. We titrate the concentration of ligand till all the receptor is bound, and this will result in the plateau of the Em_{FRET} response. Based on this experiment design we can state that in the maximum formed product PD1-PDL1, $[RL]_{Max}$, is equivalent to the total receptor concentration, $[R]_{Total}$ (Equation 9). Additionally, the $Em_{FRETMax}$ will occur at the saturation of the receptor, $[RL]_{Max}$, and a ratio of Em_{FRET} observed to $Em_{FRETMax}$ is equivalent to the ratio of [RL] observed to $[RL]_{Max}$ concentration. (Equation 10). The set of equations below resolve a relationship for deriving K_D with the change in Em_{FRET} dependent ultimately on the change in ligand concentration. Given the schematic of two reactants [R] and [L] -> [RL] interacting we derive the solution for Em_{FRET} with changing $[L]_{Total}$.

$$[L]_{Total} = [L]_{Free} + [RL] \quad \text{Eq. 2.1}$$

$$[R]_{Total} = [R]_{Free} + [RL] \quad \text{Eq. 2.2}$$

$$[RL] = [L]_{Total} - [L]_{Free} \quad \text{Eq. 2.3}$$

$$[R]_{Free} = [R]_{Total} - [L]_{Total} + [L]_{Free} \quad \text{Eq. 2.4}$$

$$K_D = \frac{[R]_{Free} * [L]_{Free}}{[RL]} \quad \text{Eq. 2.5}$$

$$K_D = \frac{([R]_{Total} - [L]_{Total} + [L]_{Free}) * [L]_{Free})}{([L]_{Total} - [L]_{Free})} \quad \text{Eq. 2.6}$$

Rearrange Eq 6, $0 = [L]_{Free}^2 + [L]_{Free} * ([R]_{Total} - [L]_{Total} + K_D) - K_D * [L]_{Total}$ Eq. 2.7

Eq 7 solved for $[L]_{Free}$, $[L]_{Free} = \frac{([L]_{Total} - [R]_{Total} - K_D) + \sqrt{([R]_{Total} + K_D - [L]_{Total})^2 + 4 * K_D * [L]_{Total}}}{2}$ Eq. 2.8

Assume the maximum bound is equivalent to $[R]_{Total}$, $[RL]_{Max} = [R]_{Total}$ Eq. 2.9

Assume ratio of bound [RL] to Em_{FRET} response, $[RL] = [RL]_{Max} * \frac{Em_{FRET}}{Em_{FRETMax}}$ Eq. 2.10

$$K_D = \frac{([R]_{Total} - [RL]) * [L]_{Free}}{[RL]} \quad \text{Eq. 2.11}$$

Substitute Eq 9 for $[R]_{Total}$, $K_D = \frac{([RL]_{Max} - [RL]) * [L]_{Free}}{[RL]}$ Eq. 2.12

$$[RL] = \frac{[RL]_{Max} * [L]_{Free}}{K_D + [L]_{Free}} \quad \text{Eq. 2.13}$$

$$\frac{[RL]}{[RL]_{Max}} = \frac{[L]_{Free}}{K_D + [L]_{Free}} = \frac{Em_{FRET}}{Em_{FRET_{Max}}} \quad \text{Eq. 2.14}$$

$$Em_{FRET} = Em_{FRET_{Max}} * \frac{[L]_{Free}}{K_D + [L]_{Free}} \quad \text{Eq. 2.15}$$

$$Em_{FRET} = Em_{FRET_{ax}} * \left(\frac{[L]_{Total} - [R]_{Total} - K_D + \sqrt{([R]_{Total} + K_D - [L]_{Total})^2 + 4 * K_D * [L]_{Total}}}{[R]_{Total} + K_D - [L]_{Total} + \sqrt{([R]_{Total} - [L]_{Total} + K_D)^2 + 4 * K_D * [L]_{Total}}}\right) \quad \text{Eq. 2.16}$$

Derived here at Eq. 2.16 is the relationship between the EmFRET observed dependent on titration of $[L]_{Total}$, constant $[R]_{Total}$, constant $EmFRET_{Max}$, and constant K_D. GraphPad Prism 5™ software suit is used to fit equation 16, using the non-linear regression curve fitting tool that uses the Marquardt and Levenberg method where a mix

of linear decent and Gauss-Newton both are used to step towards the least sum-of-squares.[20] The initial conditions for $[R]_{Total}$, $[L]_{Total}$, K_D, $Em_{FRETMax}$, and constraints of the curve fit are described in detail in the Methods section.

2.4 qFRET competition assay

The qFRET competition assay provides insight into the activity of inhibitor on the interacting pair of proteins CyPet-PD1 and YPet-PDL1. In this competitive binding assay, we introduce an unlabeled molecule that competes for binding to either donor or acceptor and inhibits the formation of the [CyPet-PD1.YPet-PDL1] product. The binding event of the unlabeled molecule directly decreases the amount of product formed [CyPet-PD1.YPet-PDL1] and is observed as a decrease in Em_{FRET}. The inhibitory concentration where the Em_{FRET} response drops to half of the initial uninhibited Em_{FRET} value, is called the half maximal inhibitory concentration (IC_{50}). The IC_{50} value is resolved by the nonlinear regression fit of the Em_{FRET} response to equation 17 and is a commonly used metric to evaluate an inhibitors potency. However, the IC_{50} of an inhibitor can change depending on the concentration of the substrate, resolving the equilibrium constant for the inhibitor, K_i is an alternative approach to characterize the potency of an inhibitor. Applied here is the Cheng Prusoff relationship to resolve K_i based on IC_{50}, equation 18, we can solve both relationships together, by solving for "$LogIC_{50}$" using equation 17, applying the dissociation constant K_D, and the substrate concentration at which the assay was ran.[21,22] We solved equation 18 for IC_{50}, in log form to get equation 19, and applied non-linear regression fit to equations 17 and 19 GraphPad Prism5^{TM} to obtain K_i. The

assay details and setup such as the acquisition settings of the Em_{FRET} response, and the non-linear regression fit setup of equations 17 and 19 are outlined in the methods section.

$$Em_{FRET} = Em_{FRETMin} + \frac{(EmFRET_{Max} - EmFRET_{Min})}{1 + 10^{([I]_{Total} - LogIC50)}} \quad \text{Eq 2.17}$$

Cheng Prusoff relationship of IC_{50} to K_i, $\mathbf{K_i} = \frac{IC_{50}}{\left(1+\frac{[R]}{K_D}\right)}$ Eq 2.18

$$Log(IC_{50}) = \log\left(10^{\log(\mathbf{K_i})} * \left(1 + \frac{[R]}{K_D}\right)\right) \quad \text{Eq 2.19}$$

2.5 Results

qFRET Assay Alpha (α) and Beta (β) Results

We observed the crosstalk parameter of donor CyPet to be at 0.34 (α) standard deviation of 0.003 and a CV of 0.1 %. The accepter YPet to be at 0.03 (β) standard deviation of 0.001 and a CV of 3 % n = 9 for both parameters. The beta cross talk in comparison to alpha has a much lower cross channel signal, this is by design and provides the support for using both CyPet-YPet as the acceptor and donor pair. The fluorescent plate reader utilized in this study is made by Molecular Devices and is called SpectraMax M3™ packaged with SoftMax Pro 7™ software suite. Plotted in Prism™. The plate reader by Molecular Devices, Flexstation II 384(™), the alpha and beta values, for CyPet and YPet were determined to be 0.360 (α) and beta of 0.026 (β) in a previous publication by Song *et al.*.[23]

Table 5: Listed mean, standard deviation, Coefficient of variation over 9 measurements at constant concentration of both acceptor and donor.

Parameter	Alpha (α)	Beta (β)
Mean	0.34	0.03
Standard Deviation	0.003	0.001
Coefficient Of Variation	0.001	0.03
n	9	9

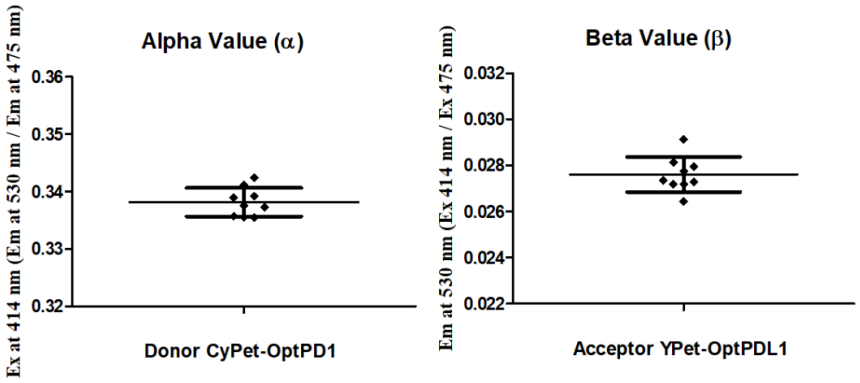


Figure 7:The scatterplots shown with standard deviation T bars, and average denoted with a line. Molecular Devices SpectraMax3 was used to measure the fluorescence response for both acceptor and donor molecule.

Codon optimization and expression of CyPet-PD1 and YPet-PDL1

The human programmed death 1 and human program death ligand 1 and both can be classified as difficult-to-purify proteins.[24] The main factors that contributes to their low purification yields is their protein structure complexity and their transmembrane region. Additionally, expression of human genes within a different host organism such as *E. coli* the codon bias plays a pivotal role in protein translation levels.[25,26] The pET system (Novagen™) developed in tandem with BL21(DE3)™, allows for the strict control of protein expression by the engineered lac operon to control the binding of T7 RNA polymerase on pET vector and the host expression of exogenous T7 RNA

polymerase.[27,28] We observe the codon selection for *E. coli* BL21(DE3) and optimize the full length cDNA sequence of hPD1 and hPDL1 to fit the BL21(DE3) *E. coli* host organism.

The results of the expression levels were analyzed on an SDS-PAGE gel with coomassie stain. We observed a band at 60 kDa, for 6XHisCyPet-PD1 and 63 kDa 6XHisYPet-PDL1. The expression was analyzed for protein in the cell pellet as well. We observed a significant improvement in expression levels with codon optimization. Additionally, we observed a significant band for our fusion protein in cell lysate, indicating inclusion body formation.

Table 6: CAI Score across organisms for PD1 and PDL1

	CAI Score Human	**CAI Score *E. coli* Wildtype**	**CAI Score *E. coli* Optimized**
hPDL1	0.7	0.6	0.9
hPD1	0.8	0.6	0.9

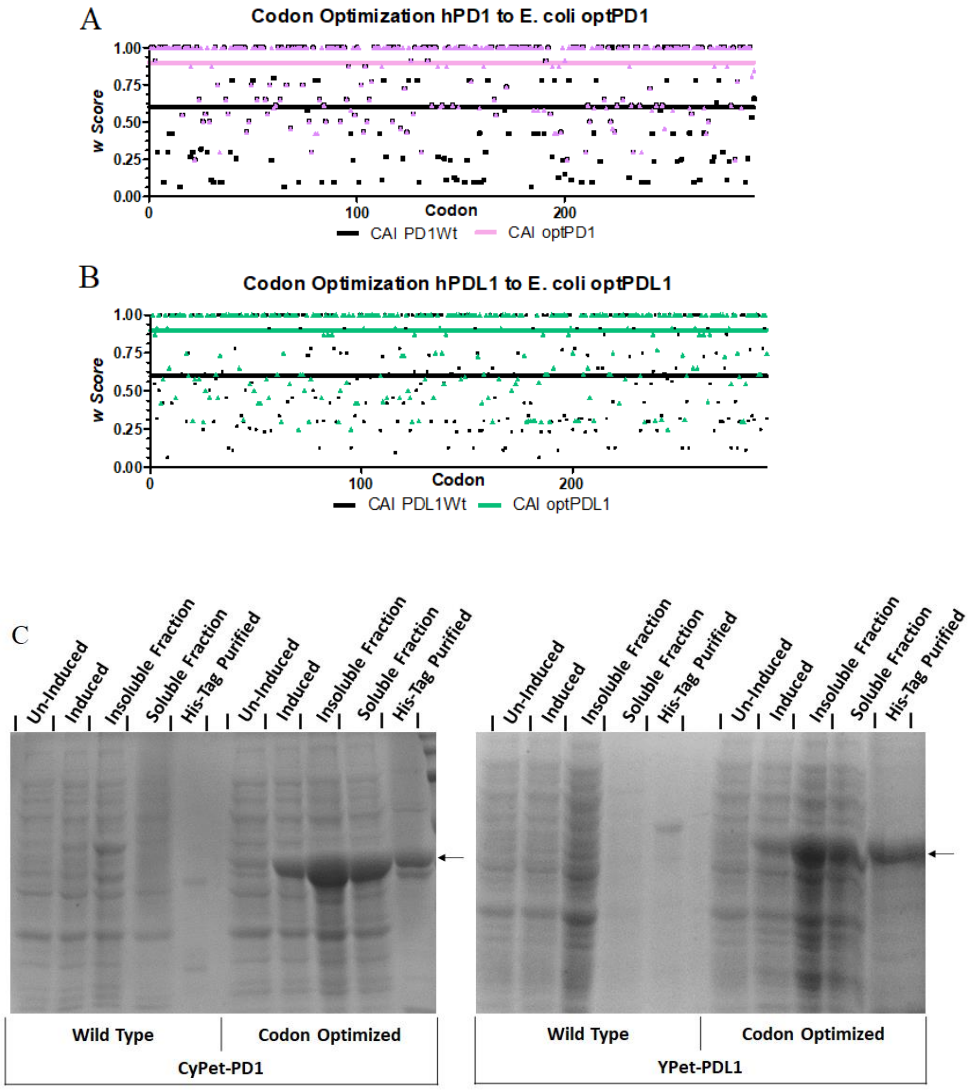


Figure 8: Overall relative codon adaptation frequency plotted for both wild type and selected codon for *E. coli*. The solid line represents the overall CAI score for wild type and the optimized sequence. The optimized PD1 sequence shifted the score to 0.9 from 0.6 in *E. coli* plot line in pink and black respectively Figure 8A. The results of optimization of PDL1 shifted the score similarly from 0.6 in E.coli to 0.9, plot in black and green respectively. (Figure 8B). The plots were generated in GraphPad Prism5^{TM}, and the scores were calculated in R.

Refolding extracellular domain CyPet-ExtPDL1 and YPet-ExtPD1

The external domains of the human PD1 (extPD1) and human PDL1 (extPDL1) were rescued from inclusion bodies. The method used was derived from a series of groups that have refolded extPD1 and extPDL1 successfully. The results show a singular

band from both CyPet-ExtPDL1 and YPet-extPD1 (Figure 9). The rescued proteins have fluorescence and were able to be quantified by fluorometer.

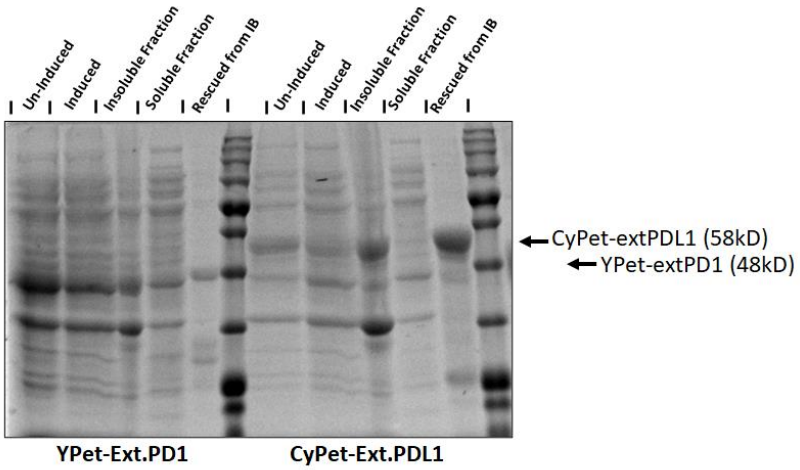


Figure 9:Commasie stain of SDS-PAGE of samples taken from various steps in the purification process. The uninduced is the cell pellet, taken from starting culture. The induced is cell pellet taken from the overnight expression. The soluble and insoluble fractions are taken after cell lysis. The rescued protein is the refolded externa domains.

Refolding Buffer Screen and Circular Dichroism Measurements

The expression results show a significant amount of protein in the inclusion body. In the past several groups have refolded the external domain of hPDL1 and hPD1.[29,30] We utilized an aggregation reporter dye from the Novagen™ refolding kit to gauge levels of aggregated proteins. Table 1 below outlines the refolding buffers used in the screen. Every component besides the denaturant in the refolding buffer stayed consistent with the buffer series 1-4. The last dialysis series was PBS and after measured for protein concentration and setup for circular dichroism (CD) measurements and aggregation dye qualification.

The aggregation reporter dye, Proteostat™ by Enzolifesciences, is applied for the initial assessment of the refolded proteins. The aggregation reporter dye binds to

aggregated amyloid structures and when excited at 550 nm emits at 600 nm. Thus, we expect to see very low signal from the aggregation in the soluble fraction and the four refolded proteins. We observed that buffers 1 and 2 had very similar aggregation response, and buffer 3 had the lowest signal among the four refolded proteins (Figure 10A).

The circular dichroism is utilized for assessment of the secondary structure of the refolded fusion protein YPet-PDL1. We observe [Θ], $deg*cm^2*dmol^{-1}$ as the response from exposing 195 nm to 250 nm range of light onto the sample. Observed a similar profile between soluble and refolded proteins Figure 10B. The protein refolded in buffer 2 has the most similar profile to the soluble fraction, with both having a profile indicative of a secondary structure with an alpha-helix. Buffer 1 shows an artifact in the measurements after 200 nm, this artifact can be caused by potential differences in the secondary structure. Buffer 3 and 4, a flat profile in the measurements is indicative of denatured protein in the sample. The final characterization of the refolded proteins is to perform the qFRET Kd assay to determine integrity of protein structure.

Table 7: Buffer conditions used to refold YPet-OptPDL1 from inclusion body.

	Buffer Components
Buffer 1	0.5 M Guanidine HCl, 50 mM Tris pH 8, 300 mM NaCl, 500 mM Arginine, 0.8 mM KCl, 0.5/0.25 mM Glutathione R/O
Buffer 2	0.5 M Guanidine HCl, 50 mM Tris pH 8, 300 mM NaCl, 250 mM Arginine, 0.8 mM KCl, 0.5/0.25 mM Glutathione R/O
Buffer 3	1.0 M Guanidine HCl, 50 mM Tris pH 8, 300 mM NaCl, 500 mM Arginine, 0.8 mM KCl, 0.5/0.25 mM Glutathione R/O
Buffer 4	1.0 M Guanidine HCl, 50 mM Tris pH 8, 300 mM NaCl, 250 mM Arginine, 0.8 mM KCl, 0.5/0.25 mM Glutathione R/O
Control Sol F.	Soluble Fraction Purified Protein in PBS

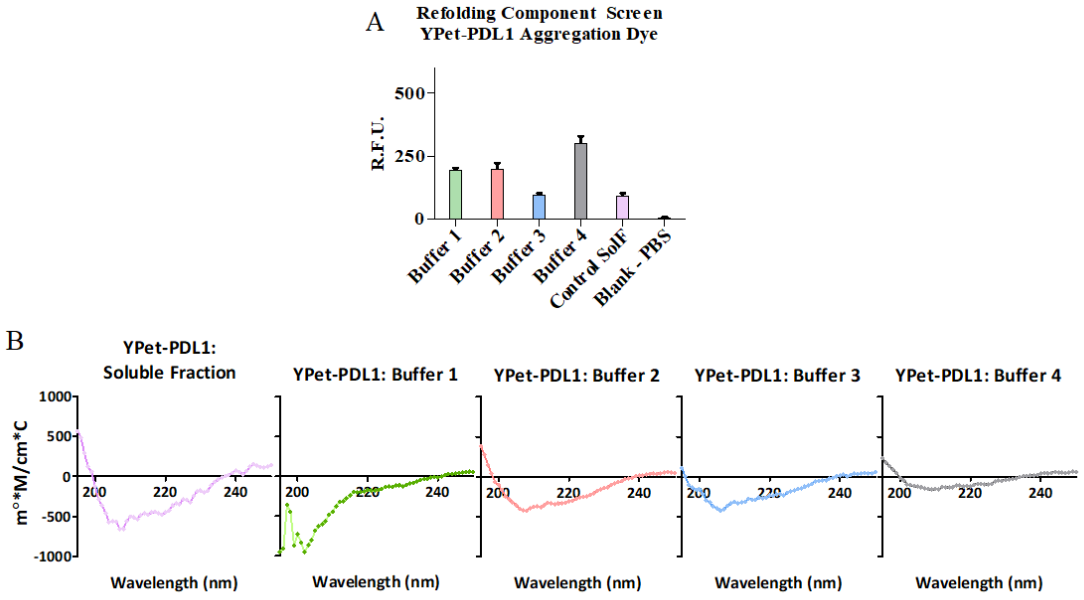
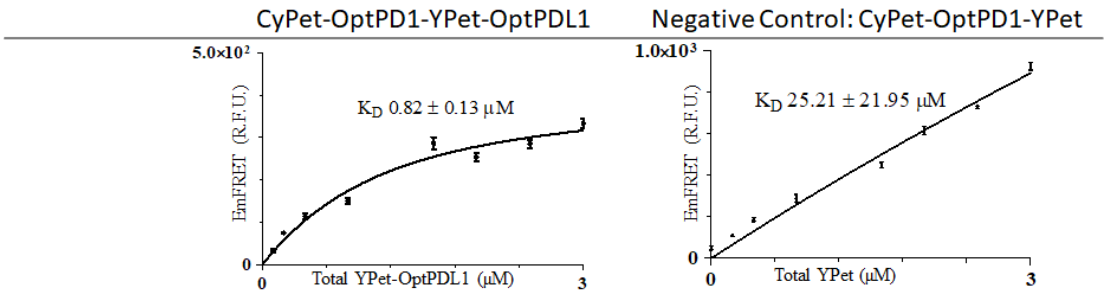


Figure 10:The aggregation dye ProteoStat by Enzolifesciences is used to report protein aggregation, it is added to equivalent concentration of protein across each buffer at a 1:50 volumetric ratio, of dye to protein solution, using the soluble fraction purified YPet-OptPDL1 as control. Excitation at 550 nm and emission at 600 nm A. The CD plots are of the refolded proteins with soluble protein as control. The plots were made on GraphPad Prism5™.

qFRET based dissociation constant K_D determination

The results of the qFRET based K_D is shown in figure 11, we observe the fitted curve and a saturation of Em_{FRET} signal, and a saturation in binding is reached within 3 µM of ligand introduction. Negative control is the interaction assay with only the fluorescent FRET pairs and without the interacting molecule, OptPDL1, fused at the C-

Terminal. The resolved K_D of CyPet-OptPD1 and YPet-OptPDL1 from the fit of the qFRET assay is 0.81 µM and a standard error of 0.13 µM. The non-linear regression fit results also provide a 95% confidence interval of 0.55 to 1.07 µM. The R^2 value of 0.97 goodness of fit provides additional confidence in the determined constant. The negative control had a large K_D value of 25.21 µM with standard error of 21.95 µM and 95% confidence interval of 0 to 70.73, these results do not gain any confidence in the resolved K_D of negative control YPet protein alone.



K_D (µM)	**0.82**	**25.21**
Std. Error (µM)	0.13	21.95
95% Confidence Intervals (µM)	0.55 to 1.07	0.0 to 70.73
R^2	0.97	0.98

Figure 11: K_D measurement was done with CyPet-OptPD1 held constant at 0.5 µM and titration of the acceptor, YPet-OptPDL1 and YPet only, from 0 – 3.0 µM. The extracted Em_{FRET} was plotted against the titration of total YPet-OptPDL1. Equation 16 was used to fit the data from both assays. GraphPad Prism 5™ is used for resolving K_D showing the resulting non-linear regression fit line in the plot.

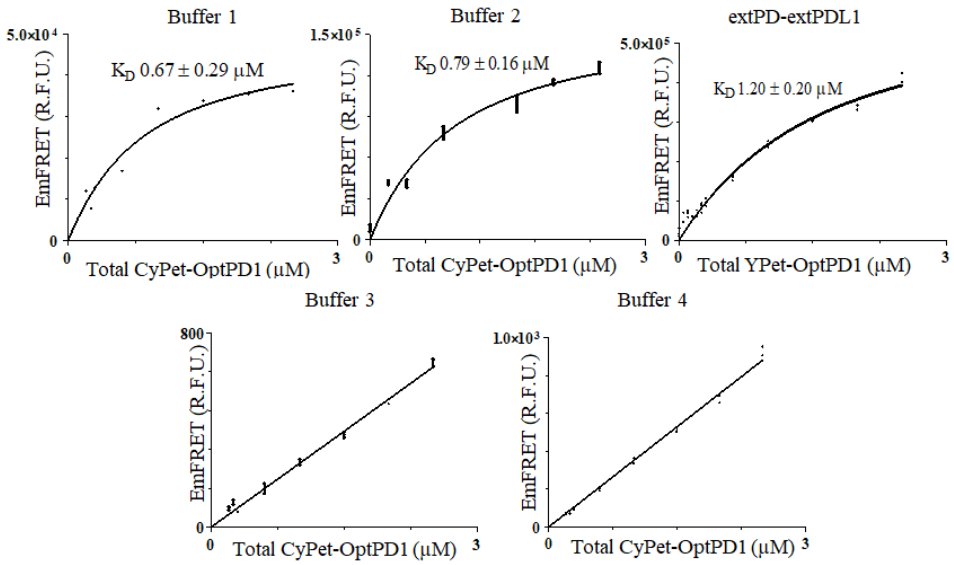
Refolding human full length Program Death Ligand 1

We screened refolding conditions for full length optPDL1 using qFRET K_D as a confirmation for folded proteins. The buffers outlined in table 1, lists a series of refolding conditions, based on refolding studies from past groups attempts to refold external

domain of hPDL1 and hPD1.[29–31]. The buffers used all had Guanidine-HCl as the denaturant, with varying amount of aggregation inhibitor and Glutathione Reduced/Oxidized to sustain the redox potential. We observed a significant difference in qFRET K_D across buffers with no qFRET K_D resolved in buffers 3 and 4 (Figure 12). The refolded proteins with buffer 1 and 2 showed a similar secondary structure profile in circular dichroism measurements when compared to the YPet-OptPDL1 purified from the soluble fraction. The results from the refolded YPet-OptPDL1 from buffer 2 estimated a K_D of 0.79 µM, standard error of 0.16 µM, with a 95% confidence interval of 0.46 to 1.12 µM (Figure 12). The soluble fraction and the refolded PDL1 both had comparable dissociation constants and provide confidence in the approach and buffer used.

Table 8:Refolding buffers screen for rescuing YPet-OptPDL1 from inclusion body

Buffer 1	0.5 M Guanidine HCl, 50 mM Tris pH 8, 300 mM NaCl, 500 mM Arginine, 0.8 mM KCl, 0.5/0.25 mM Glutathione O/R
Buffer 2	0.5 M Guanidine HCl, 50 mM Tris pH 8, 300 mM NaCl, 250 mM Arginine, 0.8 mM KCl, 0.5/0.25 mM Glutathione O/R
Buffer 3	1.0 M Guanidine HCl, 50 mM Tris pH 8, 300 mM NaCl, 500 mM Arginine, 0.8 mM KCl, 0.5/0.25 mM Glutathione O/R
Buffer 4	1.0 M Guanidine HCl, 50 mM Tris pH 8, 300 mM NaCl, 250 mM Arginine, 0.8 mM KCl, 0.5/0.25 mM Glutathione O/R



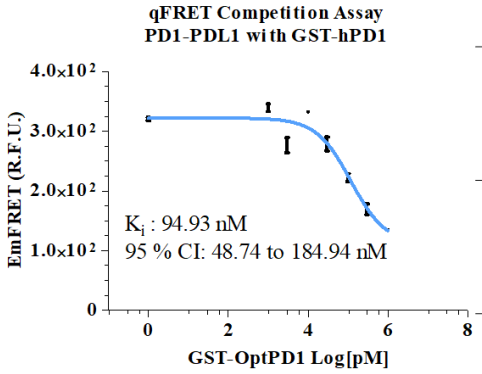
	Buffer 1	Buffer 2	Buffer 3	Buffer 4	YPet-ExtPD1 and CyPet-ExtPDL1
Kd μM	0.67	0.79	-	-	1.20
Std. Error μM	0.29	0.16	-	-	0.20
95% Confidence Intervals μM	0.06 to 1.28	0.46 to 1.12	-	-	0.77 to 1.63
R^2	0.80	0.96	-	-	0.98

Figure 12: The qFRET Refolded was setup as the constant and the titration of CyPet-hPD1 from (0 – 2.5). The extraction of Em_{FRET} from the assay is plotted against the total concentration of CyPet-hPD1. Prism™ is used for the non-linear regression solution, and all measurements were taken on Molecular Devices Flexstation II 384 using Software Max Pro 7.0™.

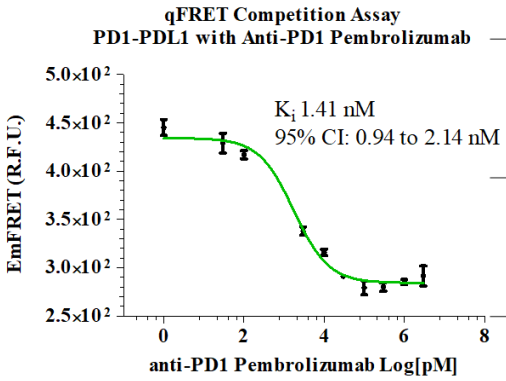
qFRET based competition assay

The two molecules characterized for inhibition of CyPet-OptPD1-YPet-OptPDL1 interaction are GST-OptPD1 without a fused fluorescent protein, and the anti-PD1 mAbs Pembrolizumab (MK-3475). The observed inhibition profiles (Figure 13) are consistent to competition assay response to an inhibitor. The fit for equations 17 and 19 were implemented directly to the Em_{FRET} data with titration of inhibitors to resolve the K_i and

IC_{50}. The determined K_i for the GST-OptPD1 from this competition assay is 94.93 nM (Figure 13A), 95% confidence interval for K_i is as low as 48.74 nM and as high as 184.94 nM with R^2 value of 0.91. We also fit equation 17 alone to estimate IC_{50} value for this assay, and found IC_{50} to be 118.38 nM, 95% confidence interval to be 60.77 nM and as high as 230.60 nM with R^2 value of 0.91. The observed K_i for pembrolizumab is 1.41 nM (Figure 13B) and the fit results show a 95 % confidence interval of 0.94 to 2.14 nM with an R^2 value of 0.97. The IC_{50} fit observed was 1.76 nM with 95% confidence interval of 1.17 to 2.66 nM.



qFRET Competition Assay: GST-OptPD1	
qFRET K_i nM	94.93
95 % Confidence Interval nM	48.74 to 184.94
R^2	0.91
qFRET IC_{50} nM	118.38
95 % Confidence Interval nM	60.77 to 230.60
R^2	0.91



qFRET Competition Assay: Pembrolizumab	
qFRET K_i nM	1.41
95 % Confidence Interval nM	0.94 to 2.14
R^2	0.97
qFRET IC_{50} nM	1.76
95 % Confidence Interval nM	1.17 to 2.66
R^2	0.97

Figure 13:qFRET Competition assay was setup with constant concentration of the substrate and titration of inhibitor. The fit results are tabulated next to the fit plot for A.) the response to the addition of GST-OptPD1 fitted to equation 17 for IC_{50} and both equation 17 and 19 for K_i estimation (Blue Plot). B.) The response to the addition of Pembrolizumab is fitted to equation 17 for IC_{50} and both equation 17 and 19 for K_i estimation (Light Green Plot). The measurements were taken on Molecular Devices SpectraMax 3™ and GraphPad Prism™ is used for non-linear regression fit for IC_{50} and K_i

2.6 Discussion

The qFRET assay is readily applied to characterize the cell surface proteins PD1 and PDL1. The equilibrium constant K_D for PD1-PDL1 was estimated to be 0.81 µM with standard error of 0.13 µM. The reported K_D of external domain PD1 and external domain PDL1 using SPR range from 8.2 µM by Cheng *et al.* (2012), 0.77 µM by Butte *et*

al. (2013) and more recently, 4.1 µM Tang *et al.* (2019).[7,32,33] The broad range of reported dissociation constants could result from the differences in protein expression, and preparation of protein for the technique utilized. The application of SPR requires immobilization and modification of one of the binding proteins. The qFRET interaction assay reported here does not require any post purification modifications, and both proteins are in solution. The refolding study demonstrated the utility of qFRET for the rapid assessment of refolded proteins. The codon optimization of both donor and acceptor fusion proteins provided a significant increase in protein yields. The resulting expression and purification of both CyPet-OptPD1 and YPet-OptPDL1 demonstrated that both proteins will form inclusion bodies in E.coli.

The observed K_D from buffer 2 condition was the most comparable to the soluble fraction. The external domains of PD1 and PDL1 after refolding were evaluated for their K_D and we observed a drop in affinity to 1.2 µM, with standard error 0.2 µM and a 95 % confidence interval of 0.77 to 1.63 µM. The slight decrease in affinity can be an artifact of the refolding process, however the full-length refolded protein demonstrated a similar affinity as the soluble fraction. The external domain of the two protein have been evaluated for their affinity numerous times and they've reported a much higher K_D, (8.2 µM, 4.1 µM, and 2.2 µM) by the groups mentioned earlier. The refolding process that includes a rapid qFRET based assay, provides once again the utility of the qFRET platform.

The qFRET competition assay evaluated the inhibition of PD1-PDL1 binding with two inhibitor molecules. The GST-OptPD1 inhibition response was evaluated with

an estimated K_i of 94.93 nM, with a confidence interval of 48.74 to 184.94 nM. The affinity of PD1 to PDL1 has been established by many groups and adding soluble PD1 to act as an inhibitor provides an evidence of the assays functionality when targeting the acceptor fusion protein. The other inhibitor is an engineered antibody, pembrolizumab, against PD1 and has been reported to have very high affinity to PD1 in in-vitro experiments. The reported SPR evaluation of this antibody to PD1 has shown to have a K_D of 2.5 nM.[34] The observed K_i of pembrolizumab based on the qFRET competition assay is 1.41 nM with a 95% confidence interval of 0.94 to 2.14 nM. SPR measured K_D of pembrolizumab to external domain of PD1 is 3.9 nM ± 0.5 nM.[34] The measured value here using the qFRET platform is within a very close range of the previously reported affinity. The qFRET platform utility is showcased here with 1.) The sample to kinetic measurement workflow, without modifying expressed proteins or attachment of proteins to a surface. 2.) One simplified platform for kinetic parameter determination, dissociation constant K_D of two proteins as well as determine K_i of a given inhibitor based on qFRET signal. Furthermore, the observed robust response of the Em_{FRET} signal between PD1 and PDL1 provide confidence of this assays ability to detect the interaction of other surface proteins. The accessible approach to the evaluation of this category of protein-protein interaction can accelerate the discovery and characterization of inhibitors.

2.7 Materials and Methods

Crosstalk parameters, alpha α, and beta β

The crosstalk parameters, alpha, and beta are part of the qFRET extraction equation resolving Em_{FRET}. The fluorescent proteins CyPet, donor, and YPet, acceptor,

are a FRET optimized pair developed to maximize FRET efficiency.[18] The donor protein CyPet, at 500 µM, is excited at 414 nm and has peak emission is at 475 nm. The acceptor protein YPet, at 200 µM, is excited at 475 nm with peak emission at 530 nm. The alpha (α) parameter is determined for the fluorometer by exciting the donor fusion protein CyPet-PD1 at 414 nm and then a ratio of emission at FRET wavelength of 530 nm divided by the emission at 475 nm. The beta (β) parameter for the acceptor protein, YPet, is derived by dividing the emission at 530 nm when excited at 414 nm by the emission at 530 nm when excited at 475 nm

The fluorescent plate reader utilized in this study is made by Molecular Devices and is called SpectraMax M3™ packaged with SoftMax Pro 7™ software suite. Plotted in Prism™. The measurements are taken in PBS and in volumes of 60 µL at room temperature (22°C), using Greiner 384 well plates, black with clear bottom. The plate reader by Molecular Devices, Flexstation II 384(™), the alpha and beta values, for CyPet and YPet were determined to be 0.360 (α) and beta of 0.026 (β) in a previous publication by Song *et al.*.[23]

Construction of qFRET fusion proteins, CyPet-PD1 and YPet-PDL1

The full-length PD1 and PDL1 are membrane proteins that fall under the category of difficult-to-purify proteins. The codon optimized full length PD1 and PDL1 and fusion cDNA constructs were constructed by the Huaxi Research Hospital. The fully constructed full length codon optimized PD1 and PDL1 were cloned into pET28 (Novagen™) vectors. Each gene is fused to the qFRET pair CyPet and YPet on the N-terminal of the fusion proteins with a linker more details of the linker can be found in the following

study (Kaur H. *et al.*).[35] The restriction sites NHEI and SALI flank the fluorescent proteins followed by a linker region to allow more freedom of movement to the interacting molecules. The genes of interest codon optimized cDNA are on the C-terminal of the linker region with the restriction sites SALI and NOTI flanking the gene coding sequence, with the stop codon at the C terminal. This results in a fusion protein with a 6xHistadine tag followed by the fluorescent protein, linker region, and the gene of interest.

The expression of full-length fusion protein CyPet-OptPD1 and YPet-OptPDL1 is demonstrated here by applying codon optimization on the native cDNA for optimization in the host expression organism Escherichia coli (*E. coli*) BL21(DE3). Amino acid translation has synonymous codons for each amino acid, and the frequency of a singular codon out of the synonymous codons for an amino acid results in a codon bias. Codon bias has downstream implication to amino acid translation, such as the availability of tRNA will match the codon bias for quick translation. In this method we apply a codon optimization technique that accounts for tRNA frequency within *E. coli* B Strain. A metric used to evaluate codons selection bias is the relative codon adaptation frequency (w) and the codon adaptation index is the average of w score across the length of the amino acid chain (CAI).[36–38] The adaptation frequency provides a singular metric for the frequency of any codon and compares that to the frequency of the most adapted codon for an amino acid within the host organism. The w plot highlights the difference between the codon optimized and the human native cDNA sequence for the *E. coli* B Strain organism. The codon optimization was constructed in R with BioSeqiner and StringR Package. The

codon selection was based on the studies that focused on measured tRNA levels.[36,37,39] The plots were completed on Prism™.

Expression and Purification of CyPet-OptPD1 and YPet-OptPDL1

The bacteria strain used to express and purify full length PD1 and PDL1, fusion proteins is The SHuffle® T7(Neb) bacteria strain (Gifted by Perry Lab). The proteins hPD1 and hPDL1 both have complex secondary structures that require disulfide bond formation, cysteine 54 – 123 on PD1, cysteine 40 – 114 and cysteine 155 – 209 on PDL1. The SHuffle® T7(Neb) strain provides the functionality of a disulfide isomerase to promote the formation of disulfide bonds. The cells are inoculated at 1:100 starting culture to expression medium, Terrific Broth™ (RPI), and once reached 0.4 OD were induced with 0.350 mM IPTG (isopropyl β-D-1-thiogalactopyranodie) and expressed at 22 °C overnight. Cell were lysed using sonication in binding buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM Imidazole, 1 mM phenylmethylsulfonyl fluoride (PMSF)) and centrifuged after sonication at 35,000 X g for 30 minutes at 4 °C. Both proteins are found in the inclusion body and in the soluble fraction during purification. All the purification steps are completed at 4 °C. The soluble fraction of the proteins was bound to NiNTA beads and washed 20x bead volume with a.) 20 mM Tris-HCl pH 7.5, 500 mM NaCl b.) 20 mM Tris-HCl pH 7.5, 1.5 mM NaCl, and 0.1 % Triton X-100 c.) 20 mM Tris-HCl pH 7.5, 500 mM NaCl, and 10 mM Imidazole. After washes are complete the proteins were eluted from the NiNTA column by elution buffer (20 mM Tris-HCl, 50 mM NaCl, and 350 mM Imidazole). The eluted proteins were dialyzed in PBS overnight at 4 °C. The

proteins were stored in 10% glycerol at -80 °C. The protein freeze thaw cycles were limited.

Similarly for GST-hPD1 with GST bead binding/lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, and 0.1 mM EDTA) and three 20x bead volume GST bead washes with the binding buffer. Eluted with GST elution buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1 mM EDTA, and 20 mM Glutathione reduced) and dialyzed in PBS over night at 4 °C. The proteins were stored in 10% glycerol at -80 °C.

Refolding extracellular domain CyPet-ExtPDL1 and YPet-ExtPD1

The extracellular domain of PD1 and PDL1 protein have been purified from inclusion bodies and reported previously. In this study the cDNA of the external domain for PD1 and PDL1 were fused to the FRET pair YPet and CyPet respectfully and rescued from inclusion bodies. The fusion genes were expressed in BL21(DE3), the 2xYT (RPMI) media was inoculated at 1:100 ratio, and at OD 600 expression was induced with 0.4 mM IPTG. The expression temperature was at 22°C overnight. The cells were pelleted by centrifugation at 8000 X g for 5 mins at 4°C. The cell pellet was resuspended in lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM Imidazole, 1 mM phenylmethylsulfonyl fluoride (PMSF)) and sonicated, after sonication the lysate was centrifuged at 35,000 X g for 30 minutes at 4 °C. The two fusion proteins are found in the inclusion body, coomassie stain SDS-PAGE gel show most of the proteins are in the precipitate. The cell precipitate after lysis and centrifugation is resuspended and sonicated in 50 mM Tris-HCl pH 8.0, 500 mM NaCl, and 0.1% Triton X-100, two times and the final third wash in lysis buffer. The washed pellet is finally resuspended in denaturing

buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 6 M Guanidine) left agitating over night at 4 °C. The dissolved pellets were centrifuged at 35,000 X g for 30 minutes at 4 °C. The individual pellets were refolded by dialysis over a guanidine step down (6,3,2,1,0.5 M), 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 250 mM Arginine. For hPD1 and hPDL1 the refolding buffer series had Glutathione Reduced/Oxidized at 0.5/0.25 mM and 1/0.5 mM hPDL1. Each dialysis was over night at 4 °C, with centrifugation at 35,000 X g for 15 minutes at 4 °C at each exchange. The final dialysis was in PBS for 36 hours at 4 °C, with fresh PBS every 24 hours. Results of the purified proteins were ran on SDS-PAGE gel, with protein samples taken at various steps in the purification process.

Refolding of YPet Fused Full length OptPDL1

Inclusion bodies of YPet-hPDL1 were collected and sonicated, at 65 Hz with 10 second pulse on and off, for 5 mins, in pellet wash buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, and 0.1% Triton X-100). After sonication, the inclusion body was centrifuged at 35,000 X g, and the process was repeated twice. Following triton wash, inclusion body was suspended in wash buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl) and, sonicated at 65 Hz with 10 second pulse on and off, for 5 mins and centrifuged each time. The washed inclusion body pellet was resuspended in denaturing buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 6 M Guanidine-HCl), and allowed to fully denature over night at 4 °C with constant agitation. The denatured fractions were setup for dialysis in four buffers, listed in table 1, with similar base (50 mM Tris-HCl pH 8.0, 500 mM NaCl, Glutathione R/O 0.5/0.25 mM) with varying concentration of denaturant (Guanidine-HCL), aggregation inhibitor (Arginine), and Glutathione oxidized/reduced at 4 °C for no less than 12 hours.

After each overnight dialysis the fraction was centrifuged at 35,000 x g for 30 minutes at 4 °C. Each dialysis condition stepped down the denaturant from 6 M to the refolding buffers. The last refolding dialysis buffer consisted of the four buffers outlined in table 1 for 12 hours at 4 °C. Finally, the protein was dialyzed in PBS for 72 hours with fresh PBS every 24 hours at 4 °C. At every buffer exchange the fraction was centrifuged. The refolded protein was stored in PBS with 10 % glycerol and limited the freeze thaw cycles.

We utilized an aggregation reporter dye from the Novagen^{TM} refolding kit to gauge levels of aggregated proteins. Table 1 below outlines the refolding buffers used in the screen. The buffering system is 50mM Tris at pH 8.0 throughout the process till the final PBS dialysis at pH 7.4. The last dialysis series was PBS and after measured for protein concentration and setup for circular dichroism (CD) measurements and aggregation dye qualification.

Circular Dichroism Setup and Measurements

The circular dichroism is utilized for assessment of the secondary structure of the refolded fusion protein YPet-PDL1. The technique applies two circular polarized Electric € waves, right-handed (E_R) and lefthanded (E_L), onto a sample, and the sample will absorb the two opposite circular polarized waves differently depending on the secondary structure of the molecule. After interacting with the sample, the two E waves are added together to produce a now elliptical E wave which is reported as degrees ellipticity [Θ]. We observe [Θ], deg*cm^2*dmol^{-1} as the response from exposing 195 nm to 250 nm range of light onto the sample. For the samples to be measured down to 195 nm, fractions of the protein samples were dialyzed in 10 mM potassium phosphate buffer (pH 7.4) with 100

mM NaCl overnight. Measurements taken on Jasco J-810™ circular dichroism/spectropolarimeter with a 0.1 mm cuvette. MilliQ water was used to rinse the cuvette between measurements. The samples were measured at 5 µg/ml, as some sample yields were low due to precipitation of protein.

qFRET K_D Assay

The Molecular Device instrument SpectraMax 3™ and Software Max Pro 7 software suit set for "Endpoint" reads, with fluorescence measurements and the PMT gain set to a constant at "Low" setting all at 22°C. The Em_{FRET} vs YPet-OptPDL1$_{Total}$ ($[L]_{Total}$) series of reaction are setup in-vitro in PBS with the $[R]_{Total}$ at 0.50 µM and titrating concentration of ligand from 0 – 3.0 µM, post mixing the reactions are incubated at 37°C for 30 minutes. The post processing of the EmTotal signal to resolve Em_{FRET} requires the measurement of three signals. First is the EmTotal, emission at 530 nm when excited at 414 nm, the raw signal with the embedded donor and acceptor crosstalk contribution. The donor crosstalk contribution is resolved by measuring the CyPet-OptPD1 emission at 475 nm when excited at 414 nm, $FL_{D,}$ and multiplying Fl_D with alpha (α). The acceptor contribution, YPet-OptPDL1 emission at 530 nm when excited at 475 nm, FL_A, and multiplying by beta (β). The post processing of the EmTotal signal resolves the qFRET data, Em_{FRET} vs YPet-OptPDL1$_{Total}$ ($[L]_{Total}$).

GraphPad Prism 5™ is used to resolve the K_D and $Em_{FRETMax}$ values by fitting equation 16 onto the data set Em_{FRET} vs YPet-OptPDL1$_{Total}$ ($[L]_{Total}$). The regression solution constraints are set as following, $[R]_{Total}$ concentration is kept as a constant at 0.5 µM, K_D and $Em_{FRETMax}$ must be greater than 0. The initial value for K_D is set to 0.1 µM

and $Em_{FRETMax}$ is set to 1. The fit results include the value for $Em_{FRETMax}$ and K_D with a standard error, 95% confidence interval, and a R^2 value for the goodness of fit.

qFRET Competition Assay

The competition assay is setup in a similar manner with CyPet-PD1 at 0.2 µM and YPet-PDL1 at 1 µM and is kept constant with varying concentration of inhibitor. The range of concentration of inhibitor varies with each inhibitor molecule. For GST-PD1 we varied the concentration from 0 – 1 µM. For pembrolizumab (MCE- HY-P9902) we used a range from 0 – 3 µM, and PBS was used to dilute the reactions.

The Non-linear regression curve fitting tool was used, to resolve K_i directly. Utilizing equation 17 and 19, the solver will resolve $LogIC_{50}$, and uses that solution for the adapted, Cheng-Prsuoff relationship, equation 19. The basis for equation 19 is derived from the K_i and IC_{50} relationship, we solve equation 18 for IC_{50}, and keep IC_{50} and K_i in log form. Using the solver with both equations together allows us to solve for K_i directly from the Em_{FRET} data. GraphPad Prism 5^{TM} custom equation allows for both equations, Equation 17 and 19, to be used within the same solution. The constants [R] and [K_D] are specified in the same units as the inhibitor, and [R] and [K_D] are kept as constants in the constraints. The $Em_{FRETMax}$ and $Em_{FRETMin}$ are initially set to Y_{max} and Y_{min}. No constraint is placed on $Em_{FRETMax}$, $Em_{FRETMin}$ and $LogK_i$ during the solve. The fit will result in a $LogK_i$ value, which is setup to be reported by the antilog of the solved $logK_i$. The IC_{50} is also resolved similarly by fitting the Em_{FRET} response to inhibitor to equation 17. With the only initial condition and constraint of $Em_{FRETMax}$ set to Y_{max} and $Em_{FRETMin}$ set to Y_{Min}.

2.8 References

1. De Sousa Linhares, A. *et al.*. Therapeutic PD-L1 antibodies are more effective than PD-1 antibodies in blocking PD-1/PD-L1 signaling. Sci Rep **9**, 11472 (2019).

2. Rotte, A., Jin, J. Y. & Lemaire, V. Mechanistic overview of immune checkpoints to support the rational design of their combinations in cancer immunotherapy. Annals of Oncology **29**, 71–83 (2018).

3. Immune checkpoint blockade therapy for cancer_ An overview of FDA-approved immune checkpoint inhibitors | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S1567576918302522?token=E470D6DCCD550D2C340680DA8010344EF7732B4AE78A7AB4193CB4B851C228CE1582A54848EEECE16681F51F5F5F64D3&originRegion=us-east-1&originCreation=20210715211119 doi:10.1016/j.intimp.2018.06.001.

4. The PD-1/PD-Ls pathway and autoimmune diseases | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S0008874914000847?token=146CFB3C490C0B3A4B789EE3E1AAF219B8BA23BA999B37A2D6E5262AC1E9475D90823D09FCCBCD8DF7AD33C3BFBE0483&originRegion=us-east-1&originCreation=20210712050330 doi:10.1016/j.cellimm.2014.05.006.

5. Zak, K. M. *et al.*. Structure of the Complex of Human Programmed Death 1, PD-1, and Its Ligand PD-L1. Structure **23**, 2341–2348 (2015).

6. Patel, S. P. & Kurzrock, R. PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy. Mol Cancer Ther **14**, 847–856 (2015).

7. Butte, M. J., Keir, M. E., Phamduy, T. B., Sharpe, A. H. & Freeman, G. J. Programmed Death-1 Ligand 1 Interacts Specifically with the B7-1 Costimulatory Molecule to Inhibit T Cell Responses. Immunity **27**, 111–122 (2007).

8. Spain, L., Diem, S. & Larkin, J. Management of toxicities of immune checkpoint inhibitors. Cancer Treatment Reviews **44**, 51–60 (2016).

9. Tang, J. *et al.*. The clinical trial landscape for PD1/PDL1 immune checkpoint inhibitors. Nature Reviews Drug Discovery **17**, 854–855 (2018).

10. Upadhaya, S. *et al.*. Combinations take centre stage in PD1/PDL1 inhibitor clinical trials. Nature Reviews Drug Discovery **20**, 168–169 (2020).

11. Park, J. A. & Cheung, N.-K. V. Limitations and opportunities for immune checkpoint inhibitors in pediatric malignancies. Cancer Treat Rev **58**, 22–33 (2017).

12. Vachali, P. P., Li, B., Bartschi, A. & Bernstein, P. S. Surface plasmon resonance (SPR)-based biosensor technology for the quantitative characterization of protein–carotenoid interactions. Archives of Biochemistry and Biophysics **572**, 66–72 (2015).

13. Wang, Y., Wang, G., Moitessier, N. & Mittermaier, A. K. Enzyme Kinetics by Isothermal Titration Calorimetry: Allostery, Inhibition, and Dynamics. Front. Mol. Biosci. **0**, (2020).

14. Berggård, T., Linse, S. & James, P. Methods for the detection and analysis of protein–protein interactions. PROTEOMICS **7**, 2833–2842 (2007).

15. Sun, Y., Wallrabe, H., Seo, S.-A. & Periasamy, A. FRET microscopy in 2010: The legacy of Theodor Förster on the 100th anniversary of his birth. Chemphyschem **12**, 462–474 (2011).

16. Algar, W. R., Hildebrandt, N., Vogel, S. S. & Medintz, I. L. FRET as a biomolecular research tool — understanding its potential while avoiding pitfalls. Nat Methods **16**, 815–829 (2019).

17. Sun, Y., Wallrabe, H., Seo, S.-A. & Periasamy, A. FRET microscopy in 2010: The legacy of Theodor Förster on the 100th anniversary of his birth. Chemphyschem **12**, 462–474 (2011).

18. Evolutionary optimization of fluorescent proteins for intracellular FRET | Nature Biotechnology. https://www.nature.com/articles/nbt1066?platform=hootsuite.

19. Optimizing fluorescent protein trios for 3-Way FRET imaging of protein interactions in living cells | Scientific Reports. https://www.nature.com/articles/srep10270#Tab2.

20. GraphPad Prism 7 Curve Fitting Guide - PRINCIPLES OF REGRESSION. https://www.graphpad.com/guides/prism/7/curve-fitting/index.htm.

21. Yung-Chi, C. & Prusoff, W. H. Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. Biochemical Pharmacology **22**, 3099–3108 (1973).

22. Krohn, K. A. & Link, J. M. Interpreting enzyme and receptor kinetics: keeping it simple, but not too simple☆☆Contribution to the Receptors Meeting, San Diego, February 2003. Nuclear Medicine and Biology **30**, 819–826 (2003).

23. Song, Y., Madahar, V. & Liao, J. Development of FRET assay into quantitative and high-throughput screening technology platforms for protein-protein interactions. Ann Biomed Eng **39**, 1224–1234 (2011).

24. Lin, S.-H. & Guidotti, G. Purification of membrane proteins. Methods Enzymol **463**, 619–629 (2009).

25. Angov, E., Hillier, C. J., Kincaid, R. L. & Lyon, J. A. Heterologous Protein Expression Is Enhanced by Harmonizing the Codon Usage Frequencies of the Target Gene with those of the Expression Host. PLoS One **3**, (2008).

26. Karlin, S., Mrázek, J. & Campbell, A. M. Codon usages in different gene classes of the Escherichia coli genome. Mol. Microbiol. **29**, 1341–1355 (1998).

27. Shilling, P. J. *et al.*. Improved designs for pET expression plasmids increase protein production yield in Escherichia coli. Commun Biol **3**, 1–8 (2020).

28. Rosenberg, A. H. *et al.*. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene **56**, 125–135 (1987).

29. Li, D.-W. *et al.*. Refolding and Characterization of Recombinant Human GST-PD-1 Fusion Protein Expressed in Escherichia coli. Acta Biochimica et Biophysica Sinica **36**, 141–146 (2004).

30. Wang, Y. *et al.*. A Systematic Protein Refolding Screen Method using the DGR Approach Reveals that Time and Secondary TSA are Essential Variables. Scientific Reports **7**, 9355 (2017).

31. Lin, D. Y. -w. *et al.*. The PD-1/PD-L1 complex resembles the antigen-binding Fv domains of antibodies and T cell receptors. Proceedings of the National Academy of Sciences **105**, 3011–3016 (2008).

32. Cheng, X. *et al.*. Structure and Interactions of the Human Programmed Cell Death 1 Receptor. Journal of Biological Chemistry **288**, 11771–11785 (2013).

33. Tang, S. & Kim, P. S. A high-affinity human PD-1/PD-L2 complex informs avenues for small-molecule immune checkpoint drug discovery. 786319 https://www.biorxiv.org/content/10.1101/786319v1 (2019) doi:10.1101/786319.

34. Brown, M. E. *et al.*. Assessing the binding properties of the anti-PD-1 antibody landscape using label-free biosensors. PLOS ONE **15**, e0229206 (2020).

35. Malik-Chaudhry, H. K., Saavedra, A. & Liao, J. A linker strategy for trans-FRET assay to determine activation intermediate of NEDDylation cascade. Biotechnology and Bioengineering **111**, 1288–1295 (2014).

36. Lee, S., Weon, S., Lee, S. & Kang, C. Relative Codon Adaptation Index, a Sensitive Measure of Codon Usage Bias. Evol Bioinform Online **6**, 47–55 (2010).

37. Wei, Y., Silke, J. R. & Xia, X. An improved estimation of tRNA expression to better elucidate the coevolution between tRNA abundance and codon usage in bacteria. Sci Rep **9**, 3184 (2019).

38. Karlin, S., Mrázek, J. & Campbell, A. M. Codon usages in different gene classes of the Escherichia coli genome. Molecular Microbiology **29**, 1341–1355 (1998).

39. Sharp, P. M. & Li, W. H. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res **15**, 1281–1295 (1987).

Chapter 3: In-Vitro qFRET Assay for SUMOylation of IAV-Matrix 1 Protein for MS Analysis of Modified Lysines

3.1 Influenza A Virus Life Cycle

The influenza A viral particle contains anti-sense viral RNA (vRNA) packaged as a viral Ribonucleoprotein(vRNP) complex.[1] The vRNP complex packages the eight viral RNA strands wrapped around numerous nucleoprotein (NP) and individual viral polymerase (P) subunits PB1, PB2, and PA. The vRNP complex has been observed to be packaged as a 1+7 configuration that demonstrates the complexity of the viral RNA genome super structure.[2] The vRNP complex NP protein interact with the M1 protein which acts as structural support and is bound on the inside of the viral particle envelope. The host derived lipid membrane is the envelope that contains the vRNP complex. The envelope is lined with M1 protein and extending through are the membrane proteins Hemagglutinin (HA), Neuraminidase (NA), and Matrix 2 (M2) proteins. The initiation of viral entry occurs when membrane protein HA binds onto the host cell surface glycoconjugate's terminal SA residue. The NA protein functions as a sialidase to improve affinity and improve HA binding.

Endocytosis occurs and the vRNP complex are released into the cell cytosol. Within the cell cytosol the vRNP is imported into the nucleus using the importin-α-importin-β nuclear import pathway. The NP protein hold nuclear localization sequence (NLS) that recruit the importin-α, and then the importin-β transport receptor that directs the vRNP to the nucleus. The translocation of the vRNP into the nucleus is essential for

negative sense viruses before any proteins can be translated the RNA must be transcribed and processed.

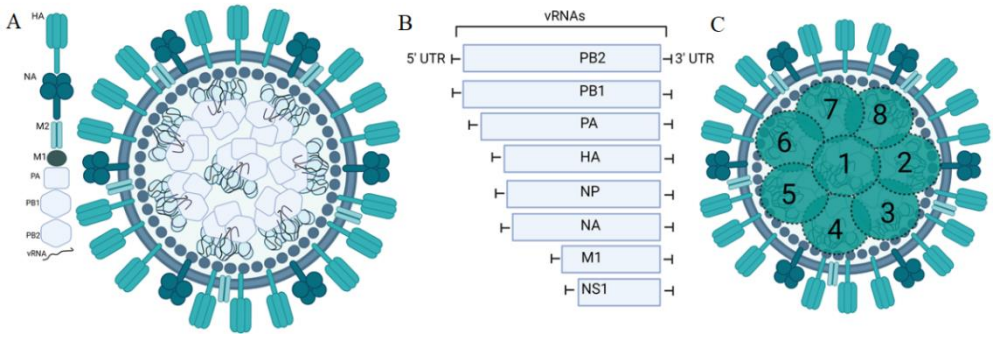


Figure 14: A.) Illustrated viral particle organization of influenza A virus. N protein compacts the vRNA and interacts with the M1 protein lined along the inner membrane. vRNP is illustrated as the complex of PA, PB1, and PB2 B.) vRNA shown for the 8 RNA segments, PB1 segment codes for PB1 and PB1-F2, PA codes for PA and PA-X, M1 codes for M1 and M2, NS codes for NS1 and NS2, and the rest are not spliced. C.) The vRNA is compacted around the N proteins that interact with M1, the super structure of the 8 segments is observed to have a 1+7 super structure shown here. The organization of the vRNP complex is not confirmed to be conserved. Created with BioRender.com

The Influenza A virus is part of the Orthomyxoviridae family and have negative sense RNA that must first be transcribed to vRNA and then polyadenylated all inside the nucleus. Followed by RNA splicing for the Matrix (M1 and M2) proteins and the Non-Structured (NS1 and NS2) vRNA and exported to the cytoplasm to be translated. The viral trans-membrane proteins HA, NA, and M2 vRNA is transferred into the endoplasmic reticulum for expression and ultimately reside on the cell membrane. The vRNP complex translated proteins M1, NP, PA, PB1, and PB2 are translocated into the nucleus to form the vRNP complex.[3] The M1 protein has two primary interactions within the vRNP complex, M1 links the NS2 to vRNP and interacts with NP to aid in transport of vRNP to the cytosol.[4] The secondary role of M1 protein is to assemble the viral

particle at the membrane of the host cell. The translated M1 protein within the cytosol, interacts with the lipid layer, and translated HA and NA proteins for viral particle generation.[5] This initiates the viral particle budding with M1 protein binding the lipid membrane of the host cell. The vRNP complex after translocation to the cytosol, travels up to the budding site and interacts with M1. Assembly of all the components at the budding site, the lipid layer begins to deform, and the viral particle takes shape. The viral particle release from the rest of the membrane is aided by the sialidase function of the NA protein.[6]

3.2 Post Translation Modification of Influenza A Virus Matrix 1 Protein

The M1 protein's role in viral pathogenesis is crucial in formation of vRNP and budding of the viral particle. Numerous reports of post translation modification (PTM) of M1 protein demonstrates the complex interaction between host cellular processes with M1 protein. Reports of M1 PTM ranging from NEDDylation, ubiquitination, phosphorylation, and SUMOylation provide ample evidence of M1 functional range modulated by post translation modifications.[4,7,8] Matrix 1 protein is one of the more abundant proteins expressed of the Influenza A proteome. The phosphorylation PTM is reported to modify eleven proteins among the influenza A virus proteome. The M1 protein phosphorylation at tyrosine 132 has been demonstrated to modulate the importin-α1 binding for M1 for translocation into the nucleus.[8] The location of phosphorylation on the M1 protein is upstream of the nuclear localization sequence (NLS) and is hypothesized to be the reason for exposing the NLS to importin-α1. The nuclear transportation of M1 is a critical part of viral particle formation as some of the M1

proteins are recruited to the HA proteins the M1 is also required in the nucleus for formation of vRNP.

The ubiquitination PTM is a powerful cellular mechanism that is utilized mainly for proteasomal degradation. The Influenza A virus utilization of this mechanism is explored in a study with interference RNA knock down of ubiquitin ligase Itch and found that the ubiquitination of M1 plays an essential role in the vRNP release into the cell.[9] The study found that Itch E3 ligase is recruited into the endosome before viral particle release, and coimmunoprecipitation experiments find M1 and Itch protein precipitated together as evidence of interaction. Insight in the ubiquitin activity is further demonstrated in the M1 translocation into the nucleus. Mehesutihan *et al.* report the ubiquitination of M1 mediated by the Ubiquitin E3 ligase AIP4 onto lysine K102 and K104.[10] The group reported that the mutation of K102R or K104R was lethal for the virus, and found this modification to be essential to IAV replication. In contrast, Neddylation investigation by Li *et al.*, discovered lysine 187 to be Neddylated and found improvement in infectivity with K187R mutant M1.[7]

The post translation modification mono-SUMOylation occurring with the SUMO paralog, SUMO1, and poly-SUMOylation occurring with SUMO2 and SUMO3 grants a target protein with the ability to interact with an array of cellular processes such as nuclear transport.[11] The global effect of SUMOylation is demonstrated when Influenza A Virus demonstrated inhibited infectivity with SUMO E2 enzyme, hUBC9, knock down cell line.[4] SUMOylated lysine modifications has been reported for the viral proteins NP, NS1, and M1.[12,13] NS1 SUMOylation sites were investigated in our lab, and one site was

determined to be a bona fide SUMOylation site, and modulation of virus infection was observed.[14] Previous reports of SUMOylation of M1 protein have shown modulation of pathogenesis, with singular M1-K242R mutants show decreased infectivity in comparison to wild type. Furthermore, Matrix 1 (M1) protein observed to have inhibited formation of M1-vRNP complex and demonstrated malformed viral particles.[4]

The M1 protein has an essential role in the pathogenesis of the virus and has adapted several host PTMs for the pathogenesis of the virus. The discovery of M1 SUMOylation from the three groups (Domingues *et al.*, Pal *et al.*, and Wu *et al.*) have provided great insight into the interaction of M1 with the SUMO PTM mechanism, and promote investigation into M1 SUMOylation.[4,15,16] The M1 protein holds 13 total lysines each have the possibility of being SUMOylated. The method used to determine possible SUMOylation test sites by Wu *et al.*, was a homology model approach, as they believed that the lysines most likely to be SUMOylated would be a conserved lysine. Thus, evaluation of lysine 187 and 242 were the only two reported in study and only lysine 242 was reported to be modified by Wu *et al.*. [17] The insight brought by this study is valuable and provides confidence in M1 and SUMO mechanism. However, due to the homology approach taken by the group, an argument can be made that there can be other potential sites that are as useful or have a lethal outcome. As loss of ubiquitination of M1 proved lethal to the pathogenesis, would loss of all SUMOylation activity be lethal as well?

Table 9: Reported Post Translation Modifications of M1.Ubiquitination and phosphorylation mutant M1 both have proven to be lethal to the virus. NEDDylation of mutant K187R has proven to surprisingly improve infectivity of the protein.

Modification	Position	Study
Ubiquitination	K102/K104	Mahesutihan *et al.* (2018)
Phosphorylation	Y132	Wang *et al.* (2012)
NEDDylation	K187	Li *et al.* (2020)

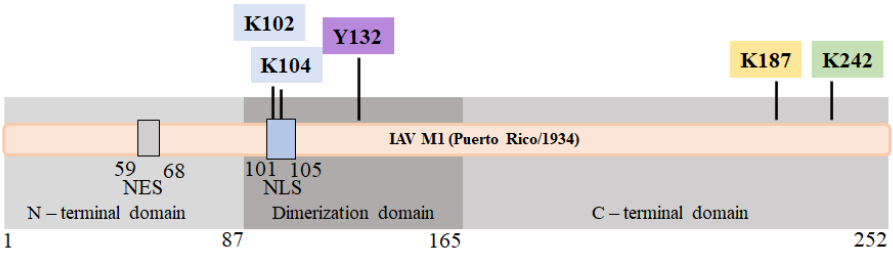


Figure 15: M1 a 252 amino acids protein has several observed PTMs. The N-terminal domain (1 – 87 aa) defined as the nuclear export signal, the dimerization domain (87-165 aa) reported to be part of the M1 oligomerization, and the c-terminal domain (166 – 252 aa) recently reported to also aid in oligomerization and stability of macro structure. The ubiquitination of M1 shown at positions 102 and 104 are found within the nuclear localization signal sequence. The phosphorylation at Y132 is found within the dimerization domain. The NEDDylation and SUMOylation modification are found within the C-Terminal domains.

Multiple covalent (multivalent) attachment of SUMO has been observed across the human proteome, and has great impact in DNA repair mechanisms.[18] The multivalent SUMO activity is observed to increase affinity of target proteins by increasing interaction surface area. Examples of discovery of multiple SUMO modifications are found with IAV-NP SUMOylation discovery. The Han *et al.* study found two sites on IAV nucleoprotein to be SUMOylated, and a lysine mutation screening method was used where all 19 lysine residues are mutated on NP individually.[13,19] Furthermore, groups

have also resorted to using SUMOylation prediction software to gain an idea of all possible sites, however this method can be labor intensive and yield only a small fraction of sites to be bona-fide SUMOylated.[14] In this study we provide an *in-vitro* method of SUMOylation that is coupled with MS to provide direct insight to site of SUMOylation. The method applies the qFRET platform to 1.) observe in real-time SUMOylation of target protein, and 2.) directly apply qFRET reaction to MS for identification of modified lysine residues.

3.3 In-Vitro SUMOylation of YPet-IAV-M1 with qFRET Reporter Design and Setup

The reconstitution of SUMOylation reaction with quantitative Förster's Resonance Energy transfer (qFRET) as a reporter is a robust and rapid method for identifying SUMOylation events and identifying SUMOylation sites. This method includes a E3 ligase for its enhanced SUMOylation activity and the design of the assay allows for direct input of a complete SUMOylated target protein for mass spectrometry (MS) identification of SUMOylated lysines. The qFRET assay utilizes a FRET optimized donor fluorescent protein tag, CyPet, on the SUMO1 protein and acceptor fluorescent protein tag, YPet, on the SUMOylation target protein, Influenza A Virus M1.[20] The fusion proteins CyPet-SUMO1 and YPet-IAV-M1 allow us to monitor the completion of the SUMOylation assay without destroying the sample and is used for MS processing directly. The qFRET platform has been applied previously for evaluation of kinetic variables such as K_D and k_{cat}/K_m.[21,22] The qFRET platform is directly applied to the

evaluation of enzymatic covalent attachment of fusion protein CyPet-SUMO1 onto YPet-IAV-M1.

The reconstituted complete SUMOylation cascade, with the E3 ligase PIAS1 which enhances the SUMOylation of target protein Matrix 1 (M1) protein. The enzymatic reaction is setup in-vitro with E1 activating enzyme hetro-dimmer, Activating Enzyme Subunit 1 (AOS1), and Ubiquitin Activating Enzyme 2 (UBA2), SUMO specific E2 Conjugating enzyme, Ubiquitin Carrier 9 (UBC9), and coupled with the addition of the target specific E3 ligase, Protein Inhibitor of activated STAT 1 (PIAS1). The enzymatic cascade is initiated by the addition of ATP to activate the E1 heterodimer, CyPet-SUMO1 Gly98 makes a thiol intermediate with Cys173 on UBA2. The E1-SUMO complex interacts with the UBC9 and transfers CyPet-SUMO1 onto the catalytic Cys93, and finally the E3 ligase specific to the target protein mediates the aminolysis reaction for covalent attachment of SUMO1 onto lysine of M1. The conjugation of the fusion protein Cypet-SUMO1 onto lysine residue on YPet-M1 is a covalent conjugation and brings the FRET pairs within the range of transfer for an observable increase in Em_{FRET} signal. Illustrated in Figure 16.

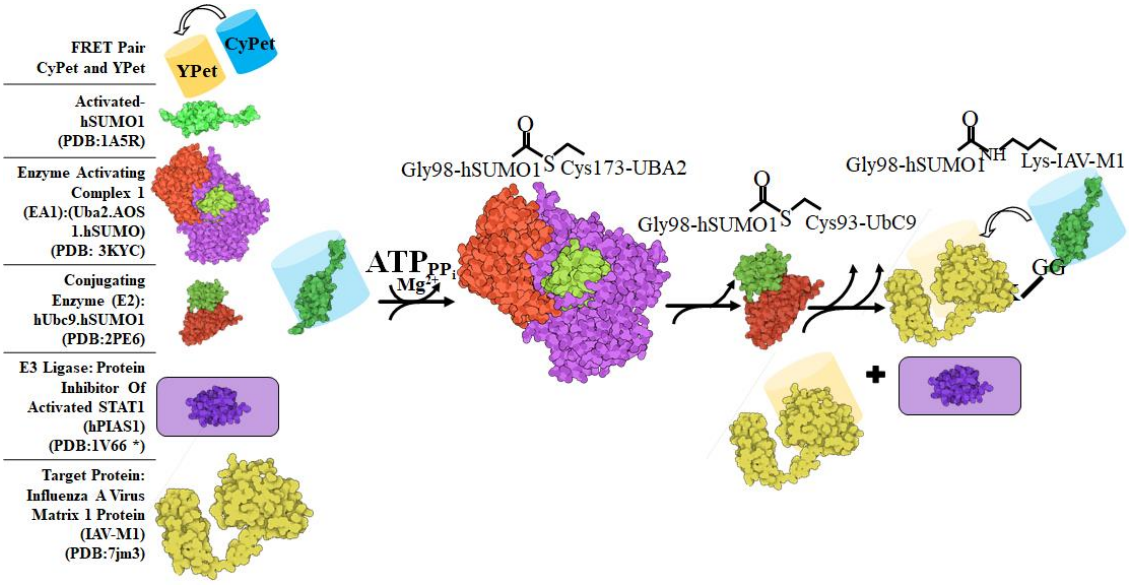


Figure 16: Illustration of SUMOylation enzymatic cascade with qFRET assay components. The activation of the assay occurs with the addition of ATP into the reaction mixture. The CyPet-SUMO1 shown in green binds to E1 activating enzyme shown in complex (3KYC) as a temporary thioester bond at Cys173 with Gly98 and transfer to E2. The E2 conjugating enzyme UBC9 (2PE6), shown in dark red, with the temporary thioester bond on the catalytic cysteine 93. The E3 ligase PIAS1 shown as a purple rectangle with the RING domain (1V66) recruits the target protein to E2. The RING domain also mediate the isopeptide bond of SUMO gly98 onto the lysine of the target protein. The FRET pairs are shown as cylinders, donor Cypet as a blue cylinder and YPet acceptor as a yellow cylinder. The FRET phenomenon occurs as a reporter of SUMO1 binding onto the target protein.

3.4 qFRET Assay Fluorescent Signal Acquisition and Processing

The in-vitro SUMOylation assay is adapted for a microplate fluorescent reader allowing rapid measurements in real time and direct processing. The FRET wavelength, EmTotal, are 414 nm excitation and 530 nm emission. Emission from unbound donor CyPet, Fl_D, 414 nm excitation and 475 nm emission, and unbound acceptor YPet, Fl_A, 475 nm excitation and 530 nm emission. The crosstalk correction parameters alpha (α) and beta (β) must be evaluated for every fluorescent spectrometer. Alpha α is the ratio of donor emission at 530 nm over the emission at 475 nm when both are excited at 414 nm, equation 2. Beta β is the ratio of acceptor emission at 530 nm when excited by 414 nm

over the emission at 530 nm when excited at 475 nm. The quantitative Em_{FRET} parameters, 0.34 ± 0.003 α and 0.03 ± 0.001 β variables are determined using the same formulation outlined in previous work from Yang *et al.*, equations 2 and 3 respectively. Equation 1 provides the relationship of Em_{FRET} that can be quantified for the FRET increase based on the direct physical SUMOylation of substrate.

Table 10: qFRET measurements taken at three unique excitation and emission points. The three wavelengths are recorded and processed for qFRET signal using equation 1.

	Excitation (λ)	**Emission (λ)**
Em_{Total}	414 nm	530 nm
FlD	414 nm	475 nm
FLA	475 nm	530 nm

$$Em_{FRET} = (Em_{Total}) - ((FL_D * \alpha) + (FL_A * \beta)) \quad \text{Equation 3.1}$$

$$\alpha = \frac{\text{Donor em}isson\ at\ 530\ nm\ with\ excitation\ at\ 414\ nm}{\text{Donor em}isson\ at\ 475\ nm\ with\ excitation\ at\ 414\ nm} \quad \text{Equation 3.2}$$

$$\beta = \frac{\text{Acceptor em}isson\ at\ 530\ nm\ with\ excitation\ at\ 414\ nm}{Acceptor\ \text{em}isson\ at\ 530\ nm\ with\ excitation\ at\ 475\ nm} \quad \text{Equation 3.3}$$

The specificity of SUMO protein to the SUMOylation target has the potential to yield a false positive FRET response. Thus, functional controls of reactions without ATP are implemented in parallel to gain confidence of SUMOylation event.

3.5 Results

In vitro qFRET assay SUMOylation of IAV M1

The observed raw FRET signal demonstrates the donor quenching with the addition of ATP. This quenching and fluorescent increase at the FRET wavelength at 530

nm is a direct result of FRET occurring between donor fusion protein CyPet-SUMO1, and acceptor YPet-IAV-M1 (Figure 17A). The ATP initiates the SUMOylation reaction and the covalent attachment of the SUMO1 onto the M1 puts both acceptor and donor within range of FRET. The real time evaluation of the Em_{FRET} signal for the duration of 60 minutes shows the saturation of Em_{FRET} at the plateau of the signal. The covalent attachment of SUMO1 at a 1:2 ratio of concentration of SUMO to target protein results in all the SUMO1 bound to substrate. The saturation of the bound SUMO1 is observed in the plateau of the Em_{FRET} signal. Observed within the 60-minute reaction the Em_{FRET} signal no longer increases past the plateau (Figure 17 B).

This observation allows us to set consistent times for each reaction, based entirely on the Em_{FRET} signal. The observation of the two tests demonstrates the robust signal of SUMOylation events. The other observations are the high background within the signal and is observed in the timed Em_{FRET} experiment. Thus, we optimized the concentration of CyPet-SUMO1 as the donor is the major contributor to the crosstalk, with alpha at 0.34. The last reaction is the reduced CyPet concentration of 0.5 μM, and all other reactants stayed the same. Furthermore, the experiment also setup a reaction without the E3 present, shown to have a significant difference from the no ATP negative control. However, the E3 ligase PIAS1 shows a very significant increase at 60-minute incubation at p value being smaller then 0.0001 with the unpaired two tailed t-test.

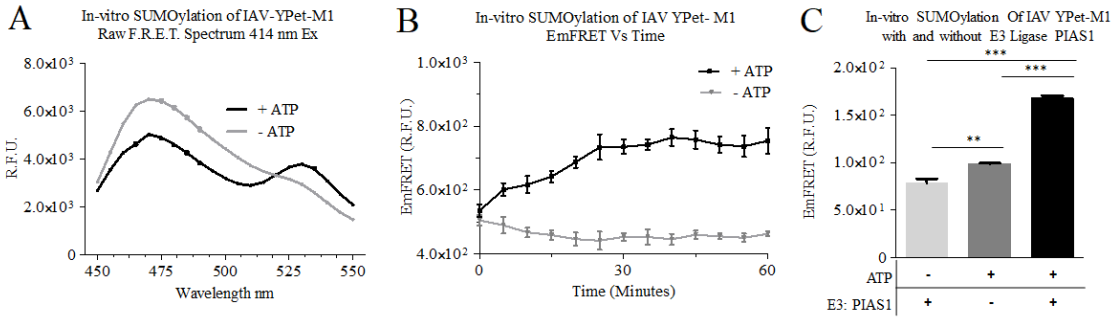
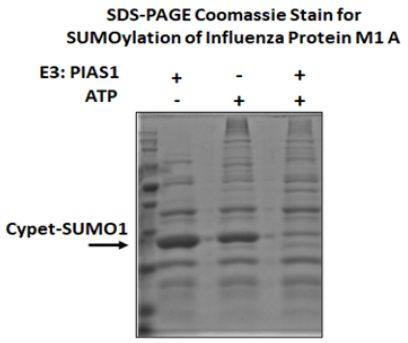


Figure 17: In-Vitro SUMOylation of YPet-1AV-M1, acceptor and donor at 1:4 ratio fluorescent spectrum was measured across wavelength 450 - 550 nm, with 414 m excitation. The gray line is the SUMOylation reaction without ATP, and the black line is the reaction with ATP(A). B.)Em_{FRET} monitored over time with acceptor and donor at 1:2 ratio C.)The in-vitro SUMOylation of IAV YPet-M1 with and without E3 ligase PIAS1, and with and without ATP at qFRET Optimized concentrations. The reaction is running under three conditions, without ATP, without E3, and one standard. An unpaired two tailed t-test was done across each reaction. $p<0.0001$***,$p<0.005$**

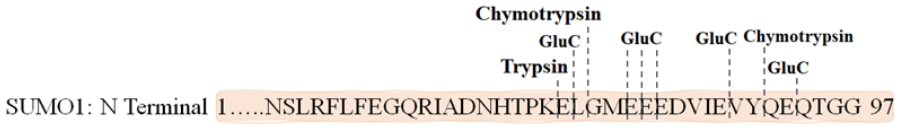
MS Sample Preparation and MS Identified SUMO Modification

The SUMOylation of IAV-M1 was completed in-vitro. The SDS-PAGE gel provides a secondary confirmation of SUMOylation as we observe the CyPet-SUMO1 band dimmish with the reaction of ATP and E3 ligase. The reaction directly digested with the three protease with their respective buffer conditions and ran over night. The expected cut pattern from each protease is shown in figure 18B. The Glu-C protease cleaves the carboxyl end of glutamic acid, and we find seven expected sites at the c-terminal end. The window of amino acid on the c-terminal considered does not extend farther then 20 - 30 base pairs, as the numerous cut sites will result in a shorter peptide, and longer peptides larger than 20 amino acids become more difficult to identify as a modification.

A



B

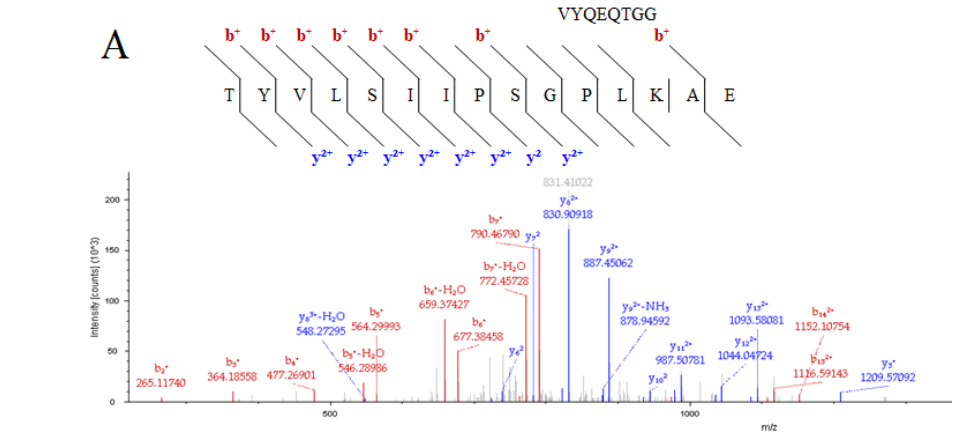


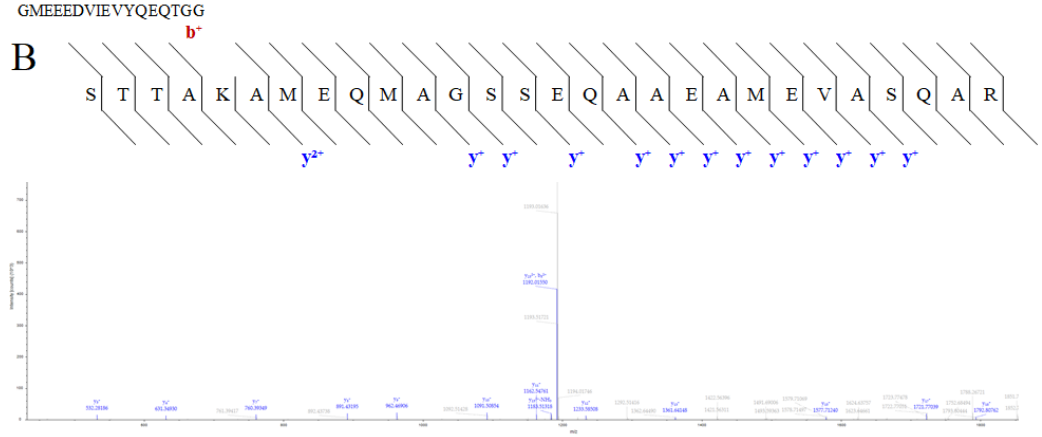
Protease	Cleavage Specificity
Trypsin	Carboxyl side of arginine and lysine
Glu-C	Carboxyl side of glutamic acid
Chymotrypsin	Carboxyl side of tyrosine, phenylalanine, tryptophan, and leucine

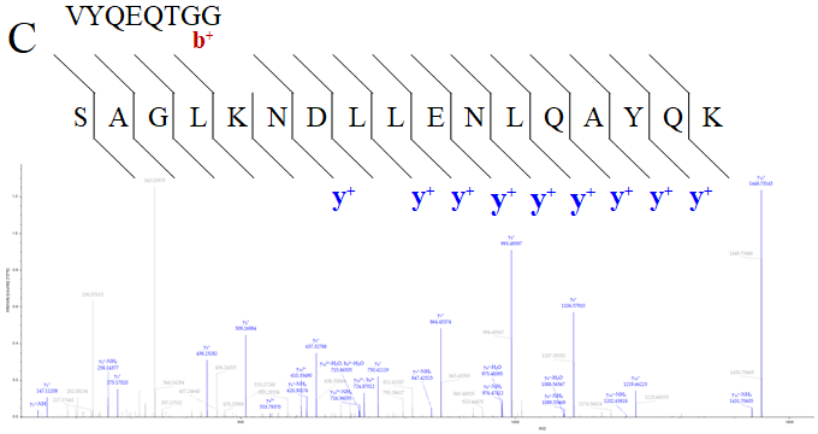
Figure 18:SDS-PAGE coomassie stain of SUMOylation reaction for MS. Lane 1 is no ATP, Lane 2 is no E3 ligase, and Lane 3 is the standard reaction. CyPet-SUMO1 band shifts to target protein, leaving low concentration of unbound CyPet-SUMO1(A). B.) Three different digestions were done on the SUMOylated YPet-IAV-M1. The expected cuts from each are provided with dash lines.

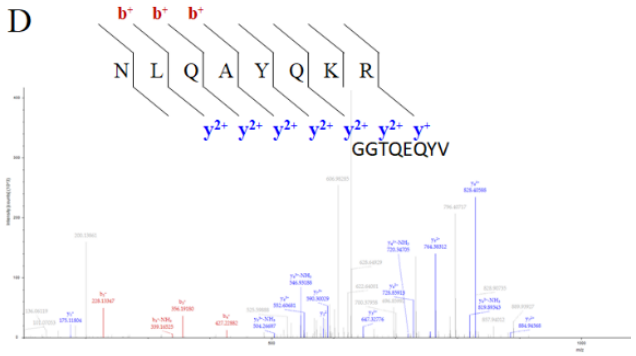
The results of identified sites with SUMO modifications are showing in figure 20, with the mass to charge (m/z) vs the intensity of the signal along with a diagram of aligned particles. The illustration of the five identified modified lysines is shown in figure 20F. The coverage of the IAV-M1 protein is observed to be 95 % of the protein with every lysine found in the MS results. The lysine sites identified with the SUMO modification were 21, 187, 230, 242, and 252. The M1 protein holds 13 lysine residues and the 8 other lysines were also confirmed to be found in peptides that were not SUMO modified. The n-terminal lysine 21, is identified with a SUMO1 peptide that follows the GluC cut at glutamic acid (E) at position 89 on SUMO1 c-terminal. Lysine 187 observed

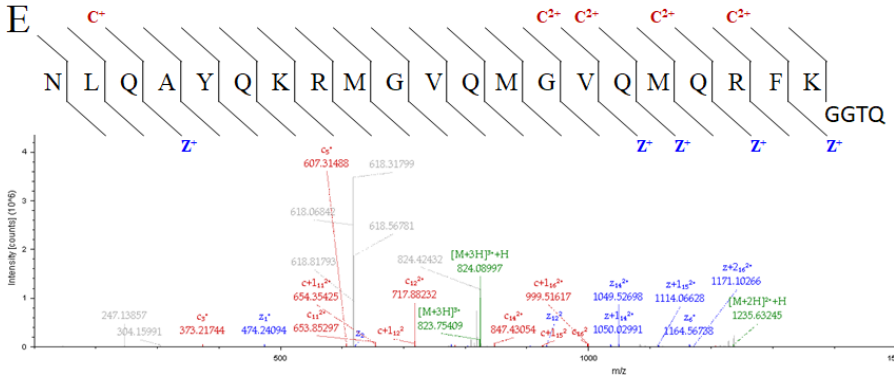
to be modified with SUMO1 peptide that follows the chymotrypsin cut at leucine position 80 on SUMO1. Lysine 230 observed to be modified with SUMO1 peptide that follows a GluC cut at glutamic acid on position 89 on SUMO1. The Wu *et al.*. identified lysine at position 242 is also found to be modified with a SUMO1 peptide following a GluC cut at position 80. The last lysine at position 252 was also identified to have a SUMO1 modification following the GluC cut at position 93 on SUMO1.


A
VYQEQTGG
T Y V L S I I P S G P L K A E
831.41022
830.90918
790.46790
887.45062
772.45728
659.37427
878.94592
677.38458
564.29993
548.27295
546.28986
477.26901
364.18558
265.11740
1093.58081
1152.10754
1044.04724
987.50781
1116.59143
1209.57092
Intensity [counts] (10^3)
m/z


GMEEEDVIEVYQEQTGG
B
S T T A K A M E Q M A G S S E Q A A E A M E V A S Q A R


C
VYQEQTGG
S A G L K N D L L E N L Q A Y Q K


D
b+ b+ b+
N L Q A Y Q K R
y2+ y2+ y2+ y2+ y2+ y2+ y+
GGTQEQYV


E
C+
C2+ C2+ C2+ C2+
N L Q A Y Q K R M G V Q M G V Q M Q R F K
GGTQ
Z+
Z+ Z+ Z+ Z+
Intensity (counts) (10^6)
m/z
607.31488
618.31799
618.06842
618.56781
618.81793
824.42432
[M+3H]3++H
824.08997
654.35425
717.88232
999.51617
1049.52698
1114.06628
1171.10266
1235.63245
[M+2H]2++H
[M+3H]3+
823.75409
847.43054
1050.02991
1164.56728
247.13857
304.15991
373.21744
474.24094
653.85297

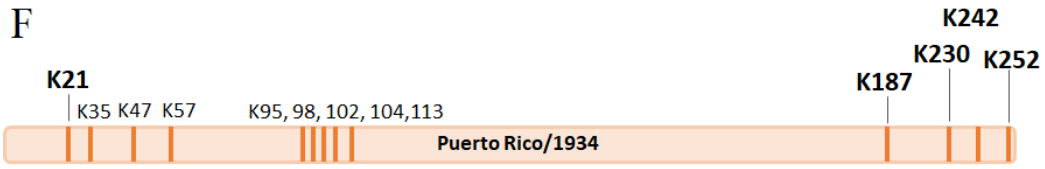

F
K21
K35 K47 K57
K95, 98, 102, 104,113
K187
K230
K242
K252
Puerto Rico/1934

Figure 19: Identified lysine modification on residues A.)Lysine 21, with identified GluC cut on SUMO1,B.)lysine 187 with identified Chymotrypsin cut on SUMO1 peptide, C.)Lysine 230 with identified GluC cut on SUMO1, D.)lysine 242 with GluC cut on SUMO1, and E.)lysine 252 with GluC cut on SUMO1. The alignment and spectrums are taken from Thermofisher Proteome Discoverer ™.F.) The illustration provides the location of all the lysine peptides found and only 5 determined to be modified by SUMO1.

In vitro SUMOylation of IAV-M1 Mutants

The qFRET assay for in-vitro SUMOylation of M1wildtype (wt) and M1 mutants provides a rapid assessment of the M1 SUMOylation sites *in vitro*. The results of the Em_{FRET} assay are calculated and plotted as bar plots with standard deviation on GraphpadPrism5^{TM} (Figure 21). The single mutants K21R, K187R, K230R, K242R, and K252R, all demonstrated a significant Em_{FRET} response to ATP and E3 ligase. The most significant drop in Em_{FRET} signal is with the C-terminal mutants. The reactions with E2 conjugating enzyme in comparison to the addition of E3 ligase PIAS1 had no significant drop with or without the addition of ATP besides in the lysine 242 mutant. A two-tailed t-test was completed on each variant reaction, with -ATP as the control group.

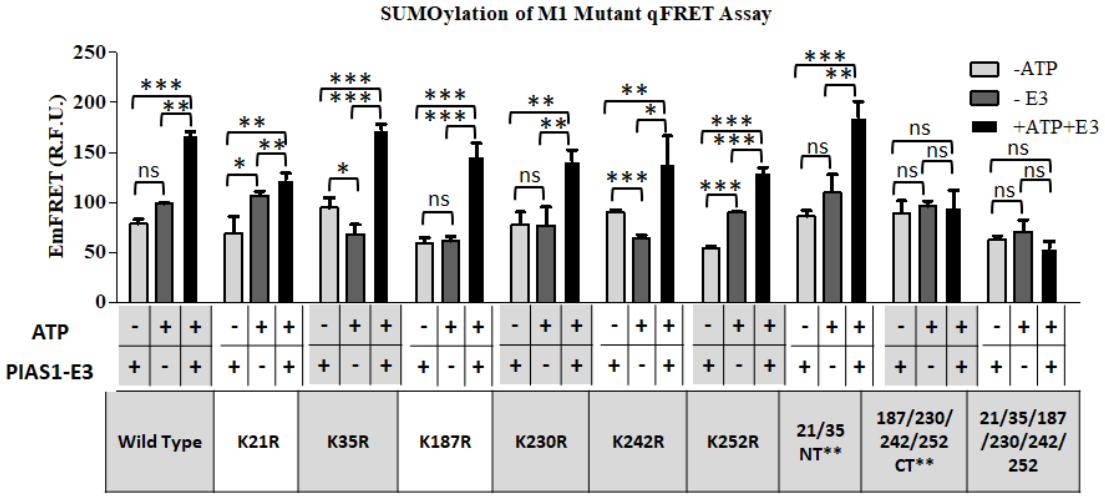


Figure 20:In-vitro SUMOylation of IAV-M1, IAV-M1 K21R, IAV-M1 K35R, IAV-M1 K187R, IAV-M1 K230R, IAV-M1 K242R, IAV-M1 K252R, IAV-M1 K21/35R, IAV-M1 K187,230,242,252R, and K21/35/187/230/242/252R mutant. The reactions all were plotted on GraphPad Prism5^{TM}. Unpaired two tailed t-test was done for all mutants, with their control reactions. $p<0.0005$*** $p<0.001$** $p<0.05$*

Viral Plaque Assay of M1 Wildtype and Mutants

The plaque assays were completed for all individual mutants, and the complete mutant M1, K21,35,187,230,242,252R. The crystal violet clearing signifies a plaque,

observed in the M1 wild type and various mutant wells. The degree of clearing is directly related to the infectivity of the viral particles generated with P2. We observe clear singular plaques at the higher dilutions and allow us to calculate plaque forming units. The plaque assays in figure 22 provide a qualitative measure of the ability of the virus to infect cells. The mutant 35, 187, 230, 242, and 252 all demonstrated infectivity to some degree in the P2 infections. The notable mutant lysine 21 to arginine displays no plaques formed at any dilution of the virus. The same result is shown with the mutant with lysine K21/35/187/242/252R mutations, no plaques observed with the direct infection.

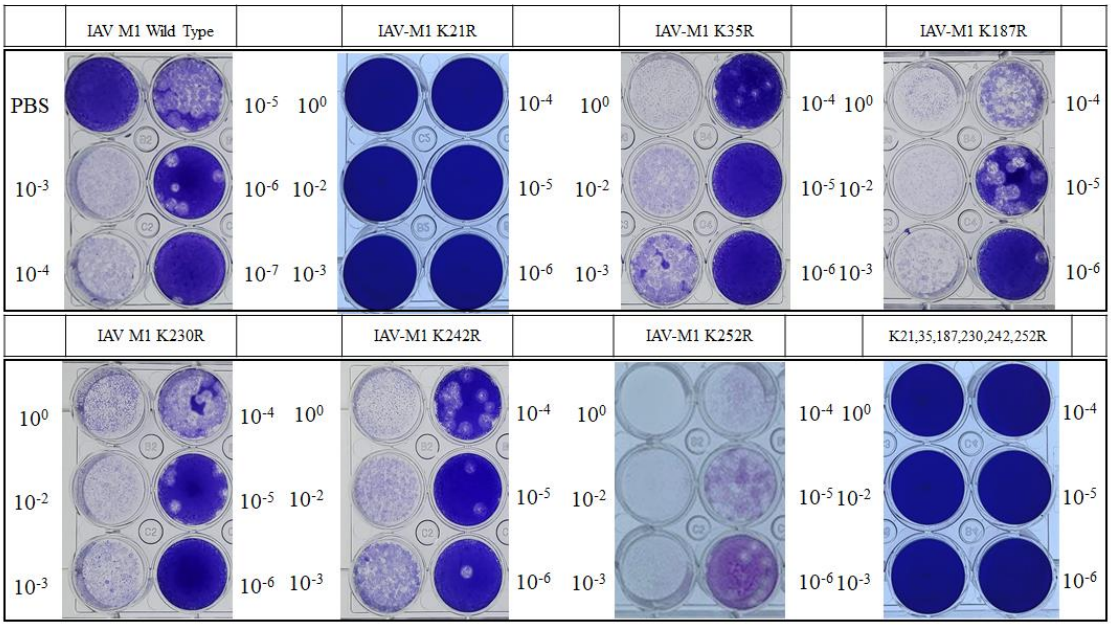


Figure 21: IAV M1 mutant plaque assays, using P2 of the generated virus particle for each mutant. The 6 ell plates were used to observe the infectivity of the virus together with serial dilutions. Noted next to the image of the plaque are the dilutions used in each well. The images were taken on a digital bright field camera with illuminated back lighting.

Cytotoxicity Assay of Viral Particles Generated from M1 Wildtype and M1 Mutants

The reconstitution assay demonstrated infectivity of the individual lysine to arginine mutant viruses, besides the single lysine 21 mutant. The plaque forming units were determined for all the viral particles that form plaques, since M1 K21R cannot form

plaques Pfu/ ml value was zero. The bar plot below is the calculated Pfu of the plaque assays (Figure 23A). The variation in the Pfu values is not indicative of infectivity, as other variables such as transfection efficiency can modulate the number of viral particles formed. The K35R mutant had a very low Pfu/ml in comparison to wild type and is notated on the graph as $0.2*10^6$ Pfu/ml. The cytotoxic assay provides a normalized method to compare infectivity of the viral particles. We compared the infectivity of wildtype M1 against lysine 242 mutant, lysine 187 mutant at an MOI of 0.005, and introduce K21R M1 mutant viral particles undiluted, and a control set with no infection, to the assay to determine if any activity can be observed (Figure 23B). We monitor the release of ATP over 48 hours with measurements taken at 0 hours, 24 hours, 36 hours, and 48 hours. The wild type M1 and the K187R mutant wells dropped in signal within 36 hours. In comparison to no infection, we see a drastic difference in signal this assay. The M1 mutant lysine 21 is added with no dilution to determine if any activity of infection is found, and we observe no observable difference between no-infection and K21R mutant M1. The K242R mutant signal shows diminished signal like the M1 wildtype, however the signal does not reach as low as M1 wildtype or M1 K187R.

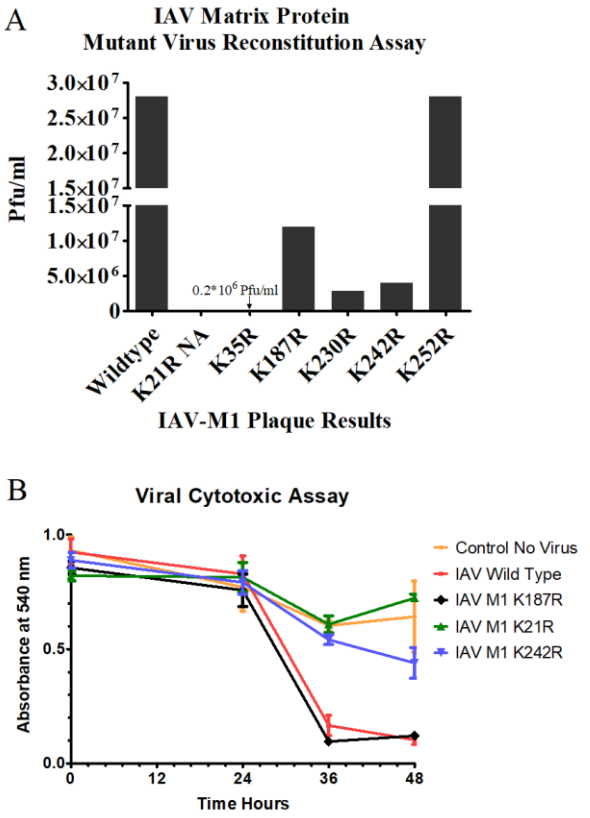


Figure 22: A.) Bar plot of Pfu/ml of generated viral particles. The M1 mutant K21R has no calculated Pfu, and the K35R Pfu/ml was much smaller than the others, thus it is noted on the graph. B.) The cytotoxic assay taken over a period of 48 hours post infection. The drop in signal was observed at 36 hours, for M1 wild type, M1 K187R and slightly for K242R. The control are wells with no infection for observing signal of natural cell death.

Cellular Translocation of wildtype M1 and M1 Mutants

The immunostaining of M1had a significant signal in the 533 nm emission and overlapped with the Hoechst stain. The response of the Hoechst stain was very high, a constant exposure of 2 milliseconds was used to obtain an observable signal. The mutant M1 K21R had a similar cellular location within the nucleus as the wildtype (Figure 24). The mutant M1 K242R had a significant signal in both the nuclear stain and the M1 stain, and the overlap in the intensity implies the location is within the nucleus.

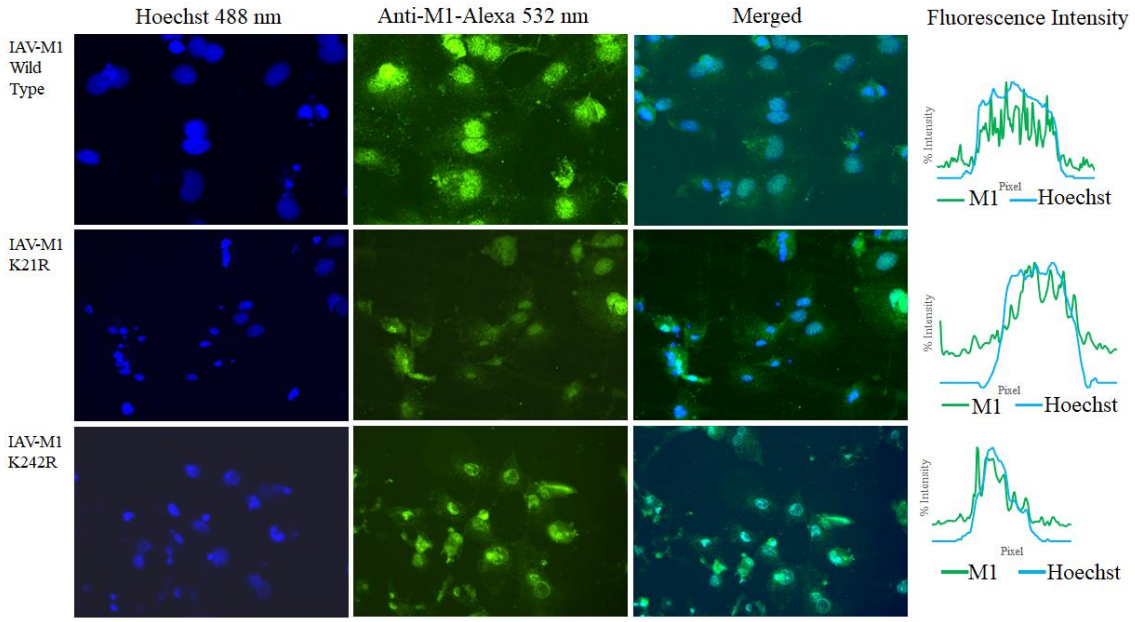


Figure 23: ImmunoStains of M1 with Alexa 532 nm stain and nuclear stain with Hoechst emission at 488 nm. The two channels are stacked in ImageJ and analyzed for intensity across the cell in both channels and plotted.

Immunoprecipitation of Wildtype M1, M1-K21R, and SUMOylated M1-wildtype

The immune staining for M1 and M1 K21R produce clear blots from protein capture from the cell. The SUMOylated M1 capture resulted in extra bands at ~100 – 130 kDa, the anti-M1 antibody is 150 kDa and can result in a similar band. The M1 wildtype and M1 K21R both in the pDZ vector are confirmed to have M1 expression. The size of the M1 protein without a tag is 28 kDa and has been demonstrated by groups to have a blot at approximately 26 – 28 kDa. The SUMOylation of M1 is confirmed with M1 blot and bands at 36 and 48 kDa show feint signal.

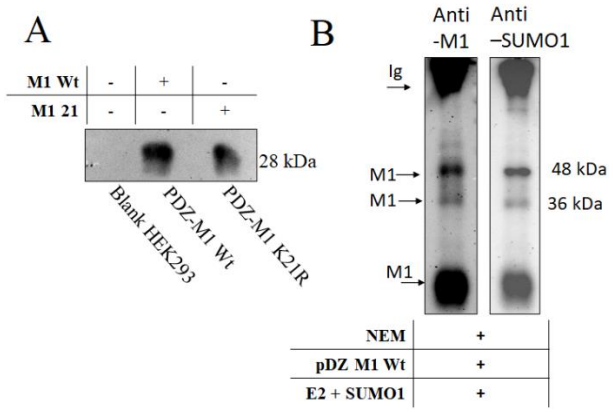


Figure 24: Immunoblot of M1 wildtype, M1 K21R, and blank HEK293 cells(A). B.)Immunoblot of SUMOylated M1, stained with M1 and with SUMO1. Images taken in UVP camera and ImageJ is used to process the images.

3.6 Discussion

The *in vitro* SUMOylation assay with qFRET reporter demonstrated here provides a rapid assessment of a target protein SUMOylation. The controlled reaction is applied to a heterologous protein influenza A virus Matrix 1 protein and demonstrated to have multiple SUMOylated lysines. The qFRET assay was quickly adapted for a timed and robust response to M1 SUMOylation in figure 18. The Em_{FRET} response to ATP qualifies the assay as a reporter of a SUMOylation event. The assay also provides insight into the activity of the E3 ligase, with a significant SUMOylation response when ATP is added. The SUMOylated target protein was followed by MS analysis with the SUMO enzymes, E1, E2, and E3 in the MS sample. The SDS-PAGE in figure 19A with coomassie stain demonstrated that majority of the CyPet-SUMO1 protein is covalently bound. Compared to the no ATP added lane, where the CyPet-SUMO1 band is still present. This observation provides confidence that majority of the 4 µM CyPet-SUMO1 is covalently attached to the target fusion protein YPet-IAV-M1.

The mass spectrometry analysis of the SUMOylated IAV-M1 provided 5 lysines that are modified by SUMO1. The sites observed includes the lysine 242 identified before by Wu *et al.* and four novel SUMOylated lysines. The coverage of IAV M1 was 95 % which is an advantage of *in vitro* reactions, allowing higher concentration of target protein in comparison to in-cell. The digested peptides resulting from the SUMO1 protein matched GluC cut pattern and based on these observations GluC alone provides great resolution of SUMO1 modification. The c-terminal of SUMO1 holds seven glutamic acid positions within thirty base pairs. The GluC protease cut profile matched for lysine positions 21, 230, 242, and 252.

Cross talk between the other lysine modification cascades such as Ubiquitin and Neddylation could interfere in the identification of SUMOylated sites in cell. In an *in vitro* reaction we have confidence that the identified modification is of SUMO1. The modification of lysine 187 detected in MS provides an example of cross talk between SUMOylation and Neddylation. The previously discovered Nedd8 modification on lysine 187 which was observed to negatively regulate pathogenesis by Li *et al.*. Demonstrated similar behavior here in figure 23B, M1 K187R mutant infectivity was slightly improved in comparison to wildtype M1.

The plaque assay and viral cytotoxicity assay both demonstrated that lysine 21 is essential for viral replication. We observed lysine 21 to be SUMOylated in the MS results, which was the only lysine on the n-terminal that was found to be SUMOylated *in vitro*. The plaque assays demonstrated infectivity in lysine mutants 35, 187. 230, 242, and 252. However, lysine 21 and the 6 mutant M1 had no infectivity. This was tested again

with the viral cytotoxicity assay and observed no virus induced cell death. The mutation of 242 to lysine

Lethal mutations have been observed in other PTMs investigations, as ubiquitin K102 and 104 mutant M1 could not regenerate the virus. Additionally, tyrosine 132 mutation also proved lethal, and no virus could be regenerated in mutation studies. Demonstrated here the *in vitro* SUMOylation of lysine 21 and observed that lysine 21 mutation to arginine alone renders the virus pathogenesis shutdown. We did not observe modulation in translocation of M1 with the mutation, as M1 was found in the nucleus with wild type M1, lysine 21 mutant, and 242 mutant. The translocation of M1 has been reported to be dependent on the phosphorylation of tyrosine residue near the nuclear localization sequence.[8] The immune precipitation of M1 provides a secondary confirmation of M1 K21R mutant expression in the cell. We observe a band for M1 at ~28 kDa in both lanes of M1 wildtype and M1 K21K. The secondary confirmation of M1 expression is the immunostaining within the cell in figure 24. These results confirm expression of mutant M1 in the cell, but in the reconstitution assay no viral activity was observed.

The *in vitro* SUMOylation of M1 demonstrated 5 lysine residues to be SUMO1 modified *in vitro*, and the immunostaining of M1 with SENP inhibitor demonstrated multiple bands with M1 blot and SUMO1 blot. The multiple bands are indicative of multiple SUMOylation modifications, and the band shifts of approximately 10 kDa indicates SUMO1 modifications. The *in vitro* SUMOylation with qFRET reported demonstrated multiple SUMOylation events, as the single mutants all had significant

Em_{FRET} signal with E3 ligase and ATP added. These findings corroborate the conclusion of multiple SUMO1 attachments of M1 and provide evidence of essential lysine 21 modification.

3.7 Methods and Materials

Expression and purification of recombinant SUMOylation proteins

The in-vitro Förster's Resonance Energy Transfer (FRET) SUMOylation reaction is completed with the E1, E2, and E3 enzymes in the SUMOylation cascade. The E1 activation enzyme complex, UBA2 and AOS1, E2 conjugating enzyme UBC9, and E3 ligase PIAS1 were all cloned into pET28B vector for expression in BL21(DE3) cells. The FRET pairs CyPet and YPet are N-terminal tagged to SUMO1 and the substrates respectively and cloned into pET28B for expression in BL21(DE3). Each BL21(DE3) cell line with individual proteins were inoculated at 1:100 and grown to 0.4 OD at 600 nm at 37 °C, and then induced at overnight at 22°C with 0.25 mM IPTG. The cells were lysed, lysis buffer (20 mM Tris-HCl (pH 7.5), 0.5 M NaCl, 5 mM Imidazole), by sonication and centrifuged at 35,000 x g. The soluble fraction was purified by 6XHis tag to NiNTA beads affinity chromatography through a gravity column. The bound proteins were washed with, buffer 1 (20 mM Tris-HCl (pH 7.5), 0.3 M NaCl), buffer 2 (20 mM Tris-HCl (pH 7.5), 1.5 M NaCl, and 0.5% Triton X-100), and buffer 3 (20 mM Tris-HCl(pH 7.5), 0.5 M NaCl, and 10 mM Imidazole). The proteins eluted using the following buffer, (20 mM Tris-HCl, 300 mM NaCl, and 450 mM Imidazole) and dialyzed in 20 mM Tris-HCl (pH 7.5), 50 mM NaCl, and 1 mM DTT.

In-vitro qFRET SUMOylation Reaction Setup and Measurement

The SUMOylation assay is evaluated for its response to various concentrations of acceptor fusion protein, CyPet-SUMO1. The enzyme concentrations are kept constant at 0.1 µM E1 hetro-dimer, 0.2 µM E2 conjugating enzyme, and 0.25 µM E3 ligase. The raw signal of in-vitro reactions allows us to observe the quenching of the acceptor molecule at 475 nm and the fluorescence's of the acceptor molecule at 530 nm. The raw signal spectrum is implemented at 1 µM of donor, CyPet-SUMO1, and 2 µM of acceptor, YPet-IAV-M1. The same reaction setup for Em_{FRET} resolution and monitored in real time over a period of 60 minutes at 37°C. Three wavelengths monitored over the 60 minutes are tabulated in table 2, the EmTotal, FL_D, and FL_A are measured and processed according to equations 1 to resolve Em_{FRET}. The reaction is setup without ATP as a control to observe only FRET signal resulting from SUMOylation of target protein.

The optimized concentrations of acceptor and donor with lowest observed background signal is implemented at 0.5 µM CyPet-SUMO1, E1 activating enzyme at 0.1 µM, E2 conjugating enzyme at 0.2 µM, E3 ligase at 0.25 µM, and acceptor target protein at 2 µM, with the reaction ran for 60 minutes at 37 ° Celsius. All reactions are completed in SUMOylation buffer of 20 mM Tris-HCl (pH 7.5) 50 mM NaCl, 4 mM MgCl, 1 mM DTT, and 2mM ATP. Functional controls are put in place for non-specific interaction, by a negative control reaction without ATP, and to observe E3 activity a control reaction without E3 ligase. Each reaction was incubated at 37°C for 60 minutes.

The SUMOylated samples are added to a 384 well micro plate (Greiner™ 384) for fluorescence measurements on a fluorescent plate reader. The plate reader used for

this study is the Molecular Devices SpectraMax3™ with the SoftMax Pro 7™. The fluorescence measurements are taken according to the excitation and emission of the three wavelengths, EmTotal, Fl_D, and FL_A. The PMT setting are set to a constant of "Low" with "Endpoint" set for read mode.

Sample preparation for mass spectrometry post in-vitro SUMOylation reaction

The in-vitro SUMOylation reactions of IAV M1 protein tagged YPet (M1), and CyPet tagged SUMO1 protein are added at the 1:1 ratio of 4 µM each. Activating Enzyme Complex 1 (E1) is at 0.1 µM, and Conjugating Enzyme 2 (E2) at 0.2 µM, E3 ligase at 0.5 µM. The reactions were completed at 37 ° Celsius for 4 hour to insure highest percentage of SUMOylation of M1 protein. Additionally, two reactions were implemented in parallel without ATP as negative controls. The following day 5 ug of each reaction was ran on an SDS-PAGE gel with Coomassie stain to observe band shift with and without the activating reagent ATP, and the rest of the sample was stored at – 80 ° C till they were ready for MS preparation.

The proteolytic digestions were performed with three different enzymes separately. Trypsin (Pierce™ Trypsin Protease-MS Grade, Thermofisher) , Chymotrypsin (Pierce™ Chymotrypsin Protease-MS Grade, Thermofisher), and Glu-C (Pierce™ Glu-C Protease-MS Grade, Thermofisher) were the enzymes selected to provide a large diversity in digested peptides for MS (Figure 19B). Each enzyme was digested at 1:100 ratio for sample to enzyme ratio and ran overnight (16 hours) at 37 ° Celsius. Each completed digestion was acidified to a final concentration of 0.1% v/v TFA, and speed

vacuumed to dry product, and then reconstituted to 0.1% v/v TFA readied for MS loading.

LTQ Orbitrap Xl Loading and Run

Samples consisted of approximately 1000 nM of in-solution digested product from each proteolytic enzyme digestion. The samples were loaded and desalted at a flow rate of 50 ul/min for 5 min on a C18 trap column (Waters Symmetry C18 180µm x 2cm) in acetonitrile at 2% v/v, water at 97.9% v/v, formic acid 0.1% v/v. The samples were separated on a C18 reversed-phase analytical column (Waters BEH C18 1.7umx75umx200mmOrbitrap Column Dimensions) using a Waters nanoAcquity UPLC (Orbitrap Ramping System) over a 70 min gradient. Mobile phase A was 0.2% formic acid in water and mobile phase B was 0.2% formic acid in acetonitrile. The gradient was as follows: from 5% to 45 % buffer acetonitrile 80% v/v, water 19.9% v/v, formic acid 0.1% v/v at a flowrate of approximately 300 nL/min.

The Thermo Orbitrap FusionLTQ Orbitrap mass spectrometer was operated in parallel acquisition Data-dependent acquisition (DDA)IDA mode. Masses from 400-2000 Da were acquired in the orbitrap with nominal resolution of 120k (FWHM) at m/z 200 for MS. Peaks above intensity 5e8 with charges 2-8 were selected for sequential CID, HCD, and ETD fragmentation in the orbitrap at 30,000 resolutions followed by dynamic exclusion for 15 seconds.

Bioinformatic Analysis of MS Results

The LTQ-orbitrap XL (.raw) raw data was analyzed on Thermofisher Proteome Analyzer^{TM}. The complete amino acid sequence of each protein was provided as a

reference for analysis. The SUMOylated lysines proteolytic products are tabulated and searched for using both software suits. Precursor ion peptide tolerances were set at 5 ppm, and MS/MS peptide tolerances were set at 1 Dalton. Table 3 outlines the predicted mass of SUMO1 cleave with protease digestion. We utilized every possible protease cut to allow identification of any SUMO modifications with as many possible protease processing.

Table 11: SUMO1 proteolytic peptides from Trypsin, Chymotrypsin, and V8 enzyme

Peptide Amino Acid Sequence	Mass (Da)	Peptide Description
ELGMEEEDVIEVYQEQTGG	2155.27	C90H139N21O38S1
LGMEEEDVIEVYQEQTGG	2026.16	C85H132N20O35S1
GMEEEDVIEVYQEQTGG	1913.00	C79H121N19O34S1
MEEEDVIEVYQEQTGG	1855.94	C77H118N18O33S1
EEEDVIEVYQEQTGG	1724.75	C72H109N17O32
EEDVIEVYQEQTGG	1595.64	C67H102N16O29
EDVIEVYQEQTGG	1466.52	C62H95N15O26
DVIEVYQEQTGG	1337.41	C57H88N14O23
VIEVYQEQTGG	1222.32	C53H83N13O20
IEVYQEQTGG	1123.18	C48H74N12O19
EVYQEQTGG	1010.03	C42H63N11O18
VYQEQTGG	880.91	C37H56N10O15
YQEQTGG	781.78	C32H47N9O14
QEQTGG	618.60	C23H38N8O12
EQTGG	490.53	C18H30N6O10
QTGG	361.40	C13H23N5O7
TGG	233.25	C8H15N3O5
GG	132.13	C4H8N2O3
G	75.07	C2H5NO2

Construction and design of M1 Mutants

The mass spectrometry results provided a total of five lysine residues that were SUMOylated. The figure 20F illustrates that out of the 13 lysine residues that were found on M1, Lysine 21, 187, 230, 242, and 252 were found to be SUMOylated in the in-vitro

reaction. To determine which lysine residues are critical for the progression of the viral infection, M1 mutants were made. The M1 mutants MS identified SUMOylated lysines mutated to arginine, to maintain charge properties. The mutants were constructed through PCR and Gibson ligation, with point mutations at the lysine coding sequences. Tabulated PCR primers shown below for each mutation. The M1 wild type and M1 mutants are ligated onto pET28B-YPet construct with linker, backbone pET28B-YPet-linker construct taken from Malik *et al.*.[23] The PDZ vector for gifted from García-Sastre lab for IAV reconstitution assays.[24]

Table 12: The list of Primers used to make point mutations on IAV-M1 protein.

Amino Acid	Primer
k21r fwd	aggccccctccgagccgagatcgcacagag
k21r rev	tctcggctcggaggggcctgacgggatga
k35r fwd	ctttgcagggcggaacactgatcttgaggttct
k35r rev	cagtgttccgccctgcaaagacatcttcaa
187r fwd	cactacagctcgg gctatggagcaaatggctg
k187r rev	ccatagcccgagctgtagtgctggctaaaa
k230r fwd	tgctggtctgcgaaatgatcttcttgaaaatttgcaggccta
k230r rev	gatcatttcgcagaccagcactggagctag
k242r fwd	ggcctatcagcga cgaatgggggtgcagatg
k242r rev	ccattcgtcgctgataggcctgcaaatttt
k252r fwd	atgggggtgcagatgcaacggttccggtga
k252r rev	tggtggtggtgctcgagtgcggccgctcaccggaaccgttgcatct

In vitro SUMOylation of IAV-M1 Mutants

The in-vitro SUMOylation assay of IAV M1 is an initial screening of lysine sites essential for SUMOylation. We observed 5 lysines in the MS analysis with the SUMO modification, and we include lysine 35 which was initially identified with low probability to have the SUMO modification. The *in vitro* SUMOylation assay screening includes individual mutants as well as all sites in the N-terminal, C-terminal, and all tested lysines. The assay is setup at the same concentration as the optimized conditions, 6xHisCyPet-SUMO1 500 nM, 6xHisYPet-M1 wildtype and mutants 2000 nM, E1 hetro-dimmer AOS1/UBA2 at 100 nM, E2 conjugating enzyme UBC9 200 nM, E3 ligase PIAS1 250 nM, and SUMOylation buffer of 20 mM Tris-HCl (pH 7.5) 50 mM NaCl, 4 mM MgCl, 1 mM DTT. Functional controls are put in place for non-specific interaction, by a negative control reaction without ATP, and to observe a significant boost in FRET a control reaction without E3 ligase. Each reaction was incubated at 37°C for 60 minutes. Following Equation 1 we measure the three wavelengths, EmTotal, FL_D, and FL_A. The measurements are taken on Molecular Devices Spectra M3^{TM}, with "Endpoint" settings, with PMT at constant gain set to "Low".

Reconstitution of Viral Particles M1wt and M1 Mutants

The reconstitution of Influenza A virus is completed by following the virus reconstitution assay constructed by Adolfo Garcia-Sastre lab.[24] The eight influenza virus genes were cloned into the pDZ vector gifted by Sastre lab. The pDZ plasmid design is outlined in the protocol as a bidirectional plasmid with human RNA polymerase I and terminator sequence. Additionally, chicken β-actin promoter and terminator within the

same plasmid. The cytomegalovirus promotor will work in human and Madin-Darby Canine Kidney (MDCK) cells. The chicken β-actin promoter is also added to allow reconstitution within chicken eggs.

The pDZ plasmids are transformed in Top10™ bacteria cells under ampicillin selection. The Biomiga™ and Qiagen™ midi-prep kits were used to obtain transfect DNA for each pDZ vector. Lipofectamine 3000™ was used to co-transfect 1 ug of each pDZ vector (PB2, PB1, PA, HA, NP, NA, M and NS) into HEK293-MDCK cell co culture on 6 well plate the first day. Post 24 hours of transfection, the media is replaced with media without FBS, infectious media, DMEM 0.3% Bovine Albumin (BA), 1% Penicillin/Streptomycin (PS) with 1 ug/ml of TPCK-trypsin. After the media change and 48 hours of incubation and viral particle generation the entire media is collected. The content of collected material is the media without FBS and TPCK/trypsin, cell debris, and viral particles. The collected media is centrifuged at 1000 x g to pellet cell debris and is now used as passage 0 (P0). The collected P0 virus is introduced to fresh 6 well MDCK cells at 80 % confluency. The cells are first rinsed with 1xPBS to remove any residual FBS media and incubated with 250 µL P0 for 1 hour with gentle agitation every 10 minutes. The volume of infecting material can vary by the size of the plate, as the aim is to not allow the cell to dry during the 1-hour incubation. Once the 1-hour infection with P0 is completed the P0 material is removed and replaced with DMEM 0.3% Bovine Albumin (BA), 1% Penicillin/Streptomycin (PS) with 1 ug/ml of TPCK-trypsin, the volume used varies and for a 6 well plate 2 mL post infection media is added. The virus is incubated for 48 hours, and the infection is repeated now with Passage 1 (P1) material

making passage 2. Each passage event exponentially increases the infectious viral particle count. The resulting material is viral particles from P0 to P2 of viral particle in FBS free infection media.

Plaque Assay M1wt and M1 Mutants

The viral plaque assays are implemented to quantify and determine viability of virus generated from the reconstitution assay. The list of reconstituted viruses differs in the M1 construct, each pDZ M1 and M1 mutant is used to generate virus. The P2 virus generated from M1wt and M1 mutants is titrated in increments of 1:1000 to determine over all plaque forming units. The P2 virus is diluted in 1xPBS with 0.5% BSA, and each viral titer is prepared before in 300 µL volumes. Fresh MDCK cells are plated onto a 12 well plate and brought to 80 % confluency. The cells are washed with 1xPBS, and 150 µL of titer is added to the wells and incubated at 37°C, with gentle agitation every 10 minutes, and after 1 hour the titer is removed. Post infection overlay medium, is added onto the wells, and incubated at 37°C for 48 hours. After 48 hours the overlay medium is aspirated and rinsed with ddH_2O. Fixing solution (4% Paraformaldehyde (PFA)) is added to the cells and incubated for 30 minutes and discarded. Post fixing the cells are covered with cell stain, 1% Crystal Violet in 20 % ethanol is added to each well. The cell stain is incubated in each well for 30 minutes and is gently washed off with ddH_2O. Images are collected on a bright field digital camera with a back light illumination.

Cytotoxicity Assay of viral particles with M1wt and M1 Mutants

The viral titer is first analyzed for plaque forming units. Utilizing equation 4 the plaque forming units are calculated. The multiplicity of infection (MOI) is then

calculated using equation 5 B. The cell survival assay used is a type of stain of adhered cells that is measured by absorbance at 540 nm. The Neural Red Cell Cytotoxicity Assay™, by BioVision, kit uses a dye that is taken up by the cell and binds onto liposomes within the cell cytosol. The amount of dye binding to each well will depend on the number of cells within that well. The cells are then washed, and the dye-stained cells are exposed to an acid buffer, that releases the bond dye from the cells. The difference in the absorption from each well directly points to the number of cells survived after the infection. The comparison of M1 wildtype and M1 mutants proved insight into the virus infectivity with M1 mutant.

The cytotoxicity assay is implemented on 96 well plate. The plate is seeded with MDCK cells at 80 % confluency. The cells are introduced to M1 viral particle, M1wt, M1K21R, M1K35R, M1K187R, M1K230R, M1K242R, and M1K252R. The cells infections are stopped at time points 0, 24, 36, 48 hours and measured for cell survival. The reporter for the assay measured the amount of free-floating ATP in the medium, the higher the ATP the more cells that have lysed due to infection.

$$\text{Pfu/ml} = \frac{\text{Average \# of Plaques}}{\text{Dilution} * \text{Volume of diluted virus added}} \qquad \text{Equation 3.3}$$

$$\text{M. O. I.} = \frac{\text{Pfu}}{\text{Number of cells}} \qquad \text{Equation 3.4}$$

Translocation of wildtype M1 and M1 Mutants

Immunostaining of M1 is used to investigate the dependency of SUMOylation of M1 on translocation between cytosol and nucleus. Glass coverslips are coated with L-lysine overnight at 22°C under UV light in a 12 well plate. Post coating HUH7 cells are seeded onto the coverslips and grown till 50 % confluent. The cells are transfected with

M1wt, M1K21R, and M1K242R. Post 24 hours of transfection, the cells are washed with DPBS, and fixed in 4% Paraformaldehyde (PFA) for 15 minutes with rocking. Post fixing the PFA is aspirated, and the cells are washed with DPBS. After fixing the cells are blocked (1XDPBS, 1 % BSA, 0.1 % Triton x-100) for 60 minutes at 22°C with rocking. Post blocking the antibody is diluted 1:100 in blocking buffer and is stained overnight at 4°C with rocking. The cells are rinsed with DPBS for 5 minutes and repeated 3 times. The cells are then incubated for 60 minutes with the secondary anti-mouse 488 Alexa-dye (Invitrogen) in 1XDPBS, 1 % BSA, 0.1 % Triton x-100. The cells are rinsed with DPBS for 5 minutes and repeated 3 times. Post-secondary stain the cell nucleus stain Hoechst 33342, by Thermofisher (H1399) is applied and incubated for 15 minutes at room temperature. Post nuclear stain, the cells are washed 4 times with DPBS with 5-minute incubation at room temperature. The cells were imaged on Olympus upright fluorescence microscope, and images were stacked and analyzed using ImageJ software.

Immunoprecipitation of wild type M1, M1 K21R, and SUMOylated M1 wildtype

The M1 plaque assays and cytotoxicity assay provide insight in decrease in viral activity in mutants M1K242R, and no activity in M1K21R. Confirmation of M1 K21R and wildtype expression in cells is demonstrated through immune precipitation. The wildtype M1 and mutant M1 K21R is transfected in 10 cm plates with 70% confluent HEK293 cells. Lipofectamine 3000 is used to transfect 15 µg of M1 wildtype and M1K21R in separate plates. Additionally, confirmation of M1 SUMOylation in cells is demonstrated by transfection of wildtype M1 with hSUMO1 and E2 conjugating enzyme UBC9. Lipofectamine 3000 is used to transfect HEK293 cells with 5 µg of pDZ-IAV-

M1, 2.5 µg of pEF1-hSUMO1 (Addgene : 27290), and 2.5 µg of pcDNA3.1-YFP-UBC9 (Addgene 2093) in 10 cm plate.

After 48 hours of transfection the 10 cm plates are moved to 4 °C and washed with 1xPBS with PMSF that was preincubated at 4 °C. The cells are then scraped off the plate and transferred into a 1.5 mL microtube. The cells are centrifuged and all 1xPBS is removed from the tube. The cells are resuspended in Cell-Lytic M^{TM} (Sigma Aldrich) with 1xPMSF and setup on a shaker at 4 °C for 30 minutes. The SUMOylated M1 has lysis buffer has 20 mM of NEM (N-ethylmaleimide) SENP inhibitor and 1x PMSF. The deSUMOylation activity of SENP is inhibited by NEM, but NEM cannot pass the cell wall, thus NEM is added to lysis buffer. Post cell lysis the cell debris is separated by centrifugation of tubes at 3000xg for 5 minutes. The cell lysate is transferred to another tube where anti-M1 antibody is added at 1:50 dilution. The primary anti-body is incubated with the cell lysate for 60 minutes at 4°C. The agarose beads with Protein A/G (Santa Cruz Antibody sc-2003) are equilibrated with lysis buffer at 4°C. The beads are then centrifuged at 500 x g for 3 minutes. The equilibrated beads are added to the cell lysate with anti-M1 to capture any M1 protein. The tubes are rocked overnight at 4°C and centrifuged the next morning at 500xg for 2 minutes. The beads now potentially have bound M1 protein to them and is precipitated for both M1 wildtype and M1 mutant. The beads are then washed two times with Cell-Lytic M^{TM} (Sigma Aldrich) and one time with 1xPBS.

Post isolation of agarose beads 40 µL of lamenelli buffer is added. The beads and buffer require denaturing by heat with incubation at 98°C for 3 minutes. Post denaturing

the cell samples are ran on SDS-PAGE. The sample set up included the wild type M1 and M1 mutant, along with blank HEK293 cells. The transfer of protein to nitrile membrane was completed under 100 mV and 450 Amp. The membrane is then incubated with blocking buffer, (1 x TBST, 1% BSA), for 60 minutes. After the 60-minute incubation with blocking buffer is used to dilute the primary antiM1 1:500 dilution. The primary antibody blotting is completed over night at 4°C with gentle rocking. The next day the membrane is washed in 1 X DPBS, and secondary HRP anti-Mouse (Thermofisher G-21040) antibody is added to the cells and gently rocked for 60 minutes. The stain is aspirated from the membrane and rinsed 3 times with TBST. SuperSignal™ West Pico PLUS (Thermofisher 34579) is used as substrate for HRP bioluminescence. The UVP camera and software suit is used to image the membrane.

3.8 References

1. Baudin, F., Petit, I., Weissenhorn, W. & Ruigrok, R. W. H. *In vitro* Dissection of the Membrane and RNP Binding Activities of Influenza Virus M1 Protein. Virology **281**, 102–108 (2001).

2. Noda, T. *et al.*. Importance of the 1+7 configuration of ribonucleoprotein complexes for influenza A virus genome packaging. Nat Commun **9**, 54 (2018).

3. Dou, D., Revol, R., Östbye, H., Wang, H. & Daniels, R. Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. Front. Immunol. **0**, (2018).

4. Wu, C.-Y., Jeng, K.-S. & Lai, M. M.-C. The SUMOylation of Matrix Protein M1 Modulates the Assembly and Morphogenesis of Influenza A Virus. Journal of Virology **85**, 6618–6628 (2011).

5. Enami, M. & Enami, K. Influenza virus hemagglutinin and neuraminidase glycoproteins stimulate the membrane association of the matrix protein. J Virol **70**, 6653–6657 (1996).

6. Sakai, T., Nishimura, S. I., Naito, T. & Saito, M. Influenza A virus hemagglutinin and neuraminidase act as novel motile machinery. Sci Rep **7**, 45043 (2017).

7. Li, Y. *et al.*. Neddylation of M1 negatively regulates the replication of influenza A virus. Journal of General Virology **101**, 1242–1250.

8. Wang, S. *et al.*. Tyrosine 132 Phosphorylation of Influenza A Virus M1 Protein Is Crucial for Virus Replication by Controlling the Nuclear Import of M1. J Virol **87**, 6182–6191 (2013).

9. Su, W.-C. *et al.*. Pooled RNAi screen identifies ubiquitin ligase Itch as crucial for influenza A virus release from the endosome during virus entry. PNAS **110**, 17516–17521 (2013).

10. Mahesutihan, M. *et al.*. CypA Regulates AIP4-Mediated M1 Ubiquitination of Influenza A Virus. Virol Sin **33**, 440–448 (2018).

11. Flotho, A. & Melchior, F. Sumoylation: A Regulatory Protein Modification in Health and Disease. Annual Review of Biochemistry **82**, 357–385 (2013).

12. Santos, A. *et al.*. SUMOylation Affects the Interferon Blocking Activity of the Influenza A Nonstructural Protein NS1 without Affecting Its Stability or Cellular Localization. Journal of Virology **87**, 5602–5620 (2013).

13. Han, Q. *et al.*. Sumoylation of Influenza A Virus Nucleoprotein Is Essential for Intracellular Trafficking and Virus Growth. Journal of Virology **88**, 9379–9390 (2014).

14. Way, G. *et al.*. A novel SUMOylation site in the influenza a virus NS1 protein identified with a highly sensitive FRET assay. Journal of Biotechnology **323**, 121–127 (2020).

15. Domingues, P. *et al.*. Global Reprogramming of Host SUMOylation during Influenza Virus Infection. Cell Rep **13**, 1467–1480 (2015).

16. Pal, S., Santos, A., Rosas, J. M., Ortiz-Guzman, J. & Rosas-Acosta, G. Influenza A virus interacts extensively with the cellular SUMOylation system during infection. Virus Research **158**, 12–27 (2011).

17. Furuse, Y., Suzuki, A., Kamigaki, T. & Oshitani, H. Evolution of the M gene of the influenza A virus in different host species: large-scale sequence analysis. Virol J **6**, 67 (2009).

18. Aguilar-Martinez, E. *et al.*. Screen for multi-SUMO–binding proteins reveals a multi-SIM–binding mechanism for recruitment of the transcriptional regulator ZMYM2 to chromatin. PNAS **112**, E4854–E4863 (2015).

19. Garcia-Dominguez, M. & Reyes, J. C. SUMO association with repressor complexes, emerging routes for transcriptional control. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms **1789**, 451–459 (2009).

20. Evolutionary optimization of fluorescent proteins for intracellular FRET | Nature Biotechnology. https://www.nature.com/articles/nbt1066?platform=hootsuite.

21. Song, Y., Madahar, V. & Liao, J. Development of FRET assay into quantitative and high-throughput screening technology platforms for protein-protein interactions. Ann Biomed Eng **39**, 1224–1234 (2011).

22. Liu, Y., Song, Y., Madahar, V. & Liao, J. Quantitative Förster resonance energy transfer analysis for kinetic determinations of SUMO-specific protease. Analytical Biochemistry **422**, 14–21 (2012).

23. Malik-Chaudhry, H. K., Saavedra, A. & Liao, J. A linker strategy for trans-FRET assay to determine activation intermediate of NEDDylation cascade. Biotechnology and Bioengineering **111**, 1288–1295 (2014).

24. Martínez-Sobrido, L. & García-Sastre, A. Generation of Recombinant Influenza Virus from Plasmid DNA. J Vis Exp 2057 (2010) doi:10.3791/2057.

Chapter 4: *In vitro* qFRET Assay for the SUMOylation of SARS-CoV-2 Nucleocapsid protein with MS analysis of Modified Lysines

4.1 SARS-CoV-2 Life Cycle

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a global pandemic responsible for the upper respiratory disease Coronavirus Disease 2019 (COVID-19). The disease in 2020 developed the delta variant which has been reported to have 1,260 times higher viral load in comparison to the original strain. The SARS-CoV-2B.1.617 lineage identified originally in India, containing the functional mutations to the spike protein, NTD (T19R, G142D, Δ156, Δ157, R158G), RBD (L452R, T478K), S2 region (D950N) and P681R.[1] The delta variant has a reduced neutralizing immune response when compared to the previous variants after a complete first and second dose Pfizer and AstraZeneca vaccine. Furthermore, investigation on vaccine dose response show no neutralizing response to a single dose of Pfizer or AstraZeneca vaccine to the delta variant. The rapid development of variants with onset of functional mutations highlights to the global science community to purse all available avenues of neutralizing this viral infection and the viral proteome interaction with host factors are a potential avenue left relatively undiscovered.[2]

SARS-CoV-2 Life Cycle Overview

The SARS-CoV-2 viral particle is composed of starting from the core the positive sense single strand RNA (~30 kb) packed around nucleocapsid (N) proteins. The compacted RNA nucleocapsid complex is enveloped in a lipid membrane with embedded membrane proteins (M) and envelope protein (E).[3] The glycoprotein spike protein (S)

exposed outside on the viral particle surface is found to have high affinity for human angiotensin-converting enzyme 2 (hACE2).[4]

The binding event initiates the mechanism of viral entry into the cell. Recent studies have demonstrated that the post translation modification (PTM) glycosylation plays an essential role in S and hACE2 interaction.[5] Once bound the human cell surface protein transmembrane protease serine 2 (TMPRSS2) cleaves the spike protein which results in the cleaved spike protein binding to the host cell surface.[6] The S protein unravels and merges the two lipid membranes together and a endocytosis mechanism delivers the particle into the cell.[7]

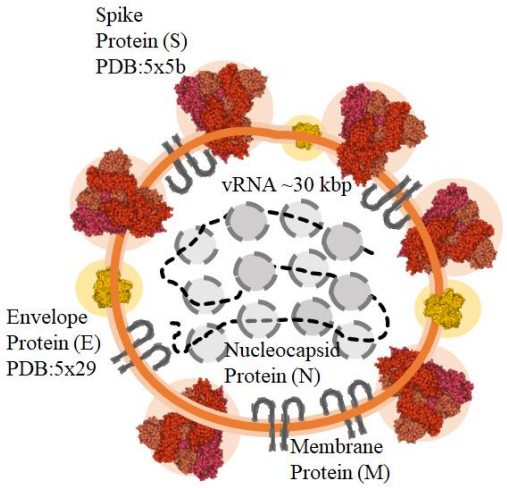


Figure 25: Viral particle organization. Spike protein depicted by the crystal structure (5x5b) found on the membrane, along with the envelope protein (PDB:5x29), and membrane protein (M). Within the particle are nucleocapsid proteins that compact the vRNA, shown in a beads on a string formation.

The viral particle releases all N protein and the viral positive sense RNA into the host cytosol. The vRNA is then unpacked and translated by the host ribosome. The translation results in an amino acid chain which is processed into proteins by two viral proteases NSP3 and NSP5. The result is an array of SARS-CoV-2 Non-structural proteins

(NSP) 1 - 16(Figure 26A). NSPs hijack host functions to inhibit host RNA expression and inhibit host anti-viral response elements from being translated. NSP1 protein has been reported to interact with host factors to inhibit anti-viral response and remodel the cellular environment for virus pathogenesis.[8,9] The spike protein infiltrates the host cell membrane and recruits adjacent host cells expressing hACE2 and fuses the cells together creating super structures called syncytia.[10] Meanwhile the viral proteins NSP 3 and 5 attack the endoplasmic reticulum (ER) and eventually snip off pieces of the ER creating double membrane vesicles (DMVs).[11] These DMVs house vRNA till the viral protein NSP3 has been observed to release that vRNA out into the cytoplasm.[11] The released RNA is packaged into a lipid envelope constructed from lysosome.[12] This newly constructed viral particle within the lipid envelope moves towards the cell surface. The S protein is said to be processed by a host protease Furin which activates the S protein, but is not essential for release of infectious particles.[13] This complex series of events have been characterized by groups of scientists across the globe and provide insight into the viral pathogenesis.

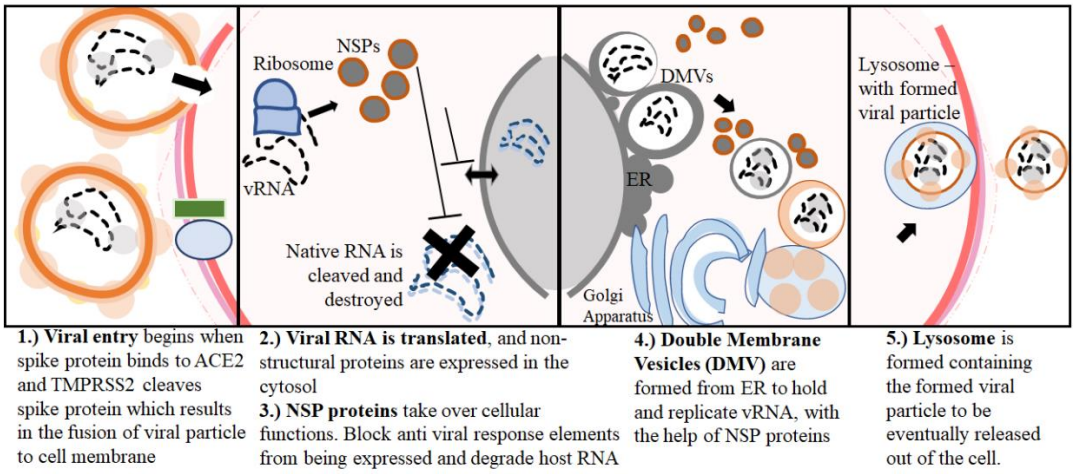


Figure 26: SARS-CoV-2 pathogenesis, viral entry by S and ACE interaction, translation of vRNA, and expression of NSP proteins. NSP proteins inhibit mRNA translocation from the nucleus to cytosol to inhibit anti-viral response and cleave host RNA in the cytosol. The vRNA is also observed to be packaged in double membrane vesicles (DMV) made by NSP5 protease from the endoplasmic reticulum. The vRNA is replicated and packaged again within the DMV and released into the cytosol. The Golgi Apparatus aid in expression of viral membrane proteins which encapsulate the vRNA and a form a viral particle within a lysosome. The lysosome releases the viral particle out of the cell.

SARS-CoV-2 Nucleocapsid Protein Structure and Functional Packaging of vRNA

SARS-CoV-2 viral pathogenesis highlights the modulation and interaction of host factors that enable the virus to infect and spread throughout the host. The focus of this study is on the N protein and the intersection of its role in pathogenesis with host factors. The N protein is the most abundant proteins of the viral proteome observed in patient samples of live virus.[14] The major known function of N proteins is to form viral RNA-protein complex (vRNP) with compacted vRNA genome and interact with M protein within the viral particle.[15] The innate organization of the N protein within the viral particle is reported to compact into a varying degrees of condensate that aid in circumventing the host immune response to vRNA and enable packaging of vRNA into a viral particle.[16]

N protein is responsible for binding and conforming vRNA and making a structural bond with M protein within the viral particle. Cryo-electron tomography (CryoET) studies have elucidated the complex overall structure of N protein bound to vRNA and N protein bound to M protein. Recent investigations concluded the oligomerized N protein compacts RNA into a structure that forms a phase-separated condensate within the viral particle that binds to M protein.[3,16–18] Highlights of CryoET and sub tomogram averaging (STA) work has investigated the organization of vRNPs within the viral particle. They observed the average diameter of a viral particle to be 80 nm and contained 30-35 vRNPs. The group predicted each vRNP is 15 nm in diameter and holds 12 copies of N protein wrap and coated in RNA, creating a condensate that encapsulates the genomic vRNA and interacts with M protein.[16,19] The sub-domains that enable protein-protein complexes of the N protein oligomers form dimers and tetramers, that enable the complex formation of vRNP and ultimately enclosure within a viron.

Structure of N Protein

The SARS-CoV-1 N protein and SARS-CoV-2 N proteins have homology across their sequences, and demonstrated to have similar domains and function.[14] The N protein identified domains N-terminal domain (NTD from 1-50 aa) is followed by the RNA binding domain (RBD from 51-174 aa), the dimerization domain (247-364 aa), and the C-terminal domain (CTD from 365-419aa).[20] The CTD reported to interact with M protein within the viral envelope and supports the vRNA super structure.[16] Between the RBD and the dimerization domain is a liker region (174-245 aa), this region is a serine and arginine rich region on the protein that is reported to be phosphorylated.[21]

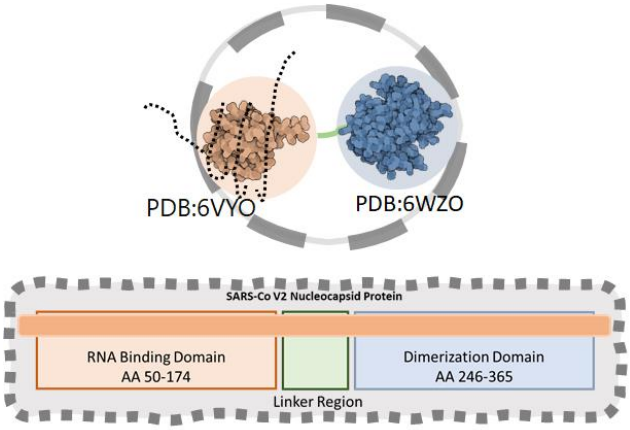


Figure 27:Nucleocapsid protein illustration of n and c terminal domains. The protein n-terminal interacts with the vRNA, the linker region is reported to be phosphorylated, and the c-terminal domain reported to be the interacting domain for oligomerization.

4.2 SARS-CoV-2 Nucleocapsid Protein and Host Proteome Interaction

Investigations of the intersection of N protein interaction network with host proteome have expanded the functional attributes of N protein beyond the vRNP complex. The anti-viral response of host cells is reported to be inhibited by N protein interaction with host immune response protein, signal transducer and activator of transcription (STAT1 & 2). N protein mediate inhibition of interferon antiviral response by sequestering activated STAT proteins within the cytosol and disrupting the INF signaling from progressing.[22] Furthermore, N protein is reported to depend on PTMs from host proteome for its functional RNA binding properties.

Recent PTM studies on the SARS-CoV-2 N protein have found potential phosphorylation on serine 197 and threonine 205.[21] The study co-expressed N protein with the catalytic subunit of the enzyme protein kinase A (PKA) in bacteria, and isolated phosphorylated N protein. The study reported a total of 20 sites modified by phosphorylation using mass spectrometry. The study ruled out 18 sites due to predicted

binding inefficiencies of phosphorylating enzymes and homology studies across variants. The study observed SARS-CoV-2 N protein modulation of RNA binding with mutations at 197 and 205 in the S/R rich region. The phosphorylation modification is inherent to the proteins function and demonstrates the dependence of virus progression on host PTMs.[16,17] Furthermore, an investigation in ubiquitination, screened in-cell modified N protein for ubiquitination and observed lysine 169, 374, and 388 to be ubiquitin modification. No follow up investigations were done in the study as identification of PTM location was the only objective.[23]

In the following study we investigate the PTM SUMOylation on SARS-CoV-2 N protein, a small ubiquitin like modification (SUMO) that has been reported with the variant SARS-CoV-1 N protein to modulate its function in cell.[24] The SUMOylation of SARS-CoV-1 N protein at lysine 62 was predicted by homology and SUMOylation motif predictions. SUMO mutant of SARS-CoV-1 N protein follow up studies demonstrated some modulation in oligomerization with SUMO mutant N proteins, and some modulation of nuclear translocation. A significant impact that was alluded to in the 2005 study by Li *et al.*, is the modulation of oligomerization of N protein with the SUMO modification. The evidence provided in the study was a western blot that observed the decrease of crosslinked N proteins. However, it has been demonstrated that N proteins can form super structures of dimers and tetramer which ultimately form condensate.[16,18] The oligomerization of N protein is described to be a critical factor in viral genome packaging. The study outlined here, is in part, to investigate the non-covalent interaction of N protein oligomerization using an *in vitro* quantitative Förster's resonance energy

transfer (qFRET) assay and determine if SUMOylation has any impact on the non-covalent oligomerization of N proteins.

SUMO proteins are small ubiquitin-related modifier (SUMO) that is a vital post translation modification utilized in our proteome. The SUMO modification of a target protein increases the target protein affinity to other cellular processes and extends the target proteins non-covalent interaction network range to proteins with affinity to SUMO proteins.[25,26] In this work we demonstrate an in-vitro qFRET assay for the SUMOylation of N protein coupled with mass spectrometry to identify the site of SUMO modified lysine. The modified lysine sites are then mutated to arginine individually making N protein SUMO mutants. The N protein mutants are evaluated by *in vitro* SUMOylation assay, and modulation of N protein cellular localization.

The investigation in modulation of dissociation constant of SUMOylated N protein by applying a qFRET based kinetic assay to determine K_D of SUMO modified N proteins. The previous investigation on SUMOylation of SARS-CoV-1 N protein used western blot to infer the decrease in binding and formation of dimers or trimers. The advantage of applying N-protein SUMO mutants to a qFRET assay to determine K_D over western blot analysis is that it provides resolution on the non-covalent interaction between N-protein - N-protein affinity. The *in vitro* assessment of the affinity of N protein to N protein with SUMO mutants provides insight of the condensate formation of oligomerized N protein with modulation of SUMO modification.

4.3 SUMOylation Enzymatic Cascade of N Protein

The reconstituted complete SUMOylation cascade, with the E3 ligase PIAS1 which enhances the SUMOylation of target SARS-CoV-2 N protein. The fusion proteins CyPet-SUMO1 and Ypet-N are constructed with the fluorescent protein at the n-terminal (Figure 29A). The enzymatic reaction is setup in-vitro with E1 activating enzyme hetro-dimer, Activating Enzyme Subunit 1 (AOS1), and Ubiquitin Activating Enzyme 2 (UBA2). The SUMO specific E2 Conjugating enzyme, Ubiquitin Carrier 9 (UBC9), and coupled with E3 ligase, Protein Inhibitor of activated STAT 1 (PIAS1). The enzymatic cascade is initiated by the addition of ATP to activate the E1 hetero-dimer, CyPet-SUMO1 Gly98 makes a thiol intermediate with Cys173 on UBA2. The E1-SUMO complex interacts with UBC9 and transfers CyPet-SUMO1 onto the catalytic Cys93, and finally the E3 ligase specific to the target protein mediates the aminolysis reaction for covalent attachment of SUMO1 onto lysine of N protein. The conjugation of the fusion protein Cypet-SUMO1 onto lysine residue on YPet-N is a covalent conjugation and brings the FRET pairs within the 10 nm range of transfer for an observable increase in Em_{FRET} signal.

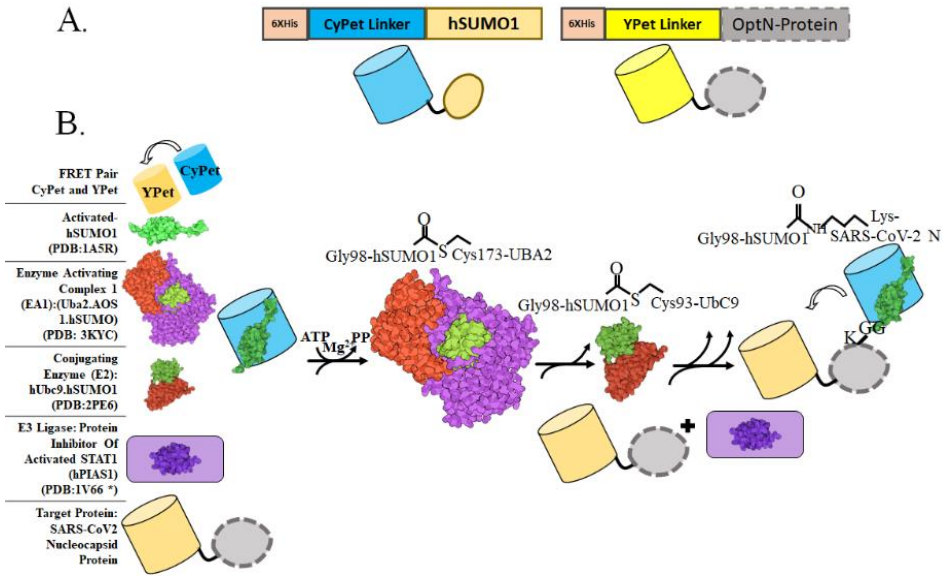


Figure 28:A. Diagram of the fusion protein CyPet-SUMO1 and YPet-N. B.) In-vitro SUMOylation assay with qFRET as a reporter illustrated here. The fusion protein CyPet-hSUMO is first bound to E1 activating enzyme, for the intermediate E1-Cypet-SUMO1 thioester bond at cys-173 to glycine 98 on SUMO1. The SUMO is then transferred to the catalytic cystiene93 of E2 conjugating enzyme. The E3 ligase and target protein are said to non-covalently interact with the E2-SUMO1 complex. The CyPet-SUMO1 is then shuttled to a lysiene on the target protein, to be covalently bound by a isopeptide bond. The covalent binding of SUMO onto N protein closes the proximiy of CyPet-Ypet to less then 10 nm allowing FRET to occur.

4.4 qFRET Assay for the In-Vitro SUMOylation of N protein

The reconstitution of SUMOylation reaction with qFRET as a reporter is a robust and rapid method for identifying SUMOylation events, characterizing enzymatic activity, and identifying SUMOylation sites.[27,28] This method includes the E3 ligase PIAS1 for its enhanced SUMOylation activity and is coupled with mass spectrometry to provide insight into identifying multiple SUMOylation sites.[29–31] The qFRET assay utilizes a FRET optimized donor fluorescent protein tag, CyPet, on the SUMO1 protein and FRET optimized acceptor fluorescent protein tag, YPet, on the N-protein.[32] The CyPet-YPet fluorescent proteins experience the non-radiative FRET phenomenon, when within 10 nm

distance between them which is applicable for observing SUMO1 attachment. The FRET phenomenon has been described and applied in various studies of protein-protein interaction, and its response to distance between donor acceptor, r, is demonstrated to directly impact FRET efficiency by r^6.[33–35] We have developed a fluorescence based method in determining covalent attachment of Cypet-SUMO1 to YPet-N protein. The method demonstrated here applies a "three cube FRET" fluorescence reporter that extracts the emission of FRET signal, Em_{FRET}, from the raw fluorescent signal, EmTotal, at the FRET wavelength. The extraction applies three different fluorescent measurements (Table 1), to extract the Em_{FRET} response from a FRET reaction. The method filters out cross channel signal of the unbound donor or acceptor from the FRET wavelength. The relationship applied in equation 1, determines the contribution of cross talk form both acceptor and donor by applying ratiometric constants alpha (α Equation 2) and beta (β Equation 3), to subtract the crosstalk signal from the EmTotal. The details on the development of the qFRET method can be found on a previous study on development of the qFRET signal by Song *et al.*.[28] The method has been applied previously to determine kinetic values of protein-protein interactions such as dissociation constant K_D, enzymatic constants k_{cat}/K_m, and applied to assess SUMO modification of viral proteins.[28,36,37]

Table 13: Em_{FRET} relationship to raw Em_{FRET} signal, EmTotal, the donor cross talk contribution, and acceptor cross talk contribution.

	Excitation (λ)	Emission (λ)
Em_{Total}	414 nm	530 nm
Fl_D	414 nm	475 nm
FL_A	475 nm	530 nm

$$Em_{FRET} = (Em_{Total}) - ((FL_D * \alpha) + (FL_A * \beta)) \quad \text{Equation 4.1}$$

$$\alpha = \frac{\text{Donor em}isson\ at\ 530\ nm\ with\ excitation\ at\ 414\ nm}{\text{Donor em}isson\ at\ 475\ nm\ with\ excitation\ at\ 414\ nm} \quad \text{Equation 4.2}$$

$$\beta = \frac{\text{Acceptor em}isson\ at\ 530\ nm\ with\ excitation\ at\ 414\ nm}{Acceptor\ \text{em}isson\ at\ 530\ nm\ with\ excitation\ at\ 475\ nm} \quad \text{Equation 4.3}$$

4.5 Results

In vitro qFRET Assay for SUMOylation of N Protein

The observed Em_{FRET} response for the SUMOylation of SARS-CoV-2 N protein demonstrates a significant difference in the magnitude of Em_{FRET} with the addition of ATP (Figure 29). A t-test analysis was completed on the triplicate measurements, compared with and without ATP, as well as with and without E3 ligase. There was also a significant Em_{FRET} signal without E3 ligase, which demonstrates a robust SUMOylation event. A one-way ANOVA was used to analyze the significance between the control group without ATP (-ATP), with ATP and no E3 ligase, and with ATP and E3 ligase.

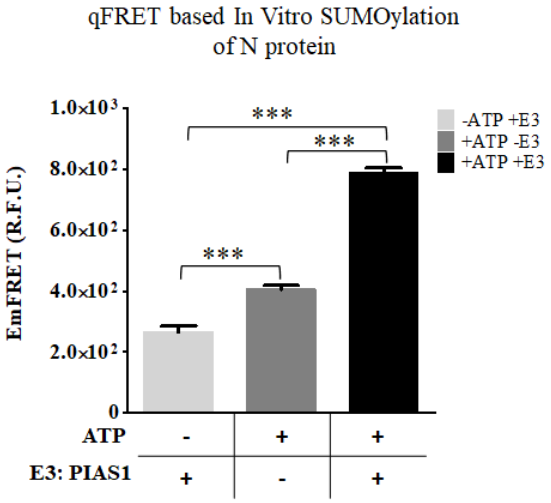


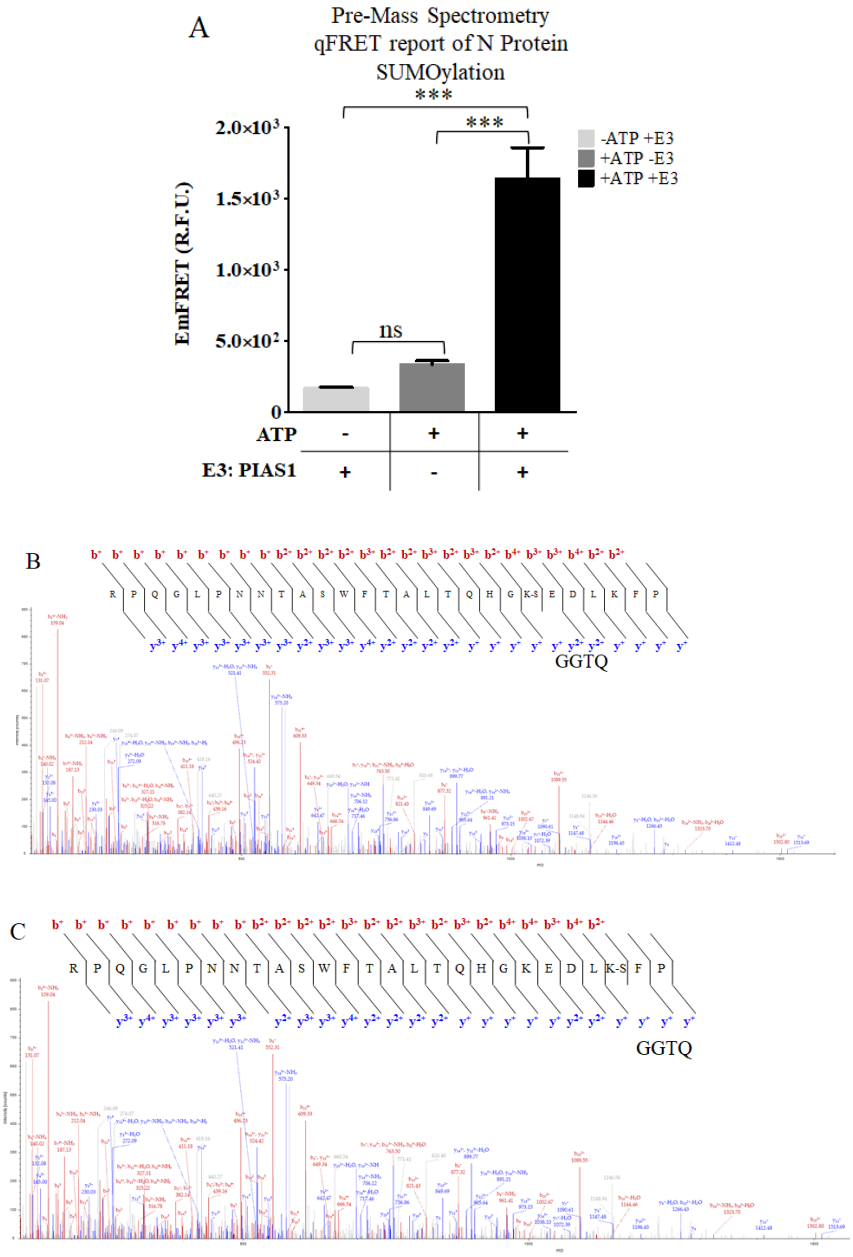
Figure 29: *In vitro* SUMOylation of N protein, without ATP, without E3 ligase PIAS1, and with both ATP and E3 ligase(A). p<0.0001***

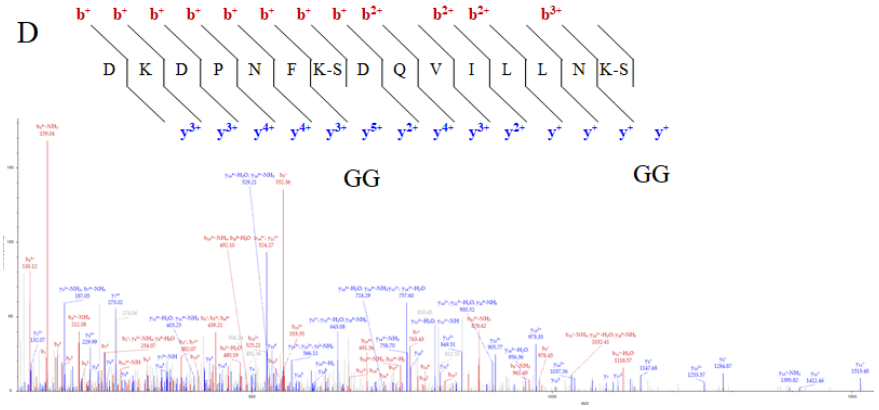
Mass Spectrometry Analysis to Determine SUMO Modified Lysine on N Protein

The identified modified lysines are illustrated in figure 31D along with the other 25 lysines found unmodified by SUMO1. The coverage of the N protein was 95 %. The spectrum of the identified peptide with lysine 61 modification was a large section of N protein from position 41 to 67, which had lysines 61 and 65 contained within it. This cut pattern matches for n-terminal arginine-41, and the cleave at arginine-68 after the proline-67. The GluC protease cuts the carboxyl ends of glutamic acids and can also cut the n terminal of arginine (R) and c terminal of aspartic acid (D). The SUMO1 peptide "GGTQ" followed by a glutamic acid at position 93 is cleaved and matches the expected GluC cut. Figure 31B has a similar peptide however with lysine-65 holding the SUMO1 modification. The GluC cut positions on the peptide are the same as the previous peptide, N protein position 41 - 67. Figure 31C, the N protein peptide from positions 341 - 355, with SUMO1 GG peptide. This peptide match the GluC cut pattern at aspartic acid

position 341 – 357. However, some discrepancies can be noted in figure 31c, the c terminal of the peptide should have an isoleucine after the modified lysine. This could be an aberration in the MS analysis or an artifact of GluC non-specific cut. We still included these two sites in the study as the other 25 lysines had no detected modification.

The secondary analysis of each lysine site can be done through evaluation of each site matching the SUMO consensus motif. Based on the SUMO consensus motif, of a hydrophobic residue (Ψ), the modified lysine (K), any amino acid, and either an aspartic acid or a glutamic acid. The Ψ-K-x-D/E motif has been applied by numerous groups to ascertain SUMOylation site. Using two only servers, GPS-SUMO and JASSA were used to determine if these sites matched the SUMO consensus.[38,39] Both servers only pointed to lysine position 338 as the highest probability, and lysine 61, 65, 347, and 355 all were low probability. This result did not match the SARS-CoV-1 N protein SUMOylation site, also does not match the *in vitro* SUMOylation sites, and was not included in this study.


A
Pre-Mass Spectrometry
qFRET report of N Protein
SUMOylation
***
***
ns
EmFRET (R.F.U.)
2.0×10³
1.5×10³
1.0×10³
5.0×10²
0
-ATP +E3
+ATP -E3
+ATP +E3
ATP - + +
E3: PIAS1 + - +
B
R P Q G L P N N T A S W F T A L T Q H G K-S E D L K F P
GGTQ
C
R P Q G L P N N T A S W F T A L T Q H G K E D L K-S F P
GGTQ


D
D K D P N F K-S D Q V I L L N K-S
GG
GG

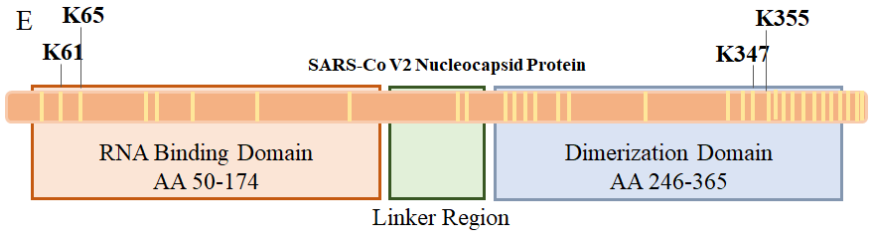

E
K65
K61
K355
K347
SARS-Co V2 Nucleocapsid Protein
RNA Binding Domain
AA 50-174
Dimerization Domain
AA 246-365
Linker Region

Figure 30: The *in vitro* MS sample was measured for qFRET signal before processing for MS, with and with ATP and E3 (30A). MS spectrum of peptide containing modified lysine 61 (*30B*), with a GluC cut on SUMO1, GGTQ. Lysine 65 (*30C*) with the same GluC cut on the SUMO1 peptide. The lysine 347 and 355 are both found in the same peptide, with SUMO1 peptide GG identified mass (3*0D*). The illustration of the location of the four discovered lysines, along with a total of 31 lysines on the protein shown as yellow lines (*30E*). Spectrums were generated by Thermofisher Proteome Discoverer™

qFRET Assay for In-Vitro SUMOylation N Protein Mutants

The results of the *in vitro* SUMOylation of N protein with qFRET as a reporter is shown below (Figure 32). We observe significant signal in all proteins with E3 ligase and ATP, this indicates the activity of PIAS1 ligase, to SUMOylate N protein is high. The significant differences can be seen in reactions without E3 ligase, we see a significant difference without ATP and without E3 ligase. This indicates some level of modulation in SUMO1 activity. In comparison the wildtype N protein observed significant differences when ATP was added, in both -E3 and +E3. The 61 and 65 mutants showed a drop in the signal in -E3, and this pattern followed in the double mutant without E3 ligase. The one-way ANOVA analysis with Tukey Test, results in significant differences when ATP is added across all E3+ reactions. However, it should be noted that modulation in minus E3 (-E3) reactions was detected.

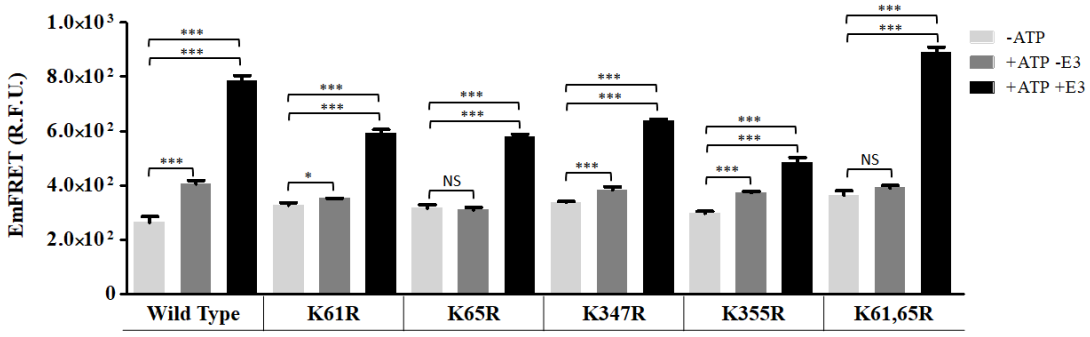


Figure 31: *In vitro* SUMOylation with qFRET as the reporter. Comparison of no ATP (-ATP), with no E3 ligase PIAS1 (-E3), and a complete reaction with ATP and E3 ligase PIAS1 (+ ATP+ E3). The reactions were all done under the same conditions, and the measurements were all taken on the same instrument, Molecular Devices SpectraMax3^{TM}. One-Way ANOVA was done on the data sets of -ATP/-E3/+ATP+E3, the -ATP was the control group. Tukey test was used as the post-HOC analysis, and displayed p values are $p<0.0001$***, $p<0.05$*, and no significant difference (ns), n=3

Cellular Translocation Nucleocapsid Protein

The YPet signal on the N protein demonstrated a condensate like formation within the cytosol. We observe fluorescent granules forming within the cell which is inherent to the N protein function. The first image at the top are the nuclear stains using, Hoechst, that provide the location of the nucleus, Figure 32. The second image is in the YPet channel, where we observe the Ypet bound N protein. The formed bright spots are observed to be condensate or granules of N protein. We observe modulation of N protein shuttling in and out of the nucleus which can be seen as the individual bright spots within the nucleus. The first set of images are of the wild type N protein, and we observe granules/condensate within the nucleus and the cytosol. The N protein K61R we observe granules within the cell and the cytosol. The N protein K65R showed minimal spotting in the nucleus but presented signal in the cytosol. The c terminal mutations K347R and K355R presented small condensates in both cytosol and nucleus.

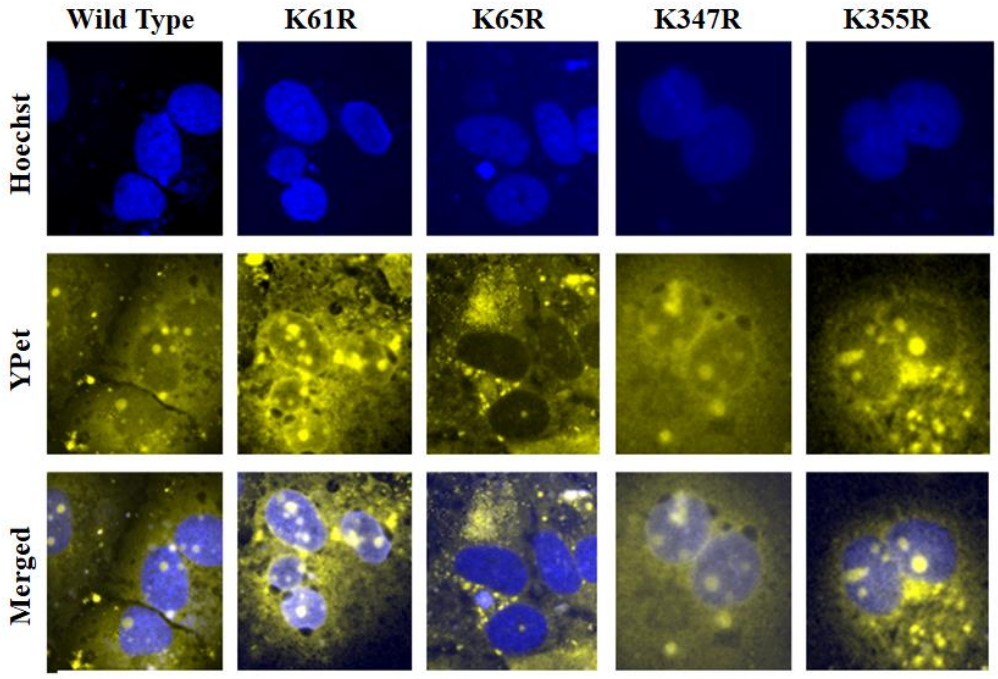


Figure 32: Fluorescent Imaging and Immunostaining of Nucleocapsid protein taken on a fluorescent microscope, Olympus BX43. The nuclear stain Hoechst was imaged at 488 nm and YPet tag was imaged at 533 nm. Images were processed on ImageJ™.

qFRET K_D of SUMOylated N Protein

The results of the wildtype N protein affinity assays are listed in table 16, with SUMO1 modified and unmodified N protein. The calculated Em_{FRET} from the titration of acceptor fusion protein YPet-Nwt is plotted in points (circle/orange) for unmodified, and (diamond/green) for SUMO1 modified. We observe a K_D of Wt N protein when SUMO1 modified to be 0.46 µM, standard error of 0.34, and 1.46 µM standard error of 0.09 when not modified. The Em_{FRET} response from the titration of total acceptor protein and the fit profile plotted in figure 33. The results of qFRET K_D assay on the N protein mutants are listed in table 4. The results of the fit including the K_D values are listed, for both modified and unmodified mutant proteins. The Em_{FRET} response

across 0 to 2.5 µM of protein demonstrates modulation of affinity with and without SUMOylation. The affinity of the un-modified N proteins ranged from 0.65 to 2.45 µM, and the range for modified was 0.30 to 1.70 µM. The lysine 65 and 355 mutants had variation in their data, the reported R^2 values are lower than the rest of the data set.

Table 14: Non-Linear Regression Fit of Equation 4, to determine K_D of SUMO1 modified and SUMO1 unmodified wild type Nucleocapsid protein

	Un-Modified Wt N Protein	**Modified Wt N Protein**
K_D µM	1.46	0.46
Standard Error µM	0.34	0.09
95 % Confidence Interval µM	0.76 to 2.15	0.28 to 0.68
R^2	0.96	0.97

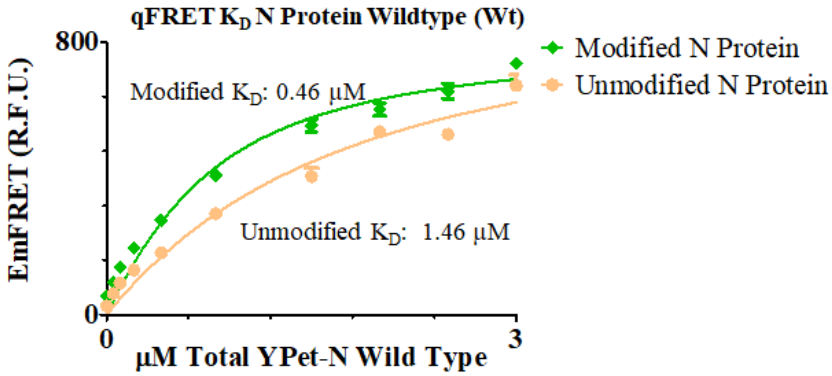


Figure 33: qFRET K_D of modified Nucleocapsid Protein, plot on GraphPadPrism5™. The determined fit of SUMO modified(Diamond/Green), and unmodified (Circle/Orange) Em_{FRET} to total acceptor fusion protein, YPet-Nc wt.

Table 15: qFRET K_D results of Nucleocapsid protein SUMO mutants, K61R, K65R, K347R, and K355R. The fit results are tabulated here, for each mutant with and without SUMO modification.

	K61R -ATP	K61R +ATP	K65R -ATP	K65R +ATP	K347R -ATP	K347R +ATP	K355R -ATP	K355R +ATP
K_D (µM)	0.92	0.64	1.38	1.84	1.98	1.70	0.51	0.17
Standard Error (µM)	0.14	0.11	0.44	0.58	0.61	0.37	0.20	0.07
95 % Confidence Interval (µM)	0.62 to 1.21	0.41 to 0.87	0.46 to 2.30	0.64 to 3.03	0.71 to 3.21	0.99 to 3.97	0.08 to 0.90	0.01 to 0.32
R^2	0.99	0.98	0.95	0.91	0.97	0.97	0.90	0.90
n	27	27	27	27	27	27	27	27

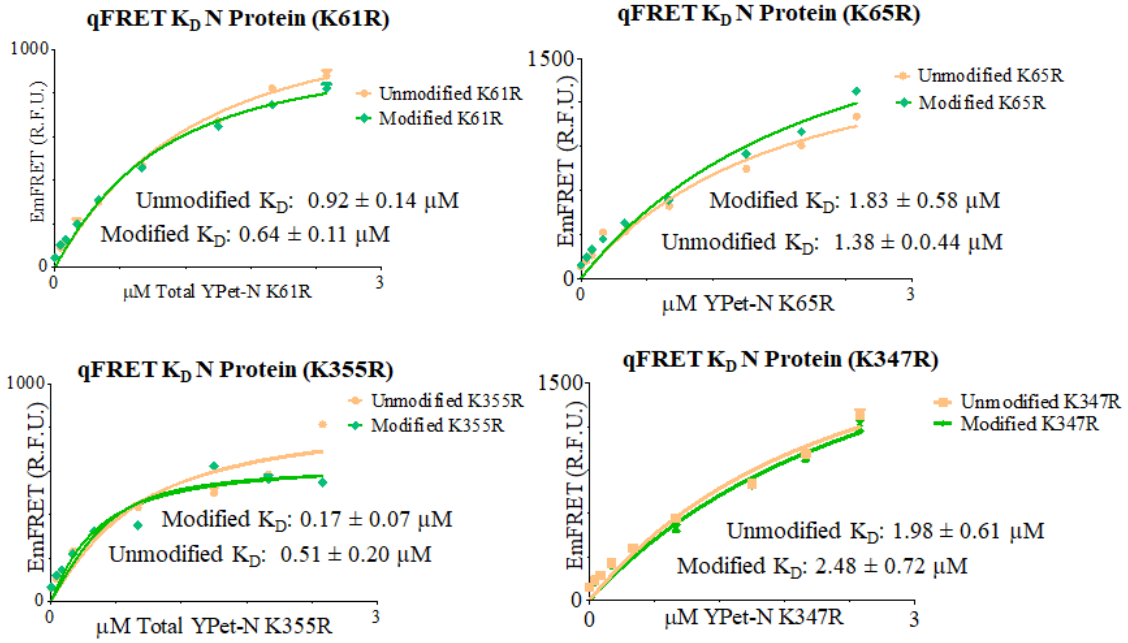


Figure 34: Fit plots of each qFRET K_D assay, with Em_{FRET} vs. total acceptor concentration. The SUMO modified mutant is shown in (Diamond/Green) and the unmodified (Circle/Orange). The plots and the fits were generated on GraphpadPrism5™.

4.6 Discussion

The *in vitro* SUMOylation assay of SARS-CoV-2 Nucleocapsid protein demonstrated here provided four SUMOylation sites, K61, 65, 347, and 355 determined *in vitro*. The qFRET based SUMOylation assay, used alternatively to immunoblot, is used in the evaluation of covalent modifications of proteins. The SARS-CoV-2 proteome is relatively new to the scientific community, and access to versatile techniques that can evaluate protein modifications and evaluate protein properties is difficult. Demonstrated here is a versatile method to evaluate the covalent modification and non-covalent interactions using the same platform. An advantage of the *in vitro* evaluation of SUMOylation modifications is it also provides certainty in the observations. Due to the numerous in-cell lysine modifications the probability of missing or false negative classification of a lysine modification is high. Furthermore, the yield of a modified protein from an in-cell pull-down assay can be challenging and have low yields. Thus, researchers look to over express the SUMOylation mechanism, however within a cellular environment, over SUMOylation can bring about unwanted consequences to the cellular proteome. The *in vitro* qFRET assay used here provided a fluorescent reporter for SUMOylation and was directly used in the identification of modified lysine residues. The coverage of the protein identified in MS was up to 95 %, the other 5 % is assumed to be degraded during the digestion and sample preparation. The high certainty and the overall coverage of the protein in MS analysis provide confidence in the identified SUMOylation sites.

The lysine resides identified were evaluated using qFRET and observed the reaction to be very robust, and sensitive to lysine mutations without the E3 present. The one-way ANOVA analysis of the reactions demonstrated a significant difference in the minus E3 reactions from mutant N proteins 61 and 65. The in-cell analysis of cellular translocation we observed formation of bright spots that are granules of N protein. The N protein resides in the cytosol, however, reports of it shuttling in and out of the nucleus are consistent with our findings. We observe the wildtype N protein in bright spots in the cytosol, and in the nucleus. These findings are consistent with reports of N protein cellular translocation. The modulation of translocation was observed with lysine 65 mutation. The findings suggest a decrease in nuclear translocation, without SUMOylation.

The qFRET K_D assessment of the N protein provides insight into the N proteins non-covalent interaction. The N protein is found to crosslink at high concentrations in the cell, but forms tetramers and dimers. The overall organization of the vRNP complex formation by the N protein requires non-covalent interaction of the various forms of N protein. Demonstrated here the modulation of affinity of N protein with the SUMO modification. Though the difference in affinity is not large, it is observed, and found to be dependent on SUMOylation *in vitro*. The qFRET platform provides the rapid assessment of a heterologous protein and is applied to evaluate *in vitro* covalent modification of a viral protein. The next step for this analysis is to determine the impact of the SUMOylation of N protein on the viral pathogenesis.

4.7 Methods and Materials

Expression and Purification of SUMOylation Enzymes and SARS-CoV-2 N protein

The in-vitro qFRET SUMOylation reaction is completed with the E1, E2, and E3 enzymes in the SUMOylation cascade. The E1 activation enzyme complex, UBA2 and AOS1, E2 conjugating enzyme UBC9, and E3 ligase PIAS1 were all cloned into pET28B vector for expression in BL21(DE3) cells. The FRET pairs CyPet and YPet are N-terminal tagged to SUMO1 and the substrates respectively and cloned into pET28B for expression in BL21(DE3). Each BL21(DE3) cell line with individual proteins were inoculated at 1:100 and grown to 0.4 OD at 600 nm at 37 °C, and then induced at overnight at 22°C with 0.25 mM IPTG. The cells were lysed, lysis buffer (20 mM Tris-HCl (pH 7.5), 0.5 M NaCl, 5 mM Imidazole), by sonication and centrifuged at 35,000 x g. The soluble fraction was purified by 6XHis tag to NiNTA beads affinity chromatography through a gravity column. The bound proteins were washed with, buffer 1 (20 mM Tris-HCl (pH 7.5), 0.3 M NaCl), buffer 2 (20 mM Tris-HCl (pH 7.5), 1.5 M NaCl, and 0.5% Triton X-100), and buffer 3 (20 mM Tris-HCl pH 7.5, 0.5 M NaCl, and 10 mM Imidazole). The proteins eluted using the following buffer, (20 mM Tris-HCl, 300 mM NaCl, and 450 mM Imidazole) and dialyzed in 20 mM Tris-HCl (pH 7.5), 50 mM NaCl, and 1 mM DTT.

In-vitro SUMOylation Assay Setup

The in-vitro SUMOylation assay is completed with 6xHisCyPet-SUMO1 500 nM, 6xHisYPet-N protein wildtype 2000 nM, E1 hetro-dimer AOS1/UBA2 at 100 nM, E2 conjugating enzyme UBC9 200 nM, E3 ligase PIAS1 250 nM, and in SUMOylation

buffer (20 mM Tris-HCl (pH 7.5) 50 mM NaCl, 4 mM MgCl, 1 mM DTT). Functional controls are put in place for observing non-specific interaction, by a negative control reaction without 2 mM adenosine triphosphate (ATP). Each reaction was incubated at 37°C for 60 minutes and measured in a 384 well microplate (Grenier 384 M6811). The FRET wavelength, EmTotal, are 414 nm excitation and 530 nm emission, Fl_D, 414 nm excitation and 475 nm emission, and Fl_A, 475 nm excitation and 530 nm emission. The quantitative Em_{FRET} parameters, α of 0.34 ± 0.003, and β 0.003 ± 0.001 variables are determined using the formulation outlined in previous work from Yang *et al.*. Equation 1 provides the calculation of Em_{FRET} that quantifies the FRET increase based on the SUMOylation of substrate. The specificity of SUMO protein to the SUMOylation target has the potential to yield a false positive FRET response. Thus, functional controls of reactions without ATP are implemented in parallel to observe differences between ATP and no ATP. Samples from each reaction of -ATP/-E3/+ATP/+E3 was also immune blotted with anti SUMO1.

Mass spectrometry analysis to determine SUMO modified Lysine on N protein

The in-vitro SUMOylation reactions the substate, YPet-SARS-CoV-2 Nucleocapsid protein is added at 3000 nM, and CyPet tagged SUMO1 protein are added at 1000 nM. Activating Enzyme Complex 1 (E1) is at 100 nM, and Conjugating Enzyme 2 (E2) at 100 nM, E3 ligase at 500 nM in SUMOylation buffer (20 mM Tris-HCl (pH 7.5) 50 mM NaCl, 4 mM MgCl, 1 mM DTT) and 2 mM ATP. The reactions were completed at 37 ° Celsius for 4 hours. The in-solution proteolytic digestions were performed with Pierce™ Glu-C Protease. Samples were digested at 1:100 ratio for

sample to enzyme ratio and ran overnight (16 hours) at 37 ° Celsius. Each completed digestion was acidified to a final concentration of 0.1% v/v TFA, and speed vacuumed to dry product, and then reconstituted to 0.1% v/v TFA readied for MS loading.

LTQ Orbitrap Xl Loading and Run

Samples consisted of approximately 1000 nM of in-solution digested product from each proteolytic enzyme digestion Liquid chromatography was performed on a Thermo nLC1200 in single-pump trapping mode with a Thermo PepMap RSLC C18 EASY-spray column (2 µm, 100 Å, 75 µm x 25 cm) and a Pepmap C18 trap column (3 µm, 100 Å, 75 µm x 20 mm). Solvents used were A: water with 0.1% formic acid and B: 80% acetonitrile with 0.1% formic acid. Samples were separated at 300 nL/min with a 250-minute gradient starting at 3% B increasing to 30% B from 1 to 231 minutes, then to 85% B at 241 minutes, holding for 10 minutes.

Mass spectrometry data was acquired on a Thermo Orbitrap Fusion in data-dependent mode. A full scan was conducted using 60k resolution in the Orbitrap in positive mode. Precursors for MS^2 were filtered by monoisotopic peak determination for peptides, intensity threshold 5.0e3, charge state 2-7, and 60 second dynamic exclusion after 1 analysis with a mass tolerance of 10 ppm. Higher-energy C-trap dissociation (HCD) spectra were collected in ion trap MS^2 at 35% energy and isolation window 1.6 m/z.

Bioinformatic Analysis of MS Data

The LTQ-orbitrap XL (.raw) raw data was analyzed on Thermofisher Proteome Analyzer™. The complete amino acid sequence of each protein was provided as a

reference for analysis. The SUMOylated lysines proteolytic products are tabulated and searched for using both software suits. Precursor ion peptide tolerances were set at 5 ppm, and MS/MS peptide tolerances were set at 1 Dalton.

Table 16: SUMO1 proteolytic peptides for identification of lysine modification.

Peptide Amino Acid Sequence	Mass (Da)	Peptide Description
ELGMEEEDVIEVYQEQTGG	2155.27	C90H139N21O38S1
LGMEEEDVIEVYQEQTGG	2026.16	C85H132N20O35S1
GMEEEDVIEVYQEQTGG	1913.00	C79H121N19O34S1
MEEEDVIEVYQEQTGG	1855.94	C77H118N18O33S1
EEEDVIEVYQEQTGG	1724.75	C72H109N17O32
EEDVIEVYQEQTGG	1595.64	C67H102N16O29
EDVIEVYQEQTGG	1466.52	C62H95N15O26
DVIEVYQEQTGG	1337.41	C57H88N14O23
VIEVYQEQTGG	1222.32	C53H83N13O20
IEVYQEQTGG	1123.18	C48H74N12O19
EVYQEQTGG	1010.03	C42H63N11O18
VYQEQTGG	880.91	C37H56N10O15
YQEQTGG	781.78	C32H47N9O14
QEQTGG	618.60	C23H38N8O12
EQTGG	490.53	C18H30N6O10
QTGG	361.40	C13H23N5O7
TGG	233.25	C8H15N3O5
GG	132.13	C4H8N2O3
G	75.07	C2H5NO2

Construction and Design of N Protein Lysine to Arginine Mutants

The mass spectrometry results provided a total of four lysine residues that were SUMOylated. There are a total of 31 lysine residues on SARS-CoV-2 N protein, Lysine 61, 65, 347, and 355 were found to be SUMOylated in the in-vitro reaction. The mutants DAN templates were constructed through PCR, with point mutations at the lysine to arginine coding sequences. Final Gibson reaction of bacterial expression pET28B vector SALI and NOTI and for mammalian expression pCDNA3.1-FLAGtag-Nprotein-YPet were created. Tabulated PCR primers shown below for pET28B and for pcDNA3.1.

Table 17: Primers listed for constructing N Protein Mutants in *E. coli*

pET28B Primers	
K61RFor	ccagcatggcagagaagacctgaaattt
K61RRev	caggtcttctctgccatgctgggtcag
K65RFor	gaagacctgagatttccgcgcggccag
K65RRev	ctggccgcgcggaaatctcaggtcttc
K347RFor	gatccgaattttcgagatcaggtgatt
K347RRev	aatcacctgatctcgaaaattcggatc
K355RFor	attctgctgaatagacatattgacgcg
K355RRev	cgcgtcaatatgtctattcagcagaat

Table 18: List of primers for mutations on N protein in HUH7 cells

pCDNA3.1 Primers	
pcD_Ncwt_61For	ctcactcaacatggcagggaagacctt
pcD_Ncwt_61Rev	aaggtcttccctgccatgttgagtgag
pcD_Ncwt_65For	gaagaccttagattccctcgaggacaa
pcD_Ncwt_65Rev	ttgtcctcgagggaatctaaggtcttc
pcD_Ncwt_347For	aaagatccaaatttcagagatcaagtcatt
pcD_Ncwt_347Rev	aatgacttgatctctgaaatttggatcttt
pcD_Ncwt_355For	gtcattttgctgaataggcatattgacgcatac
pcD_Ncwt_355Rev	gtatgcgtcaatatgcctattcagcaaaatgac

In-Vitro SUMOylation with qFRET Reporter for N Protein Mutants

The in-vitro SUMOylation assay of SARS-CoV-2 N protein mutants is an initial screening to determine impact of lysine sites on SUMOylation. The assay is setup at the same concentration as the optimized conditions, 6xHisCyPet-SUMO1 500 nM, 6xHisYPet-N Protein wildtype and mutants 2000 nM, E1 hetro-dimer AOS1/UBA2 at 100 nM, E2 conjugating enzyme UBC9 200 nM, E3 ligase PIAS1 250 nM, and SUMOylation buffer of 20 mM Tris-HCl (pH 7.5) 50 mM NaCl, 4 mM MgCl, 1 mM DTT. Functional controls are put in place for non-specific interaction, by a negative control reaction without ATP, and to observe a significant boost in FRET a control reaction without E3 ligase. Each reaction was incubated at 37°C for 60 minutes. Following Equation 1 we measure the three wavelengths, EmTotal, FL_D, and FL_A. The

measurements are taken on Molecular Devices Spectra M3^{TM}, with “Endpoint” settings, with PMT at constant gain set to “Low”. The post analysis was completed on GraphpadPrism7^{TM}, One-way ANOVA with post HOC Tukey Test was done with minus ATP as the control group.

Cellular Translocation of N protein

Immunostaining of N protein is used to investigate the dependency of SUMOylation of N protein on translocation between cytosol and nucleus. Glass coverslips are coated with L-lysine overnight at 22°C under UV light in a 12 well plate. Post coating HUH7 cells are seeded onto the coverslips and grown till 50 % confluent. The cells are transfected with M1wt, M1K21R, and M1K242R. Post 24 hours of transfection, the cells are washed with DPBS, and fixed in 4% Paraformaldehyde (PFA) for 15 minutes with rocking. Post fixing the PFA is aspirated, and the cells are washed with DPBS. After fixing the cells are blocked (1XDPBS, 1 % BSA, 0.1 % Triton x-100) for 60 minutes at 22°C with rocking. Post blocking the antibody is diluted 1:100 in blocking buffer and is stained overnight at 4°C with rocking. The cells are rinsed with DPBS for 5 minutes and repeated 3 times. The cells are then incubated for 60 minutes with the secondary anti-mouse 488 Alexa-dye (Invitrogen) in 1XDPBS, 1 % BSA, 0.1 % Triton x-100. The cells are rinsed with DPBS for 5 minutes and repeated 3 times. Post-secondary stain the cell nucleus stain Hoechst 33342, by Thermofisher (H1399) is applied and incubated for 15 minutes. Post nuclear stain, the cells are washed 4 times with DPBS with 5-minute incubation. The cells were imaged on Olympus BX43, and images were stacked and analyzed using ImageJ software.

qFRET K_D of SUMOylated N Protein

The evaluation of N protein oligomerization by *in vitro* qFRET based K_D affinity assay. The individual N protein wild type and mutants are first cloned into the FRET fusion genes. Each N protein must be tagged with the donor or acceptor pair fluorescent proteins for the implementation of this assay. Thus, five pairs of interacting N proteins are made into pET-28B, the wild type pair, mutant 61, 65, 347, and 355. Each pair of proteins are *in vitro* SUMOylated following optimized concentrations of enzyme and SUMO1. 6xHisCyPet-SUMO1 6000 nM, 6xHisYPet-N protein 6000 nM, E1 hetro-dimer AOS1/UBA2 at 100 nM, E2 conjugating enzyme UBC9 200 nM, E3 ligase PIAS1 250 nM, and in SUMOylation buffer (20 mM Tris-HCl (pH 7.5) 50 mM NaCl, 4 mM MgCl, 1 mM DTT). Functional controls are put in place for observing non-specific interaction, by a negative control reaction without 2 mM ATP. The SUMOylation reaction is incubated at 37°C for 1 hour, along with the negative control of no ATP.

The reactions include the N protein bound to CyPet (donor) and the separate reaction of bound YPet (acceptor). A total of four reactions are setup for each pair of N proteins, that includes the negative control of no ATP. After the incubation the reaction is directly setup for affinity measurements. The qFRET based K_D method holds the donor fusion protein constant at 500 nM, and the acceptor is titrated up to 2500 nM from 0 nM. The series of titrations are individual measurements that provide the Em_{FRET} response to the concentration of acceptor fusion protein, YPet-N protein. For each titration of acceptor, Em_{FRET} value is calculated. Following Equation 1 we measure the three wavelengths, EmTotal, FL_D, and FL_A for each reaction. The measurements are taken on

Molecular Devices Spectra M3^{TM}, with "Endpoint" settings, with PMT at constant gain set to "Low". We obtain a data set of Em_{FRET} vs Total acceptor concentration, and with and without ATP.

Equation 4.4

$$Em_{FRET} = EmFRET_{Max_{Max}} * \left(\frac{[Acceptor]_{Total} - [Donor]_{Total} - K_D + \sqrt{([Donor]_{Total} + K_D - [Acceptor]_{Total})^2 + 4 * K_D * [Acceptor]_{Total}}}{[Donor]_{Total} + K_D - [Acceptor]_{Total} + \sqrt{([Donor]_{Total} - [Acceptor]_{Total} + K_D)^2 + 4 * K_D * [Acceptor]_{Total}}} \right)$$

This data set of Em_{FRET} vs total donor fusion protein concentration, is fitted to the nonlinear derived relationship, equation 4, described in the work by Sang *et al.*.[28] For the analysis of this data, GraphpadPrism5^{TM} is used to fit the nonlinear equation to the data set collected. The constraints for the non-linear regression fit are set to donor concentration at a constant of 500 nM, and the constraint K_D and $Em_{FRETMax}$ cannot be zero.

4.8 References

1. Planas, D. *et al.*. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature (2021) doi:10.1038/s41586-021-03777-9.

2. Reardon, S. How the Delta variant achieves its ultrafast spread. Nature (2021) doi:10.1038/d41586-021-01986-w.

3. Cubuk, J. *et al.*. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. Nat Commun **12**, 1936 (2021).

4. The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells | bioRxiv. https://www.biorxiv.org/content/10.1101/2020.01.31.929042v1.

5. Identification of 22 N-glycosites on spike glycoprotein of SARS-CoV-2 and accessible surface glycopeptide motifs: Implications for vaccination and antibody therapeutics | Glycobiology | Oxford Academic. https://academic.oup.com/glycob/article/31/1/69/5851811.

6. Hoffmann, M. *et al.*. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell **181**, 271-280.e8 (2020).

7. Wang, H. *et al.*. SARS coronavirus entry into host cells through a novel clathrin- and caveolae-independent endocytic pathway. Cell Res **18**, 290–301 (2008).

8. Finkel, Y. *et al.*. SARS-CoV-2 uses a multipronged strategy to impede host protein synthesis. Nature **594**, 240–245 (2021).

9. Jauregui, A. R., Savalia, D., Lowry, V. K., Farrell, C. M. & Wathelet, M. G. Identification of Residues of SARS-CoV nsp1 That Differentially Affect Inhibition of Gene Expression and Antiviral Signaling. PLOS ONE **8**, e62416 (2013).

10. Zhang, Z. *et al.*. SARS-CoV-2 spike protein dictates syncytium-mediated lymphocyte elimination. Cell Death Differ 1–13 (2021) doi:10.1038/s41418-021-00782-3.

11. Mohan, J. & Wollert, T. Membrane remodeling by SARS-CoV-2 – double-enveloped viral replication. Fac Rev **10**, 17 (2021).

12. Ghosh, S. *et al.*. β-Coronaviruses Use Lysosomes for Egress Instead of the Biosynthetic Secretory Pathway. Cell **183**, 1520-1535.e14 (2020).

13. Papa, G. *et al.*. Furin cleavage of SARS-CoV-2 Spike promotes but is not essential for infection and cell-cell fusion. PLOS Pathogens **17**, e1009246 (2021).

14. Supekar, N. T. *et al.*. Variable posttranslational modifications of severe acute respiratory syndrome coronavirus 2 nucleocapsid protein. Glycobiology (2021) doi:10.1093/glycob/cwab044.

15. He, R. *et al.*. Characterization of protein–protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. Virus Research **105**, 121–125 (2004).

16. Lu, S. *et al.*. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. Nat Commun **12**, 502 (2021).

17. Carlson, C. R. *et al.*. Phosphoregulation of Phase Separation by the SARS-CoV-2 N Protein Suggests a Biophysical Basis for its Dual Functions. Molecular Cell **80**, 1092-1103.e4 (2020).

18. Savastano, A., Ibáñez de Opakua, A., Rankovic, M. & Zweckstetter, M. Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. Nat Commun **11**, 6041 (2020).

19. Yao, H. *et al.*. Molecular Architecture of the SARS-CoV-2 Virus. Cell **183**, 730-738.e13 (2020).

20. Ye, Q., West, A. M. V., Silletti, S. & Corbett, K. D. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. Protein Science **29**, 1890–1901 (2020).

21. Tung, H. Y. L. & Limtung, P. Mutations in the phosphorylation sites of SARS-CoV-2 encoded nucleocapsid protein and structure model of sequestration by protein 14-3-3. Biochem Biophys Res Commun **532**, 134–138 (2020).

22. Mu, J. *et al.*. SARS-CoV-2 N protein antagonizes type I interferon signaling by suppressing phosphorylation and nuclear translocation of STAT1 and STAT2. Cell Discov **6**, 1–4 (2020).

23. Sun, Z. *et al.*. Mass Spectrometry Analysis of SARS-CoV-2 Nucleocapsid Protein Reveals Camouflaging Glycans and Unique Post-Translational Modifications. Infectious Microbes & Diseases (2021) doi:10.1097/IM9.0000000000000071.

24. Han, Q. *et al.*. Sumoylation of Influenza A Virus Nucleoprotein Is Essential for Intracellular Trafficking and Virus Growth. Journal of Virology **88**, 9379–9390 (2014).

25. Merrill, J. C. *et al.*. A Role for Non-Covalent SUMO Interaction Motifs in Pc2/CBX4 E3 Activity. PLOS ONE **5**, e8794 (2010).

26. Arriagada, G., Muntean, L. N. & Goff, S. P. SUMO-Interacting Motifs of Human TRIM5α are Important for Antiviral Activity. PLoS Pathog **7**, e1002019 (2011).

27. Liao, J., Song, Y. & Liu, Y. A new trend to determine biochemical parameters by quantitative FRET assays. Acta Pharmacol Sin **36**, 1408–1415 (2015).

28. Song, Y., Madahar, V. & Liao, J. Development of FRET assay into quantitative and high-throughput screening technology platforms for protein-protein interactions. Ann Biomed Eng **39**, 1224–1234 (2011).

29. Kaci, A. *et al.*. The E3 SUMO ligase PIASγ is a novel interaction partner regulating the activity of diabetes associated hepatocyte nuclear factor-1α. Scientific Reports **8**, (2018).

30. Galanty, Y. *et al.*. Mammalian SUMO E3-ligases PIAS1 and PIAS4 promote responses to DNA double-strand breaks. Nature **462**, 935–939 (2009).

31. Chanda, A. *et al.*. Identification of the SUMO E3 ligase PIAS1 as a potential survival biomarker in breast cancer. PLOS ONE **12**, e0177639 (2017).

32. Evolutionary optimization of fluorescent proteins for intracellular FRET | Nature Biotechnology. https://www.nature.com/articles/nbt1066?platform=hootsuite.

33. Algar, W. R., Hildebrandt, N., Vogel, S. S. & Medintz, I. L. FRET as a biomolecular research tool — understanding its potential while avoiding pitfalls. Nat Methods **16**, 815–829 (2019).

34. Chen, T. *et al.*. Application of Förster Resonance Energy Transfer (FRET) technique to elucidate intracellular and In Vivo biofate of nanomedicines. Advanced Drug Delivery Reviews **143**, 177–205 (2019).

35. Liu, Y., Song, Y., Jiang, L. & Liao, J. Quantitative analysis of FRET assay in biology—new developments in protein interaction affinity and protease kinetics determinations in the SUMOylation cascade. Front. Biol. **7**, 57–64 (2012).

36. Liu, Y., Song, Y., Madahar, V. & Liao, J. Quantitative Förster resonance energy transfer analysis for kinetic determinations of SUMO-specific protease. Analytical Biochemistry **422**, 14–21 (2012).

37. Way, G. *et al.*. A novel SUMOylation site in the influenza a virus NS1 protein identified with a highly sensitive FRET assay. Journal of Biotechnology **323**, 121–127 (2020).

38. GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs | Nucleic Acids Research | Oxford Academic. https://academic.oup.com/nar/article/42/W1/W325/2436801.

39. Beauclair, G., Bridier-Nahmias, A., Zagury, J.-F., Saïb, A. & Zamborlini, A. JASSA: a comprehensive tool for prediction of SUMOylation sites and SIMs. Bioinformatics **31**, 3483–3491 (2015).

Chapter 5: Development of *In vitro* High throughput Screening (HTS) for Inhibitors of Protein-Protein Interaction With qFRET as a Reporter

The high throughput screening approach to drug discovery requires a versatile reporter of a protein's function or interaction, and the applicability of that reporter to be applied in a high throughput fashion. The qFRET technique provides a robust solution in applying a FRET reporter of protein function or interaction and its scalability in microscale reactions that can be detected on a bench top plate reader. This chapter outlines the Z' qualification of *in vitro* qFRET based HTS for two assays, one is the inhibitor of protease activity, and second is the more traditional protein-protein interaction inhibitor.

5.1 *In vitro* qFRET Based HTS Assay for Inhibitors of Autophagy Related Protein 4A

The protease of interest is, Autophagy related protein 4 (Atg4A), part of an intercellular process that is conserved across eukaryotic cells and contributes to sustaining cellular homeostasis. Autophagy process degrades cellular material, such as misfolded proteins, and the process extends to larger cellular organelles such as damaged endoplasmic reticulum, mitochondria, and damaged nucleus.[1,2] Autophagy plays an important role in cell response to the environment, notably within oxidative stressed or starvation, and response to a pathogen. Recent studies have reported cases where resistance to chemotherapy accompanies increased levels of autophagy, additionally, inhibition of autophagy has been shown to reverse drug resistant cancer cell lines.[3,4] The

immense role that this process plays in cell survival and cell death makes the modulation of this process a potential target for drug discovery.

The construction of the autophagosome requires interaction between two protein cascade systems, Autophagy related genes 12-15 and Autophagy related gene 8 (Atg8)-phosphatidylethanolamine(PE) conjugation system. This study focuses on the Atg8 conjugation system, the conjugation protein cascade is initiated with the precursor Atg8 protein is processed by protease Autophagy related gene 4 (Atg4) to expose the glycine residue at the C-terminus, which is essential for its E1 enzyme, Autophagy related gene 7 (Atg7), to adenylate Atg8. The adenylation reaction is ATP dependent, and the intermediate consists of a thioester bond between Atg8 and Atg7, E1-Atg8 intermediate. The E1-Atg8 intermediate then forms thioester intermediate with the E2 enzyme to finally be transferred onto the amino group of the PE, completing the Atg8-PE moiety. The Atg8-PE mediates the membrane formation of the phagosome and provides a way to select for content to be degraded. Atg4 is employed by the Atg8-PE conjugation system to act as the processor of Atg8-PE moiety by deconjugation of Atg8-PE where it cleaves the moiety, referred to as delipidation. Delipidation of Atg8-PE moiety by Atg4 is for recycling Atg8 to restart the protein cascade.[1,4]

Atg4 protease role is essential for the autophagy pathway in its function as the protease of Atg8, this critical role can be leveraged as a potential drug target for the modulation of the autophagy pathway. The Atg4 protease has four homologs in humans, Atg4A, Atg4B, Atg4C, and Atg4D. The homologs differ in their affinity to the six homologs of Atg8, all six are split between the sub family's light chain 3 (LC3) and its

homologs, and the γ-aminobutyric acid receptor associated protein (GABARAP) family of proteins. The protease Atg4A has very low affinity to LC3 and its homologs but has very high affinity to GABARAP family of proteins, specifically to Golgi-associated ATPase enhancer of 16 kDa (GATE-16).

5.2 QFRET Assay Design for Observing Atg4A Activity

The first example is the in-vitro HTS assay for inhibitors of Atg4A with the GABARAP family of proteins, Golgi-associated ATPase enhancer of 16 kDa (GATE-16) based on Förster Resonance Energy Transfer (FRET). The FRET efficiency between our donor and acceptor is dependent on inverse of the sixth power of the distance, r, between the donor and acceptor. The protein construct is designed to place the substrate between the acceptor and donor molecules, which have been optimized for efficient FRET.[5] The activity of the protease can be correlated to the FRET signal decrease, with the acceptor molecule being cleaved from the substrate and separating the donor and acceptor proteins. The substrate designed is based on previous work for the study of SUMO protease activity.[6,7]

The design for the substrate consists of a donor fluorescent protein CyPet on the N-terminal of the substrate, GATE16, and the acceptor fluorescent protein, YPet, downstream of the cleavage site on the C-terminal of the substrate (Figure 1). The Em_{FRET} from the measurement after we subtract the contribution from cross channel fluorescence from the donor and the acceptor, we can establish Em_{FRET} of intact substrate, un-cleaved substrate. The activity of the protease is inversely proportional to the Em_{FRET}, as the activity of the protease increases the Em_{FRET} decrease. (Figure 1)

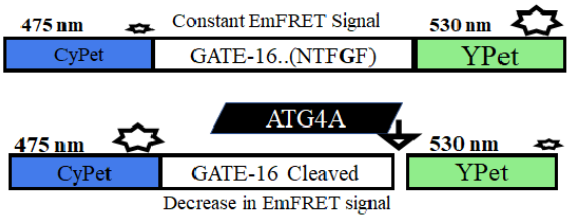


Figure 35:Illustration of GATE16 sensor for Atg4A activity, decrease in Em_{FRET} signal with Atg4 cleave of GATE16 (A).

5.3 Motivation for the HTS development for PD1-PDL1 Interaction Inhibitors

The second qFRET based HTS development is to screen for inhibitors of Programmed Death 1 (PD1) interaction with Programmed Death Ligand 1 (PDL1). PD1 interaction with PDL1 biological impact is described in detail in chapter 2. Within the past decade six total FDA approved anti-PD1/PDL therapies are currently used in treatment. One example of a novel treatment is Keytruda™ (Pembrolizumab) which is an immune-checkpoint inhibitor for treatment of cancer. The trends in clinical trials with Pembrolizumab have evolved to combination therapies, and this is mainly due to lack in efficacy and the discovery of immune-related adverse events.[8,9] The recent studies of clinical data from the combination therapies of FDA approved anti-PD1/PDL1 such as Pembrolizumab, Nivolumab, and Durvalumab is the occurrence of varying levels of pulmonary toxicity.[10] Furthermore, clinical trials recruitment involving pembrolizumab alone or in combination with other treatments are still on the rise, however only a fraction makes it to completion. Shown in the accumulated data below of clinical trials involving pembrolizumab, the number of active and or completed trials is decreasing since 2015. A novel inhibitor of PD1-PDL1 would expand the current portfolio of anti-PD1 or PDL1 therapies and provide new dimension for treatment options.

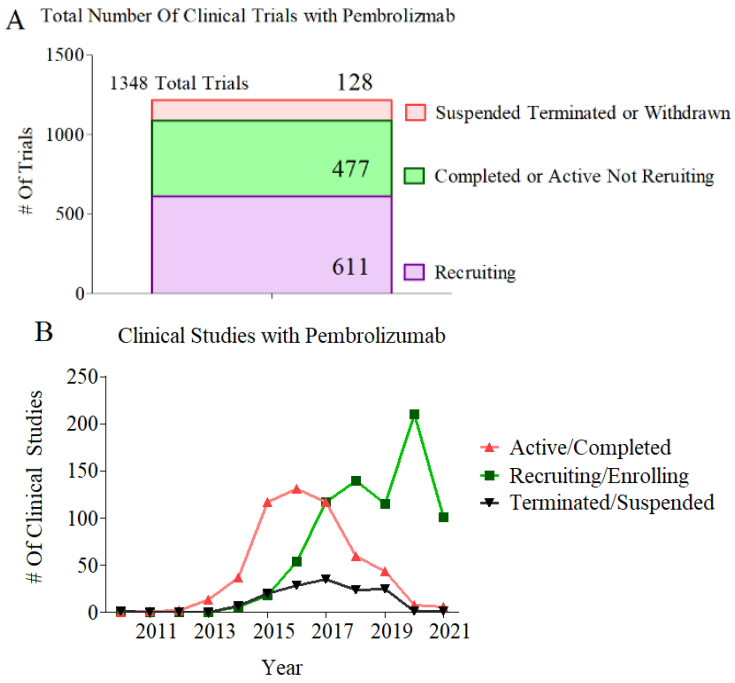


Figure 36: Total posted Pembrolizumab clinal trials on clinicaltrials.gov database. (A) Data of Pembrolizumab clinical trials were first posted, from 2010 – 2021. Active or completed clinical trials, in red triangle, of those that are recruiting or enrolling in green square, and of those that are terminated or withdrawn in black triangle pointed down.

5.4 CyPet-PD1 and YPet-PDL1 Fusion Protein Design

The qFRET based HTS for inhibitors of PD1-PDL1interaction demonstrated here is optimized for a microplate reaction, that can be measured on a 384 well plate. The overall design applies the FRET pair CyPet-YPet to the full length PD1-PDL1 proteins. The donor fluorescent protein, CyPet, is recombinantly fused to the N terminal of PD1 and acceptor fluorescent protein YPet is recombinantly fused to the N terminal of PDL1, illustrated in Figure 3. Based on the FRET principles the interaction of the fusion protein CyPet-PD1 and YPet-PDL1 will result in an increase in Em_{FRET} signal. The inhibition of the PD1-PDL1 interaction results in decrease of Em_{FRET} signal. This design is different from the Atg4 activity assay with GATE16 fusion protein, as here we observe a decrease

in signal, while in Atg4A activity assay we observe a constant Em_{FRET} signal with inhibitor activity.

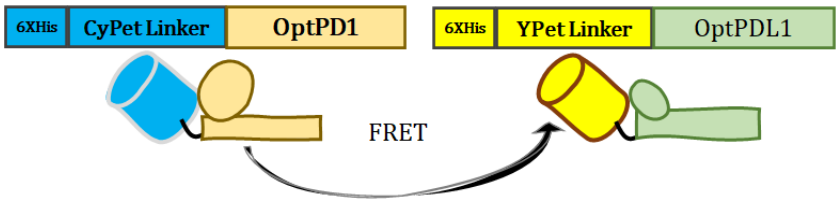


Figure 37: The qFRET assay design with CyPet-PDL1 and YPet-PDL1

5.4 Measuring Em_{FRET}

We monitor the FRET emission as well as the donor and acceptor wavelengths to extract the Em_{FRET} signal out of the combined $Em_{FRETTotal}$. Em_{FRET} is calculated based on equation 1. We measure the emission of the FL_D by excitation at 414 nm and emission at 475 nm, the FL_A excitation at 475 nm with emission at 530 nm. The EmTotal is measured by excitation at 414 nm with emission at 530 nm. We determine the donor contribution by multiplying the donor fluorescence $FL_{D,}$ with the crosstalk parameter alpha (α). The acceptor contribution by multiplying the acceptor fluorescence FL_A, with the crosstalk parameter beta (β). The alpha (α) and beta (β) parameters are unique for each fluorescence instrument and must be characterized initially. The alpha parameter is a unitless ratio of the fluorescence of the donor in the FRET wavelength 530 nm to the fluorescence of the donor in its emission wavelength 475 nm. The beta parameter is also a unitless ratio of the fluorescence of the acceptor when excited at 414 nm and measured emission at 530 nm to the fluorescence at 530 nm when excited at 475 nm. The alpha

parameter is determined previously to be 0.34 ± 0.003 and the beta parameter is determined to be 0.03 ± 0.001 for the Molecular Devices SpectraMax3™.

$$\mathbf{Em_{FRET}} = (\mathbf{Em_{FRET}}) - ((\mathbf{FL_A} * \boldsymbol{\alpha}) + (\mathbf{FL_D} * \boldsymbol{\beta}) \quad \text{Equation 5.1}$$

$$\boldsymbol{\alpha} = \frac{\textit{Donor emisson at 530 nm with excitation at 414 nm}}{\textit{Donor emisson at 475 nm with excitation at 414 nm}} \quad \text{Equation 5.2}$$

$$\boldsymbol{\beta} = \frac{\textit{Acceptor emisson at 530 nm with excitation at 414 nm}}{\textit{Acceptor emisson at 530 nm with excitation at 475 nm}} \quad \text{Equation 5.3}$$

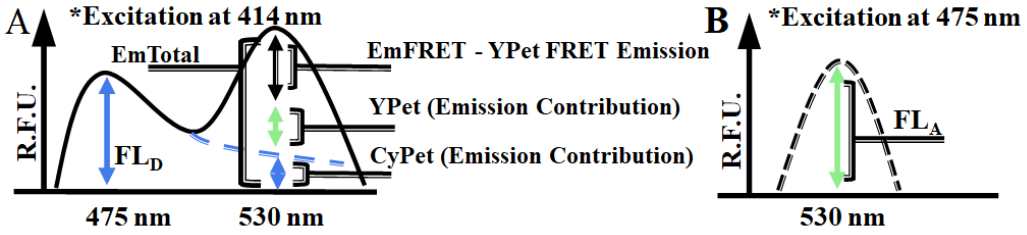


Figure 38: Illustration of the emission spectrum of the protein construct when excited at 414 nm, component EmTotal, is composed the emission from the un cleaved substrate construct, (Em_{FRET}), free YPet protein emission contribution, and CyPet emission contribution at 530 nm. The contribution of CyPet is resolved by monitoring the emission of CyPet at 475 nm when excited at 414 nm, FL_D, and multiplying by the CyPet contribution ratio, α. The YPet contribution is resolved by measuring the acceptor emission at 530 nm, FL_A, when excited at 475 nm and multiplied by β, the YPet contribution ratio (B).

5.5 Z' Qualification Assay Setup

The primary metric for confidence in the HTS assay and inhibitor "hits" is the Z'. The qualification assay is designed for a 384 well plate format and set in a sequence outlined within the assay guidance manual developed . The Z' is resolved by equation 5.4 under three different reaction conditions that are orchestrated to yield a Max, Mid, and Min Em_{FRET} signal from the assay. The Z' considers the difference between the Max signal, the low signal, and the variability in the measurements. The difference between the Max and Min signal in our Em_{FRET} signal represent the nominal distance between our

donor and acceptor "Max" Em_{FRET}, and the lowest Em_{FRET} signal, where the donor and acceptor are most separated for that assay.

The Max signal is the highest consistent Em_{FRET} we can achieve within the reaction conditions for that specific assay. The Min is the minimum Em_{FRET} we can achieve within the reaction conditions. The Mid signal is implemented to fall between the max and min values as close to 50 % of the difference between Max and Min as possible. The variation of the signal is considered by calculating the coefficient of variation, CV, equation 5.5. The criteria for Z' are that the observed Z' must be higher then 0.4, and the percent CV must be lower than 20% .The overall objective for the Z' qualification assay is to observe the assay performance as close to HTS settings as possible. The variation that occurs on a day-to-day basis with formulating assay buffers, protein freeze thaw, and the instrument measurements all are tested and presented in the Z'.

$$\boldsymbol{Z'} = \frac{\left(\boldsymbol{AVG_{Max}} - \frac{\boldsymbol{3SD_{Max}}}{\sqrt{\boldsymbol{n}}}\right) - \left(\boldsymbol{AVG_{Min}} - \frac{\boldsymbol{3SD_{Min}}}{\sqrt{\boldsymbol{n}}}\right)}{\boldsymbol{AVG_{Max} - AVG_{Max}}} \qquad \text{Equation 5.4}$$

$$\boldsymbol{CV} = \frac{\left(\frac{\boldsymbol{SD}}{\sqrt{\boldsymbol{n}}}\right)}{\boldsymbol{AVG}} \qquad \text{Equation 5.5}$$

The assay is conducted at interleaved-signal format which further qualifies the assay for non-uniformity. The format for a 384 well plate follows the outlined guide shown in figure 5. The H is the reactions that produce a Max Em_{FRET} signal, M is Mid, and L is Min. This pattern is repeated across each three plates each of the three days. The Em_{FRET} signal is calculated across the each of the three plates with the Max, Mid, and Min pattern. The robustness and confidence in the assay are directly determined through calculating the Z' and CV from the data collected over the course of three days.

Plate 1

	C1	C2	C3	C4	C5	C6	....	C19	C20	C21	C22	C23	C24
1	H	H	M	M	L	L		H	H	M	M	L	L
2	H	H	M	M	L	L		H	H	M	M	L	L
3	H	H	M	M	L	L		H	H	M	M	L	L
4	H	H	M	M	L	L		H	H	M	M	L	L
5	H	H	M	M	L	L		H	H	M	M	L	L
6	H	H	M	M	L	L		H	H	M	M	L	L
7	H	H	M	M	L	L		H	H	M	M	L	L
8	H	H	M	M	L	L		H	H	M	M	L	L
9	H	H	M	M	L	L		H	H	M	M	L	L
10	H	H	M	M	L	L		H	H	M	M	L	L
11	H	H	M	M	L	L		H	H	M	M	L	L
12	H	H	M	M	L	L		H	H	M	M	L	L
13	H	H	M	M	L	L		H	H	M	M	L	L
14	H	H	M	M	L	L		H	H	M	M	L	L
15	H	H	M	M	L	L		H	H	M	M	L	L
16	H	H	M	M	L	L		H	H	M	M	L	L

Plate 2

	C1	C2	C3	C4	C5	C6	....	C19	C20	C21	C22	C23	C24
1	L	L	H	H	M	M		L	L	H	H	M	M
2	L	L	H	H	M	M		L	L	H	H	M	M
3	L	L	H	H	M	M		L	L	H	H	M	M
4	L	L	H	H	M	M		L	L	H	H	M	M
5	L	L	H	H	M	M		L	L	H	H	M	M
6	L	L	H	H	M	M		L	L	H	H	M	M
7	L	L	H	H	M	M		L	L	H	H	M	M
8	L	L	H	H	M	M		L	L	H	H	M	M
9	L	L	H	H	M	M		L	L	H	H	M	M
10	L	L	H	H	M	M		L	L	H	H	M	M
11	L	L	H	H	M	M		L	L	H	H	M	M
12	L	L	H	H	M	M		L	L	H	H	M	M
13	L	L	H	H	M	M		L	L	H	H	M	M
14	L	L	H	H	M	M		L	L	H	H	M	M
15	L	L	H	H	M	M		L	L	H	H	M	M
16	L	L	H	H	M	M		L	L	H	H	M	M

Plate 3

	C1	C2	C3	C4	C5	C6	....	C19	C20	C21	C22	C23	C24
1	M	M	L	L	H	H		M	M	L	L	H	H
2	M	M	L	L	H	H		M	M	L	L	H	H
3	M	M	L	L	H	H		M	M	L	L	H	H
4	M	M	L	L	H	H		M	M	L	L	H	H
5	M	M	L	L	H	H		M	M	L	L	H	H
6	M	M	L	L	H	H		M	M	L	L	H	H
7	M	M	L	L	H	H		M	M	L	L	H	H
8	M	M	L	L	H	H		M	M	L	L	H	H
9	M	M	L	L	H	H		M	M	L	L	H	H
10	M	M	L	L	H	H		M	M	L	L	H	H
11	M	M	L	L	H	H		M	M	L	L	H	H
12	M	M	L	L	H	H		M	M	L	L	H	H
13	M	M	L	L	H	H		M	M	L	L	H	H
14	M	M	L	L	H	H		M	M	L	L	H	H
15	M	M	L	L	H	H		M	M	L	L	H	H
16	M	M	L	L	H	H		M	M	L	L	H	H

Figure 39: Interleaved-Signal Assay plate format for day 1, 2, and 3. The format follows the 384 well format, with double lanes for each high, mid, and low.

5.6 Results

qFRET Based HTS for Inhibitors of Atg4A

The Z' qualification assay for Atg4A activity is applied here over three days. The Atg4A response to DTT within the reaction buffer provides the Max signal, (1 mM DTT), Min (0 mM DTT), and Mid (0.1 mM DTT). The resolved Em_{FRET} mean is plotted for each plate across the rows. An average of each Max, Mid, and Min along the rows was taken and plotted with standard deviation at each row (Figure 6). The assay also requires a qualitative assessment of the variation at the edges of the plate. No trend in the reads is observed across the plates over the three days, however variation in data is observed. On Day 3 the mid signal is lower than the days 1 and 2, however it is within constraints and does not impact the outcome. The Z' for the entire plate is calculated over the days and tabulated in table 1. The Z' represents the overall score for each plate and is resolved here to be within the constraints of $Z' > 0.4$. The percent coefficient of variation is calculated across each control set and tabulated in table 2. Each value is shown in percentage and observed to be within constraints of $CV < 20\%$.

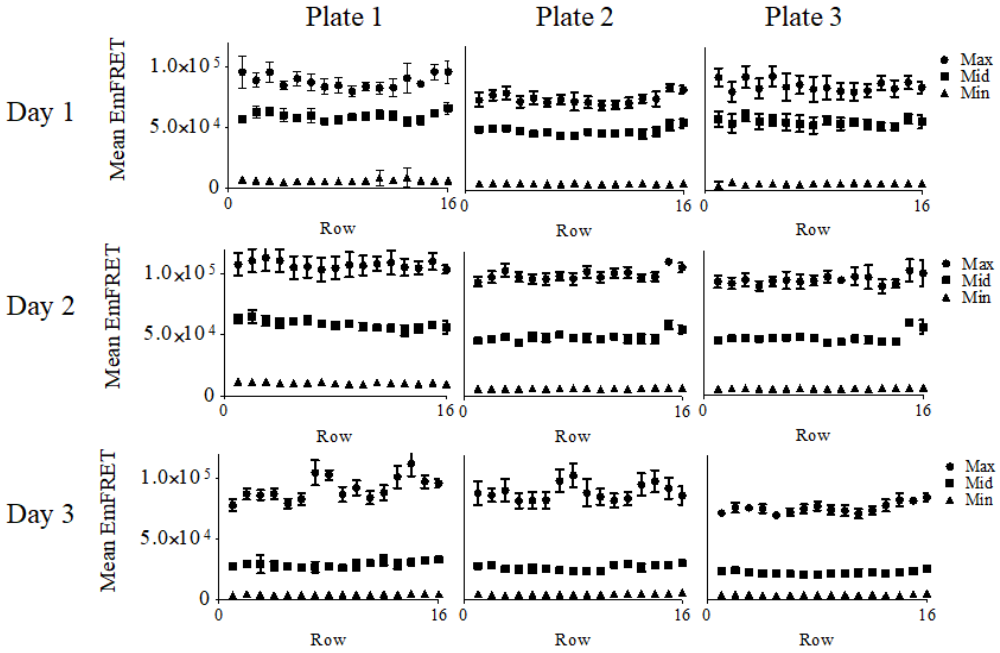


Figure 40: Z' assay of Atg4A response, plot of Em_{FRET} mean of Max, Mid, and Min with standard deviation over each row, across each plate and day. The plots were generated on GraphPadPrism5™. Each data point represents an n of 8, in each row.

Table 79: Z' results for qFRET based HTS for inhibitors of Atg4 at each day on each plate.

	Day 1	Day 2	Day 3
Plate 1	0.96	0.97	0.96
Plate 2	0.97	0.97	0.97
Plate 3	0.97	0.98	0.98

Table 80:Percent Coefficient of Variation across each control set of Max, Mid, and Min.

	Day 1			Day 2			Day 3		
	Plate 1	Plate 2	Plate 3	Plate 1	Plate 2	Plate 3	Plate 1	Plate 2	Plate 3
Max	0.86	0.71	0.88	0.7	0.5	0.64	1.08	0.99	0.62
Mid	0.7	0.69	0.86	0.69	0.81	0.91	1.14	1.01	0.9
Min	3.65	1.87	2.49	1.15	1.33	1.31	1.84	1.97	2.57

qFRET Based HTS for Inhibitors of PD1-PDL1 Interaction

The Z' qualification assay for the PD1-PDL1 HTS results are shown in figure 7. The Guanidine HCL at varying concentrations is used to induce decreased Em_{FRET} signal. An average of each Max, Mid, and Min along the rows was taken and plotted with standard deviation at each row. We observe a negative Em_{FRET} at Min signal across all plates, this could be an artifact of the Guanidine HCL at high concentrations which denatures the fluorescent proteins. The Z' for the entire plate is calculated over the days and tabulated in table 1. The Z' for each plate meet the criteria of Z'>0.4 and pass the main metric to qualify. The percent coefficient of variation is calculated across each control set and tabulated in table 2. Each value is shown in percentage and observed to be within constraints of CV < 20%.

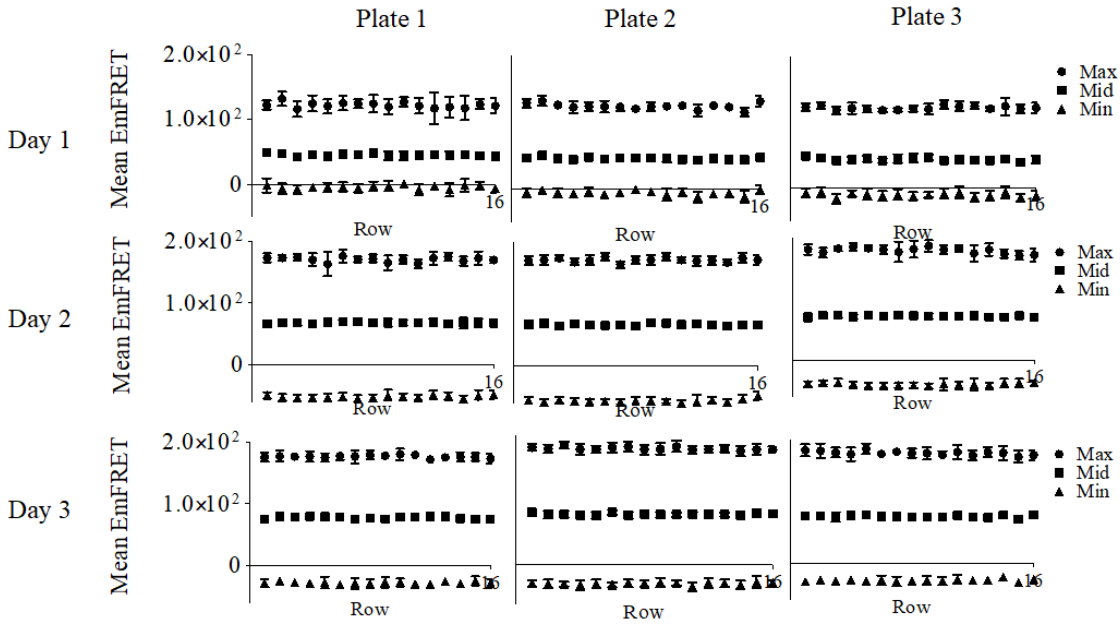


Figure 41: Z' assay for PD1-PDL1 response, plot of Em_{FRET} mean of Max, Mid, and Min with standard deviation over each row, across each plate and day. The plots were generated on GraphPadPrism5™. Each data point represents an n of 8, in each row.

Table 21: Z' Results for each plate on each day

	Day 1	Day 2	Day 3
Plate 1	0.96	0.98	0.98
Plate 2	0.97	0.98	0.98
Plate 3	0.97	0.98	0.98

Table 22:Percent Coefficient of Variation across each control set of Max, Mid, and Min.

	Day 1			Day 2			Day 3		
	Plate 1	Plate 2	Plate 3	Plate 1	Plate 2	Plate 3	Plate 1	Plate 2	Plate 3
Max	0.91	0.53	0.47	0.49	0.36	0.49	0.36	0.34	0.45
Mid	0.94	0.89	1.00	0.62	0.52	0.56	0.48	0.46	0.56
Min	-17.54	-9.70	-5.91	-1.10	-0.96	-1.27	-1.95	-1.83	-1.81

5.7 Discussion

The qFRET based HTS assay qualified here demonstrate a consistent Z' of above 0.9 across all plates tested, under both assays. The rigorous testing across the multiple day's exercises the buffer stability and protein fidelity with multiple freeze thaws. The enzyme activity of Atg4A was consistent across the freeze thaws over the three days. The Atg4A reaction buffer with DTT had to be made at time of use, as the DTT reagent degraded overnight. The response to an oxidative environment that Atg4A provides gives insight into the acute reactivity this mechanism has embedded within the protein sequence and structure. The Z' for the qFRET based HTS of inhibitors of Atg4A was higher than 0.9 across all plates. The difference between the Min and the Max is consistently large with minimal standard deviation across each row. The Min signal that approached zero or went negative is observed to have a larger CV across all plates. This variation is due to the system noise at the lower RFUs. As we approach zero or very low

signal, the system noise is higher than the actual detected signal. Thus, the standard deviation in the Min signal is as high as the calculated Min Em_{FRET}. The result is the mean of the Min signal is only a few factors higher than the standard deviation, and a higher percent CV in the Min signals. This could potentially be problematic for the Atg4A assay as the Min signal represents full protease activity. This issue is minor as the Max signal represents the "Hit" of the assay, and that is consistently several folds higher than the Min.

The qualification for the *in vitro* qFRET based HTS for the inhibitors of PD1-PDL1 interaction was implemented together with the help of Amanda Xaypraseuth. The development of the work is published in the undergraduate journal.[11] The assay utilized Guanidine-HCl to disrupt the PD1-PDL1 interaction and provide the varying signals that are required for Z' qualification. The observed Z' across all plates is within the criteria, the Z'> 0.4. Similarly, the Min values across the plate had higher CV. Though the criteria are met without fail on each plates, we can still attempt to improve the CV. The instrument noise within the system will not change, the future development for this assay can be to characterize the system noise and filter it from the observed Em_{FRET}.

5.8 Methods and Materials

DNA constructs and protein purification for Atg4A and Cypet-GATE16-YPet fusion protein

The protein construct for the substrate were made using recombinant DNA cloning techniques. The protease Atg4A was purchased from Addgene and cloned into pET28b within the expression vector pET-28b. The cDNA for GATE16 was acquired

through Addgene. The substrate protein construct was made using pET28b with CyPet at the N-terminal and YPet at the C-Terminal, with GATE16 centered between restriction sites, Xho1 and Nhe1 restriction sites flanking GATE16 substrate. The construct was previously reported in qFRET study published by Hariharan *et al.*.[12] The BL21(DE3) *E. coli* bacteria were transformed with the two pET28b constructs separately. The strains were expressed overnight at 20 degrees Celsius and were pelleted using centrifugation at 8000xg for 10 mins. The cell pellets were resuspending with lysing buffer, (20 mM Tris HCl pH 7.4, 0.5 M NaCl, and 10 mM Imidazole). The cells were lysed using a probe sonicator at 60 Hz for 5 seconds on and 10 seconds off and centrifuged at 35,000 g for 30 minutes at 4 degrees Celsius.

The supernatant was then subjected to nickel bead affinity chromatography, with three washes at 10 times the volume of the nickel beads in the column. The first wash is a high salt buffer wash, (20 mM Tris-HCl pH 7.4, 300 mM NaCl), the second wash is tween with high salt (20 mM Tris-HCl pH 7.4, 1500 mM NaCl, and 0.5 % Tween X-100), and the last wash, (20 mM Tris-HCl pH 7.4, 500 mM NaCl, and 10 mM Imidazole). After the washes the protein was eluted from the nickel beads using 20 mM Tris-HCl, 50 mM NaCl, and 200 mM Imidazole. The eluted proteins were then left in dialysis buffer (20 mM Tris-HCl pH 7.4, 50 mM NaCl, and 1 mM DTT) over night at 4 degrees Celsius on a stir plate at 200X volume of eluted proteins. The expression and purification of CyPet-PD1 and YPet-PDL1 is noted in chapter 2.

Setup of *Z'* Qualification Assay for the *In vitro* qFRET Based HTS for Inhibitors of Atg4A

The Atg4A enzyme is qualified beforehand for level of Em_{FRET} wanted within each assay. The recommended target to inhibitor concentration is at 1:100 molar ratio of enzyme to inhibitor. The HTS is implemented at 10 µM of inhibitor in each reaction, thus we setup the enzyme to 0.1 µM, lower levels of enzyme are applicable here as well. The substrate complex is added at 0.2 µM to deliver a high Em_{FRET} signal. The inhibitor is dissolved in DMSO and is added to the reaction at up to 3 % volume. The reaction implemented here uses 1.2 µL of DMSO in 60 µL of reaction. The Z' qualifications all implemented at these settings, to mimic the screening conditions as closely as possible.

The protease Atg4A requires a reducing agent in the buffer to function and cleave the fusion protein. The reducing agent used in this reaction is dithiothreitol (DTT). The reactions were done in 384 well plate format and set in a sequence outlined in section 5.5. The "Max" Em_{FRET} reaction provides the highest FRET signal, where the protease is not active, this reaction was setup at 0.2 uM CyPet-GATE16-YPet substrate with 0.1 uM Atg4, with no DTT within the reaction buffer, Tris-HCl pH 8.0 and 150 mM NaCl. The "Mid" reaction buffer is the same as "Max" (Tris-HCl pH 8.0 and 150 mM NaCl) however with 0.1 mM DTT. The "Min" reaction provided the lowest Em_{FRET} signal and the high activity of the protease at 1 mM DTT and standard reaction buffer (Tris-HCl pH 8.0 and 150 mM NaCl). Each buffer is made the day of use, as DTT degrades overnight. The reaction plates are incubated at 37°C for 30 minutes before measuring

Setup of *Z'* Qualification Assay for the *In vitro* qFRET Based HTS for Inhibitors of PD1-PDL1 Interaction

The fusion protein CyPet-PD1 and YPet-PDL1 are kept at 1:1 molar ratio at 0.5 µM. The DMSO % volume is kept below 3% and the reaction buffer is PBS. The "Max" Em_{FRET} signal is with no Guanidine-HCl and will have nominal FRET. The "Mid" signal is orchestrated with half the substrate concentration. The "Min" signal is supplemented with 50 µM Guanidine HCl to inhibit all PD1-PDL1 interactions. Each reaction volume is kept at 60 µM. The reaction plates are incubated at 37°C for 30 minutes before measuring.

Fluorescence Plate Reader

Molecular Devices SpectraMax3™ and FlexStationII^{384TM}, was used to monitor the fluorescence at three wavelengths. The EmTotal, is measured at 414 nm and excitation with emission at 530 nm. The FL_D, CyPet wavelength with excitation at 414nm and emission at 475 nm. The $FL_{A,}$ YPet contribution through excitation at 475 nm and emission at 530 nm. The measurements are taken using SoftMax Pro 7™ at constant PMT gain at "Low". The reaction results were plotted in GraphPad.

5.9 References

1. Maruyama, T. & Noda, N. N. Autophagy-regulating protease Atg4: structure, function, regulation and inhibition. *J Antibiot* **71**, 72–78 (2018).

2. Autophagy: Renovation of Cells and Tissues: Cell. https://www.cell.com/cell/fulltext/S0092-8674(11)01276-1?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867411012761%3Fshowall%3Dtrue.

3. Levy, J. M. M., Towers, C. G. & Thorburn, A. Targeting autophagy in cancer. *Nat Rev Cancer* **17**, 528–542 (2017).

4. Guo, J. Y. *et al.*. Activated Ras requires autophagy to maintain oxidative metabolism and tumorigenesis. *Genes Dev* **25**, 460–470 (2011).

5. Evolutionary optimization of fluorescent proteins for intracellular FRET | Nature Biotechnology. https://www.nature.com/articles/nbt1066?platform=hootsuite.

6. Song, Y., Madahar, V. & Liao, J. Development of FRET assay into quantitative and high-throughput screening technology platforms for protein-protein interactions. *Ann Biomed Eng* **39**, 1224–1234 (2011).

7. Liu, Y., Song, Y., Madahar, V. & Liao, J. Quantitative Förster resonance energy transfer analysis for kinetic determinations of SUMO-specific protease. *Analytical Biochemistry* **422**, 14–21 (2012).

8. Pascovici, D. *et al.*. Clinically Relevant Post-Translational Modification Analyses—Maturing Workflows and Bioinformatics Tools. *International Journal of Molecular Sciences* **20**, 16 (2019).

9. Home - ClinicalTrials.gov. https://clinicaltrials.gov/.

10. Delaunay, M. *et al.*. Management of pulmonary toxicity associated with immune checkpoint inhibitors. *European Respiratory Review* **28**, (2019).

11. Xaypraseuth, A., Madahar, V. & Liao, J. Development Of Quantitative Förster Resonance Energy Transfer (QFRET) Based High Throughput (HTS) Screening For PD-1/PD-L1 Immune-Checkpoint Assay. *UC Riverside Undergraduate Research Journal Submit* **14**, (2020).

12. Hariharan, C. *et al.*. Development of High Sensitive and Quantitative FRET Based Biosensor to Detect Atg4A Kinetics in Autophagy Cell Death Pathway. *Journal of Biomedical Engineering and Biosciences (JBEB)* **8**, 1–10 (2021).

Chapter 6: Conclusion and Future Perspectives

The aim of this thesis is to apply qFRET technology in the characterization of non-covalent and covalent interaction between the "difficult to purify" proteins PD1 and PDL1, and the SUMO modification of viral proteins. The method outlines how to process the raw data to extract the FRET signal using correction parameters that can be derived for any commercial fluorescent plate reader. The secondary aim is to present the qFRET technique not only as a plug and play but as a hands-on technique with the versatility to investigate a broad range of protein-protein interactions. The qFRET platform is an accessible technique that scientists can apply in a multifaceted way to investigate complex protein-protein interactions that are non-covalent or covalent.

The qFRET based characterization of PD1-PDL1 interaction reported a K_D of 0.82 µM. In comparison to the past investigations of PD1-PDL1 the K_D values ranged from 8.2 µM to ~ 0.5 µM. This reported variation could be due to the inherent pitfalls in immobilization requirement in SPR and could also be because here the full-length protein was used. The qFRET technique applied here provides an alternative method that removes many of the initial processing steps, such as immobilization, in the assay that can introduce variances in the measurement. Furthermore, the qFRET K_D is also applied towards the assessment of refolding proteins. This allowed for a rapid screening of rescued proteins that otherwise use a qualitative method such as aggregation reporters to gauge refolded proteins. The assay established here provides a framework for the assessment of other difficult to purify proteins. The alternative binding partners to PD1 such as Programmed Death Ligand 2 (PDL2), and their interaction has also been

demonstrated to circumvent the immune response from T cells. Furthermore, this assay is developed in a 384 well microplate format allowing seamless transition for high-throughput screening for novel inhibitors.

The detection of covalent attachment of SUMO1 to IAV M1, and SARS-CoV-2 N protein is successfully demonstrated here. The qFRET method provides a rapid assessment of the covalent interaction between the target viral proteins, and the SUMO enzymatic cascade. The evaluation of IAV-M1 resulted in the discovery of an essential lysine that is proven fatal to the virus upon mutation. The MS analysis of the qFRET assay provided 4 novel, and 1 previously discovered SUMOylation site. The method outlined here is advantageous in the versatility of using the entire SUMOylation cascade *in vitro*. The entire SUMOylation event is a complex enzymatic cascade involving the synchronized function of three enzymes. The *in vitro* application of the SUMOylation cascade yields a heavily modified protein that can be directly applied to MS analysis.

The *in vitro* SUMOylation of SARS-CoV-2 N protein provided novel SUMOylation sites, lysine 61, 65, 347, and 355. This is the first reported SUMOylation of the novel SARS-CoV-2 N protein and was also found to modulate in-cell activity. The fluorescent imaging analysis of N protein mutants demonstrated a modulation in cellular localization with lysine 65 mutation. Furthermore, the *in vitro* SUMOylation assay demonstrated a modulation in SUMO attachment with lysine 61 and 65 mutation without E3 present. The combined observations of, MS analysis, *in vitro* qFRET based SUMOylation assay, and fluorescent imaging of mutant N protein, we can conclude that lysine 61 and 65 are likely SUMOylated.

Appendix A: Curriculum Vide

VIPUL MADAHAR
Email:vipulmadahar@gmail.com

EDUCATION

- Doctor of Philosophy, Bioengineering Class of 2021*, University of California Riverside
- Master of Science, Bioengineering Class of 2011, University of California, Riverside
- Bachelor Of Science, Bioengineering Class of 2008, University of California, Riverside

PUBLICATIONS

- Cell Press: "SUMOylation is an Essential Host Pathway for Influenza Viruses Life Cycle with Minimal Mutation Escape Potential" – Submitted August 2021 Under Review
- Medicine in Drug Discovery: Volume 2, April 2020, Accepted: April 2020 " Target Virus or Target Ourselves for COVID-19 Drugs Discovery?"
- Analytical Biochemistry: Volume 422, March 2012 "Quantitative Förster resonance energy transfer analysis for kinetic determinations of SUMO-specific Protease"
- Analytical Biochemistry: Volume 422, March 2012 "Quantitative Förster resonance energy transfer analysis for kinetic determinations of SUMO-specific Protease"
- Annals of Biomedical Engineering, 2010 "Development of FRET Assay into Quantitative and High-throughput Screening Technology Platforms for Protein-Protein Interactions"
- Undergraduate Student Journal of University of California Riverside: Volume 2, 2008 "High-throughput Screening Assay of Small Molecule Inhibitor(s) of PIAS1"

CONFERENCES, PRESENTATIONS AND AWARDS

- Dissertation Year Program Fellowship 2021
- Teaching Assistant of the Year Award 2019
- National Science Foundation – Innovator Corps (NSF-ICOPRS)
  - – Participated as Leader of team 2020
- MolMed Drug Discovery Symposium: Poster 2018 – **Awarded Conference Poster Award*
  - Madahar V. *et al.* "Improve protein expression and refolding methodologies for difficult-to-express proteins to expand the capability of the qFRET technology platform"
- Presentation Bioengineering Colloquium Presentation 2018
  - "SUMOylation of Influenza Virus A Matrix 1 protein with PIAS 1 as E3 ligase"

- UC Systemwide Bioengineering Symposium: Poster 2015
    - Madahar V. *et al.* "Development of FRET based HTS for small molecule inhibitor of atg4"
- Dean Graduate Student Fellowship 2015
- Illumina Values Award Innovator Q3 2014

RESEARCH PROJECTS

Determination of the SARS-CoV-2 Nucleocapsid protein SUMO modified Lysine

- Biological significance of SUMOylation of N protein is the development of a novel drug discovery platform for the treatment of viral infection. Host factors such as SUMOylation are essential for the progression of the viral infection.
- Utilized Mass Spectrometry to determine lysine residue modification from an in-vitro SUMOylation reaction
- Developed a quantitative FRET assay for the determination of SUMOylation event

Determination of the SUMOylation of Influenza Virus A Matrix 1 protein as essential for Viral Pathogenesis

- Biological significance of SUMOylation of M1 protein is the development of a novel drug discovery platform for the treatment of viral infection. Host factors such as SUMOylation are essential for the progression of the viral infection.
- Utilized Mass Spectrometry to determine lysine residue modification from an in-vitro SUMOylation reaction
- Developed a FRET assay for the determination of SUMOylation event
- Viral Plaque assay used to determine viral particle infectivity
- Manuscript –
  "Determining SUMOylation site of Matrix 1 protein of Influenza Virus A, with PIAS1 as E3 ligase"
- Manuscript –
  "SUMOylation is an Essential Host Pathway for Influenza Viruses Life CyLe with Minimal Mutation Escape Potential" –Submitted in Cell Press

Rescue of tagged Programmed Cell Death Ligand 1(PDL1) and Programmed Cell Death 1(PD1) from Ibs

- Biological Significance of PDL1 and PD1 are used to overcome immune checkpoints by cancerous cells
- Expression of PDL1 and PD1 external domains within *E.Coli* B strain into inclusion bodies
- Successful purification of full length PDL1 and PD1 using IMAC protocol under denaturing conditions
- Successful determination of a FRET based K_i of sub nano molar
- Expressed full length fluorescently tagged PD1 and PDL1 within HEK-293

Optimization of Codon Usage based on Codon Adaptation Index and mRNA structure

- Biological significance of codon optimization provides the ability to improve expression of genes within *E coli*, CHO, and HEK293
- Established genomic harmonization algorithm based on the updated frequency of codon usage

Development of FRET based High throughput Screening (HTS) assay for small molecule inhibitors of SUMOylation E3s

- Summer 2018 - Present
- Biological significance of SUMOylation E3s is to inhibit a specific substrate and not inhibit the SUMOylation system
- Determined optimum assay conditions while minimizing concentration and achieving maximum SNR
- Successfully measured HTS metrics to validate assay for drug screening
- Established virtual screening of Maybridge Compound Library and National Cancer Institute Compound Library

Development of FRET based High throughput Screening (HTS) assay for small molecule inhibitors of Atg4A

- December 2017- April 2018
- Biological significance of Atg4A is its ability to toggle the autophagy system and inhibition of Atg4b could yield a novel treatment for cancer
- Determined optimum assay conditions while minimizing concentration and achieving maximum SNR
- Successfully measured HTS metrics to validate assay for drug screening
- Established virtual screening of Maybridge Compound Library and National Cancer Institute Compound Library
- Utilized ChemmineR library to mine for similar structures to narrow down compound library
- Manuscript –
  "Development of FRET based High throughput Screening (HTS) assay for small molecule inhibitors of Atg4B"
- Poster Presentation –
  "Development of FRET based High throughput Screening (HTS) assay for small molecule inhibitors of Atg4B"

Expression and purification of PIAS1a SUMOylation E3 from a variety of hosts

- February 2017- September 2017
- Biological Significance of PIAS1a is its function as a E3 ligase and its part in immune response within the JAK/STAT pathway
- Successful expression of PIAS1a within CHO cells
- Establish a deactivated-CRISPR Cas9 system to knock down PIAS1a

WORK EXPERIENCE

Senior Research Associate within Systems Integration at Illumina Inc.

-September 2011-September 2015

- Research Associate 2 within Chemistry Development at Illumina Inc.
- Successfully completed tasks within a technical development team on multiple projects across multiple sites
- Successfully lead team in communicating and resolving technical issues within a challenging environment
- Successfully transferred several projects from research phase onto product development

Research Assistant in bioengineering lab at U.C. Riverside

-October 2009-2011

High throughput screening assay development of small molecule inhibitors of sumo pathway and protease

- Successfully screened 12,000 compounds from U.C. Riverside Genomics compound library
- Investigation of crosstalk between two pathways using Förster's resonance energy transfer(FRET) as reporter
- Successfully cloned and tagged key genes within the ubiquitin pathway
- Optimized conditions of small-scale expression and purification of genes, involved utilizing HPLC
- Live mice study with knock out globular-protein receptor gene for lipids

Research Assistant within Ambryx Biotechnology part time -November 2008-June 2011

- Internship from November 2008 – February 2009
- Production of Cancer therapeutic peptide
- Involved with veterinarian study of peptide on treatment of dogs with cancer
- Process materials following strict FDA standards within SOP
- Retain a strict standards of quality control following FDA standards

Research Assistant Argus Bioscience -June 2008-June 2009

- Establish assay platform within the Small Ubiquitin like Modifier (sumo) pathway in vivo and *in vitro* – Lead a team on techniques of cloning and expression of genes in in-vitro and in-vivo systems
- Determination of the equilibrium constant k_D using Förster's resonance energy transfer (FRET) – Utilized Biacore™ SPR technology to investigate and measure K_d

Undergraduate Research Assistant in bioengineering lab at U.C. Riverside -June 2006-June 2008

- High-throughput screening assay development for small molecule inhibitor(s) of PIAS1 alpha

RELEVANT SKILLS & QUALIFICATIONS

- Proficient with **molecular cloning techniques such as PCR, DNA ligations/digestions, and to all that pertains**
- Proficient with **gene expression techniques within E.Coli(bl21-DE3) and mammalian cell(HEK293)**

- Proficient in **protein purification and characterization methods**(Affinity Chromatography & Western Blots)
- Proficient with **High Performance Liquid Chromatography**(HPLC)
- Proficient with **Flex station II™ and Softmax Pro Software**(High-throughput plate reader)
- Proficient with **DNA sequencing technology** such as MiSeq and HiSeq-2000 and 2500
- Proficient with **Beckman Coulter Biomek® FX Laboratory Automation Workstation and Software**
- Familiar with **Biacore™**(SPR Technology)
- Computer Software /OS: **COMSOL™, Matlab™, LabView™, R, Python, SAS** and windows/Linux

Community and Academic Outreach

- Graduate Student Mentorship Program 2019
- Gradaute Student Panel for Undergraduate-BMES 2018 and 2019
- Queens-Town Meeting Drug Discovery Shanghai, China -Volunteer for the International Conference
- BioEngineering Gradaute to Undergraduate Mentor 2015 and 2016
- BioEngineering Gradaute Student Association Board Member: Finance Chair 2016

Thank you for your consideration.