Evaluation of split-superpositive GFP and split-cerulean fluorescent proteins as

protein complementation assays for evaluating competitive binding proteins

A Senior Thesis Presented to

The Faculty of the Department of Biology,

The Colorado College

By

Wade Banta

Bachelors of Arts Degree in Biology
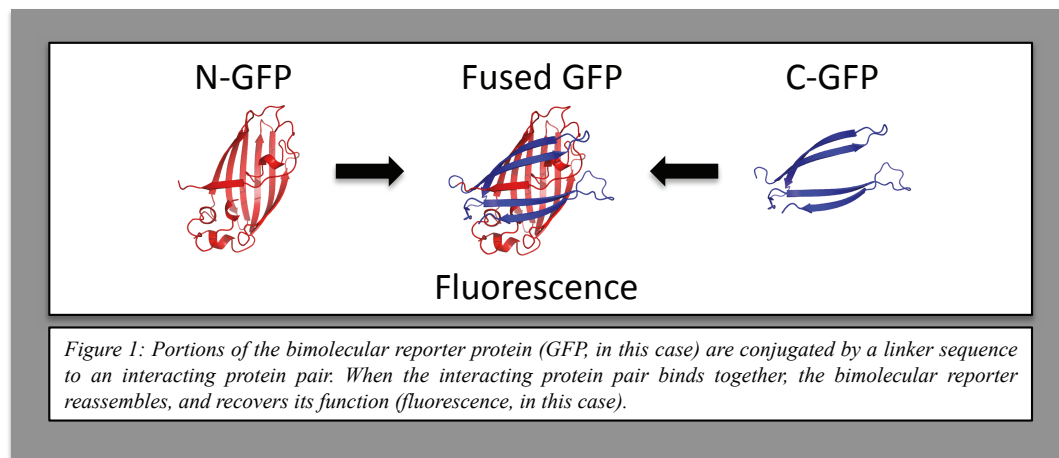
19th day of May, 2014

Dr. Darrell Killian
Primary Thesis Advisor

Dr. Murphy Brasuel
Secondary Thesis Advisor

Abstract

Superpositive green fluorescent protein (GFP) is a designed recombinant bimolecular form of sg100 GFP, the most commonly used variation of native GFP, with a high theoretical net charge, which reduces aggregation, achieved by replacement of aspartic acid residues with arginine and lysine residues. Here, we expressed superpositive GFP as separate N-terminal and C-terminal domains (split superpositive GFP; see Figure 1) in order to develop an assay for the detection of proteins binding to one another and reconstituting functional GFP fluorescence.

Figure 1: Portions of the bimolecular reporter protein (GFP, in this case) are conjugated by a linker sequence to an interacting protein pair. When the interacting protein pair binds together, the bimolecular reporter reassembles, and recovers its function (fluorescence, in this case).

This recovery of fluorescence upon protein pair binding provides an easily observed and semi-quantitatively measurable analogue for the successful interaction of the proteins. The objective of this project was to use the split-superpositive GFP protein complementation assay in conjunction with free or orthogonally reporting inhibitor proteins. We hoped to find that the presence of high affinity free inhibitor proteins would correlate with a decrease in recovered fluorescence when compared to cells in which a lower affinity inhibitor, or no inhibitor, is present. As a secondary objective, we also tested whether an inhibitor conjugated to the N terminal fragment of split cerulean fluorescent protein would be capable or providing an orthogonal cerulean signal upon

competitive binding with the C terminal fragment of split-superpositive GFP. Evaluating time dependent evolution of fluorescence within cells expressing this system using flow cytometry could provide a novel method of screening macromolecular protein-protein interactions and inhibitors of these interactions rapidly and easily.
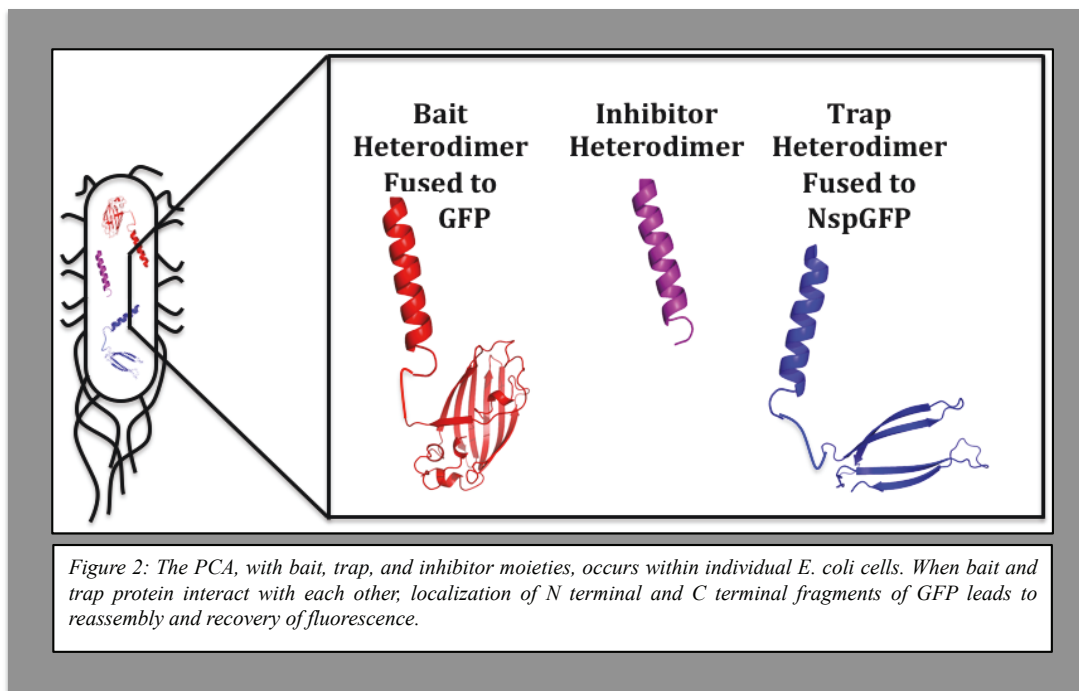
Introduction

Recent years have witnessed a fundamental shift in pharmaceutical strategies, namely, increasing emphasis placed on macromolecular drug development, in addition to the traditional drug development model of developing small organic molecules that interact with target receptor proteins and enzymes. This has led to a greater need for high-throughput screening technologies to identify evolved or engineered proteins with therapeutic potential.

Protein complementation is the main biological mechanism of control over diverse cellular activities, from responses to changes in the cellular environment to mediation of the cell cycle. As a result, monitoring protein-protein interactions provides insight into these processes and allows understanding of the macromolecular interactions that occur in a normally functioning cellular environment. In addition, because of the importance of these interactions functioning correctly in a healthy cellular environment, errors in expression level or mutations that change the affinity of interacting proteins can lead to diseases by interrupting the control these macromolecular interactions hold over cellular activities. For these reasons, the information provided by Protein Complementation Assays (PCAs) has many useful applications in cellular/molecular biology, biochemistry, and medicine.[3]
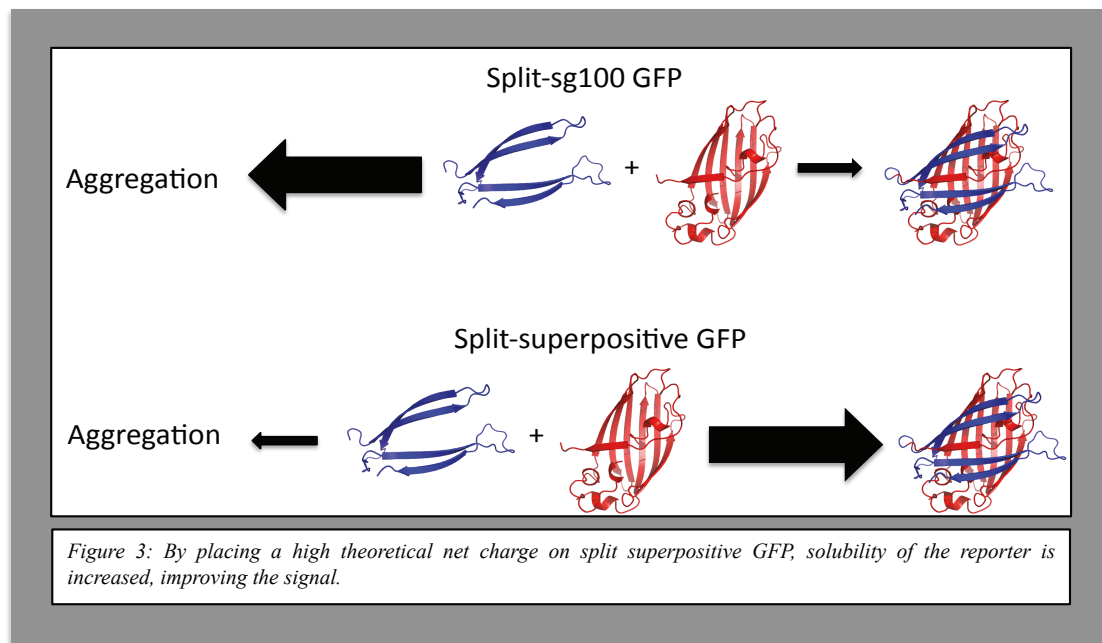
Split-superpositive GFP is a PCA reporter system that has been shown to give a binary signal in the form of fluorescence for protein-protein binding (see Figure 2).[4] The central purpose of this project is to determine if split-superpositive GFP can accurately evaluate competitive inhibitors of a bait/trap protein pair and show a decrease of fluorescence intensity associated with inhibitor interference with the reassembly of split-superpositive GFP.

The major desirable factors in the use of split superpositive GFP as a protein complementation reporter are its variable output and ability to apply *in vivo*. Other PCA's, which use bimolecular enzymes with a recovery of function producing a binary cell survival signal, such as split β-lactamase or split dihydrofolate reductase, would not be appropriate reporters as they are more difficult to generate a variable signal with, and often require the addition of exogenous material, such as dye which produces a colorimetric change, to produce a signal.

Figure 2: The PCA, with bait, trap, and inhibitor moieties, occurs within individual E. coli cells. When bait and trap protein interact with each other, localization of N terminal and C terminal fragments of GFP leads to reassembly and recovery of fluorescence.
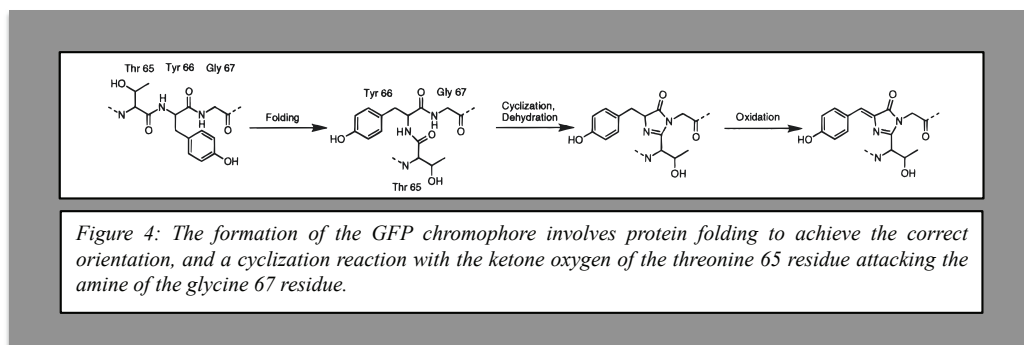
Reporter Proteins

Split superpositive GFP was chosen as a bimolecular GFP reporter over split sg100 GFP, another available isoform of recombinant split GFP. Split superpositive GFP is highly modified with the replacement of solvent exposed residues by lysine and arginine to provide a high theoretical net positive charge. The introduction of these residues raises the net charge of the N terminal fragment from -4 to +24 and the C terminal fragment from -4 to +10 (see Figure 3).

Figure 3: By placing a high theoretical net charge on split superpositive GFP, solubility of the reporter is increased, improving the signal.

Because protein solubility generally decreases as a protein approaches its isoelectric point, the additional lysine (pKa 10.53) and arginine (pKa 12.48), which raise split superpositive GFPs isoelectric point, improve the proteins resistance to aggregation. The result of this improved solubility is a dramatic increase in reassembly speed of split superpositive GFP over splitSG100 GFP, and a much higher fluorescence signal.[4]

The recovery of fluorescence of split GFP involves a covalent rearrangement of bonds in two key residues to form the GFP chromophore (refer to figure 4). As time

passes, the cells saturate with recovered GFP until all split GFP in both cells containing

inhibitor and cells not containing inhibitor contains active chromophores. It is important

to note that although the formation of the chromophore involves a covalent bond

formation, the residues involved all reside on the N terminal fragment of the split GFP.

The reassembly of the two fragments allows the formation of the chromophore because

the structure of the beta barrel of GFP stabilizes the vector of nucleophilic attack that

allows the cyclization reaction to occur and enhances the chemical environment for

nucleophilic attack. This improves the kinetics of the reaction to allow the formation of

the chromophore.

Figure 4: The formation of the GFP chromophore involves protein folding to achieve the correct orientation, and a cyclization reaction with the ketone oxygen of the threonine 65 residue attacking the amine of the glycine 67 residue.
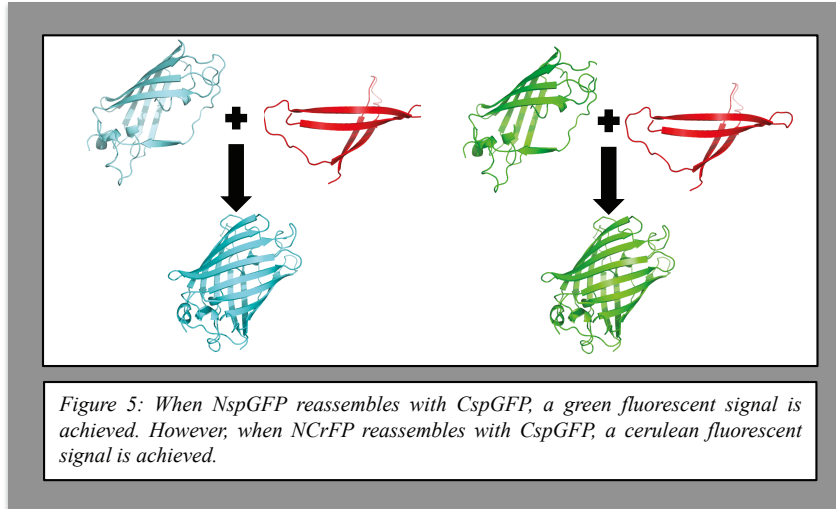
No covalent bonds form between the N and C terminal fragments of split GFP.

The result of this is that although the formation of the chromophore is not technically

irreversible, there is a large energetic barrier to free N terminal fragment GFP with an

intact chromophore reverting to inactive N terminal fragment GFP with a nonfunctional

chromophore. This makes the formation of the chromophore "functionally irreversible"

in the context of the cellular environment, and permits the saturation of cells with active

fluorescent N terminal fragment GFP to occur.

Because the accumulation of actively fluorescent N terminal fragment split

superpositive GFP raises the fluorescence intensity of the cells in the model system, it is

best to think of the parameters observed as a time dependent recovery of fluorescence approaching saturation. It is our hypothesis that introducing an inhibitor to the bait/trap protein pair in the system will lower the rate of accumulation of actively fluorescent N terminal fragment superpositive GFP by altering the kinetics of reassembly. To evaluate whether or not this was the case, it was important to observe the fluorescence intensity of the cells expressing the system at an appropriate time point after induction.

Ideally, this time point would be sufficiently long after induction to observe fluorescence intensity in the cells expressing the system clearly higher than that of the negative controls, but also not so long that the cells had approached complete saturation. The ideal time point would show maximum or near maximum fluorescence in the positive controls, but intermediate saturation in the cells expressing the inhibitor.

A second reporter molecule pairing was used in a secondary experiment to test orthogonal reporting properties of split superpositive GFP and split cerulean protein. Split cerulean protein contains a high degree of sequence similarity with split superpositive GFP. In fact, the C terminal fragments of split cerulean protein and split GFP are identical except for a single residue change, which is included in the mutation of split GFP to split superpositive GFP. This means that the N terminal fragment of split cerulean protein could potentially reassemble with the C terminal fragment of split superpositive GFP and produce a 475 nm cerulean signal, while the N terminal fragment of split superpositive GFP could, within the same cell, reassemble with the C terminal fragment and produce a 507 nm green signal (see Figure 5).

Figure 5: When NspGFP reassembles with CspGFP, a green fluorescent signal is achieved. However, when NCrFP reassembles with CspGFP, a cerulean fluorescent signal is achieved.

Flow Cytometry

The primary instrumental procedure used in this project was flow cytometry. Flow cytometry is a technique that has evolved since the first device utilizing the Coulter principle, the main principle that allows the usefulness of the first flow cytometry technique, was invented in 1953. The Coulter principle states that cells passing through an aperture between two separated chambers increase electrical resistance across the chambers. This principle allowed researchers to accurately count the number of cells in a sample by passing that sample across a small aperture separating two chambers whose electrical conductance was measured by electrodes in each chamber.
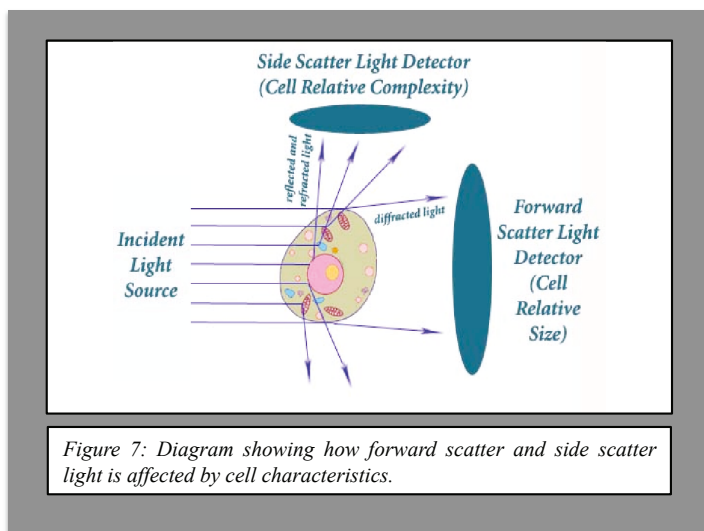

Figure 6: Diagram of flow cytometer apparatus.
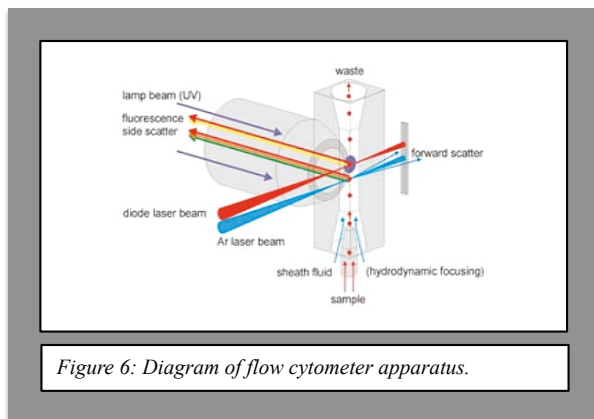
Modern flow cytometry devices utilize optical deflection techniques as a method of accurately measuring discrete characteristics about cell population. They can also utilize fluorescent markers as a further

method of distinguishing cell populations. By using modern computing technology and special software, millions of cells can be analyzed in a few minutes. Flow cytometry is also used in medical diagnostics, such as counting cell type components of whole blood samples.

Modern flow cytometry uses laser deflection to quantify two characteristics of cell populations. To accurately separate an entire cell population into single cells for analysis, flow cytometers stream a sheath fluid around the diluted sample, which spreads the sample out before passing it in front of the laser light source and detector (Refer to figure 6). The computer automatically discards data points than could be two or more cells. When cells pass in front of the laser, a detector behind the light source collect information about the cell in two parameters: Forward scatter and side scatter. Depending on the laser wavelength used, there can also be a fluorescence detector that will measure the fluorescence intensity of the cell (Refer to figure 7).

Figure 7: Diagram showing how forward scatter and side scatter light is affected by cell characteristics.

The forward scatter parameter refers to the amount of light deflected at a low angle from the path of the laser measured by a detector in line with the source of laser light. This can be imagined as a measure of the shadow cast by the cell as it passes in front of the laser light source. Forward scatter is a comparative measure of the volume of

a cell, since the area obscured from the forward scatter light detector by the cell is roughly proportional to the size of the cell.

The side scatter parameter refers to the amount of light deflected at a high angle by the internal structures of a cell, and subsequently measured by a light detector to the side of the path of the laser light source. This provides a measure of the internal complexity of the cell because when more internal structures (granules or internal membranes) are present, more light is deflected.
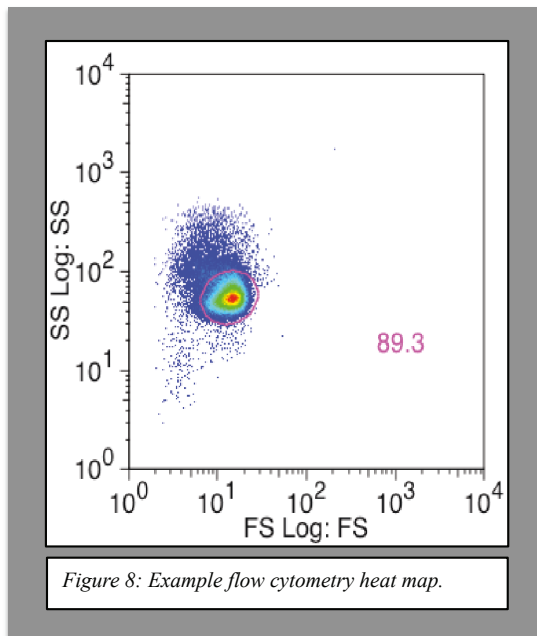

Figure 8: Example flow cytometry heat map.

The raw data collected from flow cytometry forms a heat map, where the X-axis shows forward scatter values, and the Y-axis shows side scatter values. Colors range from blue to red, where "cool" colors correspond to a low number of hits at that point of forward scatter/side scatter values, while increasingly "hot" colors correspond to higher number of hits with the same forward scatter/side scatter values (refer to figure 8). This heat map can be imagined as a 3-dimensional histogram, with red points at the peak of the histogram, sloping downwards to blue points at the base.
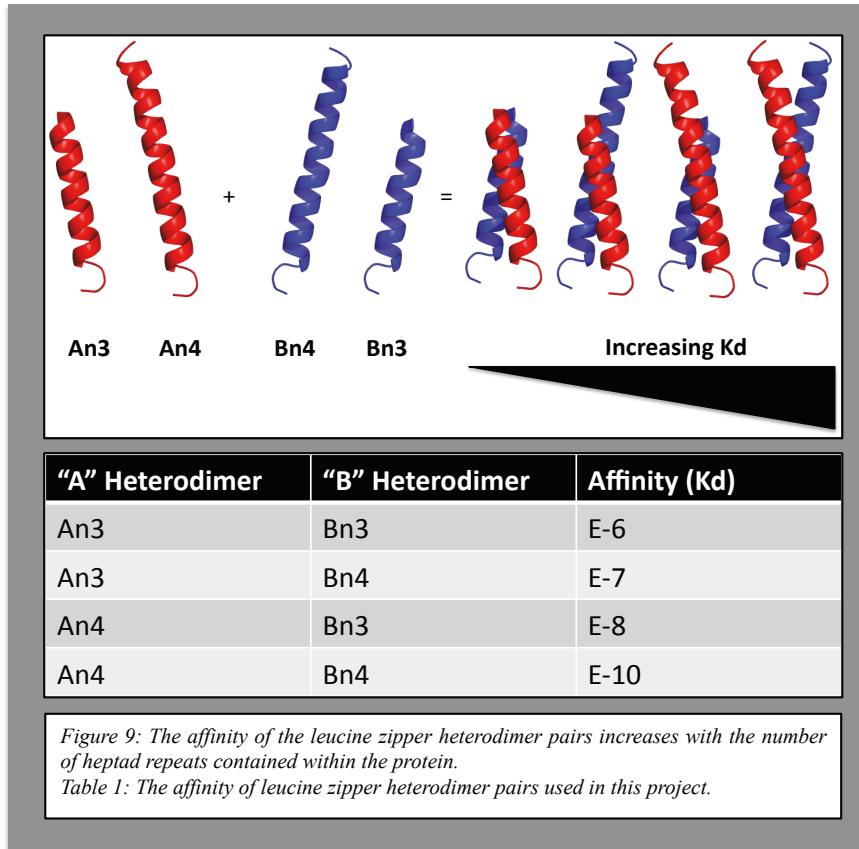
When this raw data is collected, generally only the top ~80% of the cell population is selected and segregated from the total data set by selecting the mostly green, yellow and red regions. The rest of the data is discarded as outlying data points, as

unhealthy cells and debris can be present in this fraction, and this can interfere with the results.
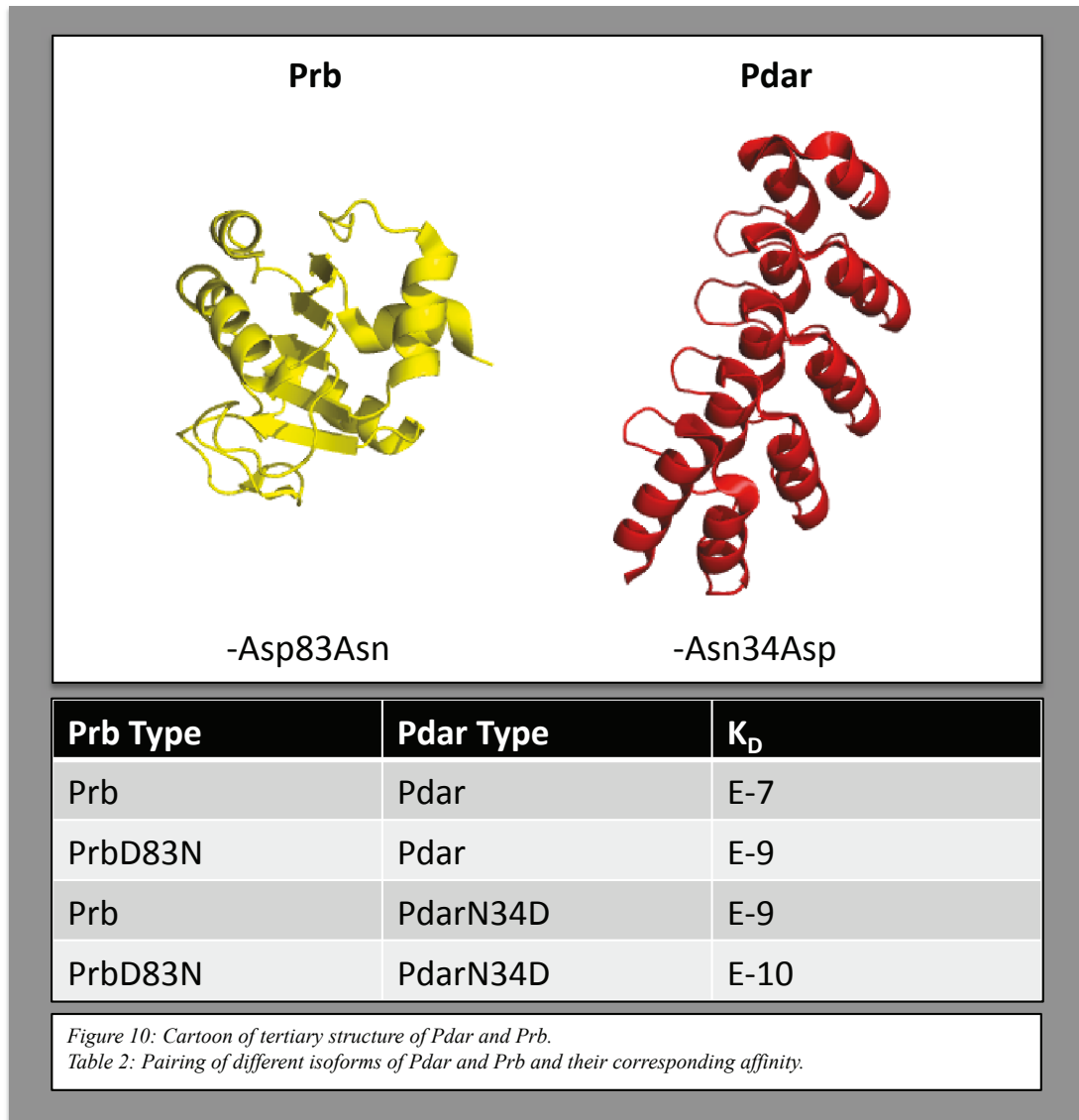
Once the desired set of cells is selected, the raw data can be manipulated to display different characteristics of interest, although usually this is a fluorescence intensity, as transgenically expressed fluorescent proteins or recombinant molecules conjugating a fluorescent molecule to a protein that recognizes a cell surface receptor are often used to unambiguously distinguish cell types that may share physical characteristics.

Bait-Trap Proteins

The bait-trap proteins designed by Thomas et al (2013) were chosen for use as a model to test the split superpositive GFP protein complementation system because the design of the proteins provides highly specific binding with a suitable range of affinity for interacting protein pairs. Leucine zipper coiled-coils are a common protein-binding motif of native proteins. Leucine zippers consist of a heptad repeat alpha coil exposing polar residues to the solvent surface and hydrophobic residues to the dimer interface. The affinity of the designed heterodimer pairs (A chain and B chain) used here increases with additional heptad repeats. Replacing isoleucine residues with asparagine (An chain and Bn chain) in the dimer interface residues of the heptad repeats reduces complex stability and increases specificity, providing predictable K_D values for heterodimer pairs (Refer to figure 9). These designed heterodimers are particularly useful because K_D of these heterodimer pairs spans the micromolar to subnanomolar range.

"A" Heterodimer	"B" Heterodimer	Affinity (Kd)
An3	Bn3	E-6
An3	Bn4	E-7
An4	Bn3	E-8
An4	Bn4	E-10

Figure 9: The affinity of the leucine zipper heterodimer pairs increases with the number of heptad repeats contained within the protein.
Table 1: The affinity of leucine zipper heterodimer pairs used in this project.

A second pair of bait/trap proteins was also used in the same experimental setup to test the split superpositive GFP system with more complex, naturally occurring protein pairs. The pair chosen for this project was Prb and Pdar, a Coenzyme A protein and ankyrin repeat protein that bind with affinity in the submicromolar region in natively binding protein pairs (Refer to figure 10). Karacinolas et al (2011) showed that site directed mutagenesis of a single residue in each protein provides a combination of protein pairs with affinities from the submicromolar region (both proteins in native form) to the subnanomolar region (mutant PrbD83N and mutant PdarN34D).

Prb	Pdar
-Asp83Asn	-Asn34Asp

Prb Type	Pdar Type	K_D
Prb	Pdar	E-7
PrbD83N	Pdar	E-9
Prb	PdarN34D	E-9
PrbD83N	PdarN34D	E-10

Figure 10: Cartoon of tertiary structure of Pdar and Prb.
Table 2: Pairing of different isoforms of Pdar and Prb and their corresponding affinity.

These protein pair combinations have a similar range of affinity to the designed leucine zipper heterodimers used in this project, as well as the added benefit of a complex tertiary structure that is most commonly the case with native biologically sourced interacting proteins. Applying the methods used in this project to these interacting proteins would further demonstrate the ability of this protein inhibition assay to apply to biologically relevant protein pairs.

Materials and Methods

Cloning Setup

To express inhibitor proteins and bait proteins in roughly equal amounts we used a dual expression system in BL21 *E. coli* to express three proteins from two vector plasmids. All cloning operations were performed by traditional cloning techniques entailing PCR amplification with New England Biolabs (Ipswich, MA) Q5 high fidelity DNA polymerase of inserts containing the coding sequences for the proteins used, restriction endonuclease digestion of PCR inserts and plasmid vectors, and ligation of the two restriction digest products with New England Biolabs Quick Ligase to complete the construct. Quality control measures taken during cloning operations included gel purification of PCR product and treatment of digested vector with New England Biolabs Calf Intestinal Phosphatase.

To express inhibitor protein and bait protein proportionally, we used an ampicillin resistant IPTG-inducible pETDuet plasmid with two multiple cloning sites (MCS). MCS1 of the pETDuet plasmid always contained bait protein conjugated to the C terminal fragment of split superpositive GFP. MCS2 of the pETDuet plasmid always contained an inhibitor protein. In the set of experiments run using the leucine zipper heterodimers and the set run using the native proteins, these proteins remained unconjugated to a reporter molecule fragment. However, with the orthogonally reporting split cerulean fluorescent protein experiments, the inhibitor protein in MCS2 was conjugated to the N terminal fragment of split superpositive GFP. Positive controls for each experiment were obtained by cloning pETDuet plasmids with bait protein/N terminal fragment superpositive GFP

fusion in MCS1 and an empty MCS2, so that no inhibitor protein was present in the system.

The trap protein motif was expressed using a kanamycin resistant L-arabinose-inducible pMRBAD plasmid vector with a single multiple cloning site. This multiple cloning site always contained a trap protein/N terminal split superpositive GFP fusion.

Plasmid sequencing

Plasmids were sequenced using Genewiz (Plainfield, NJ) commercial sequencing services. pETDuet was sequenced with pETDuet forward and reverse sequencing primers (sequences available from manufacturer). pMRBAD plasmids were sequenced with a pMRBad reverse sequencing primer. Sequence files were evaluated using SerialCloner and Geneious software programs.

Dual Plasmid Expression

To express all three protein motifs within the same cell, we used a high yield protein expression BL21 *E. coli* strain that was transformed with both a pETDuet plasmid and a pMRBAD plasmid. First, pETDuet was transformed into chemically competent BL21 *E. coli* using a standard heat shock transformation protocol. For the second transformation, these *E. coli* strains were made electrocompetent. To generate electrocompetent cells, strains were grown to saturation in 2XYT media (16g/L tryptone, 10g/L yeast extract, and 5g/L NaCl adjusted to pH 7.0) with effective concentrations of ampicillin to ensure that only cells containing both desired plasmids grew. These cell cultures were then cooled in an ice water bath, centrifuged and the supernatant was discarded. The cell pellet was then resuspended in cold 10% glycerol solution, centrifuged, and the supernatant was discarded. The completed electrocompetent cells

were then split into 140 μL aliquots, which were then transformed with pMRBAD plasmid by electroporation using a Biorad (Richmond, CA) Gene Pulser XCell cuvette electroporation apparatus. After electroporation they were rescued using room temperature SOC media and subjected to ampicillin and kanamycin selection. Any electroporation ready BL21 aliquots were labeled by the contents of their pETDuet plasmid, flash frozen in a dry ice and acetone bath, and stored at -80°C.

In preparation for flow cytometry, 5 mL cultures of each strain transformed with both plasmids were grown to optical density (OD) 0.5 at 37°C for optimum protein expression cellular conditions in media containing both ampicillin and kanamycin. Once all cultures reached OD 0.5, both plasmids were simultaneously induced using 50μL 10X 10% w/w arabinose stock solution and 5μL 100X 1M IPTG stock solution. After induction, cell cultures were incubated at variable temperature for variable time periods (indicated in the Results) depending on the experiment being performed.

Flow Cytometry

Flow cytometry experiments were performed with a Dako Colorado, Inc. (Fort Collins, CO) MoFlo Flow Cytometer and High Speed Cell Sorter. GFP fluorescent signal was excited with a solid state iCyt 488 nm blue laser. Cerulean fluorescent protein was excited with a Coherent Radius 405 nm violet laser.

<div align="center">

Results
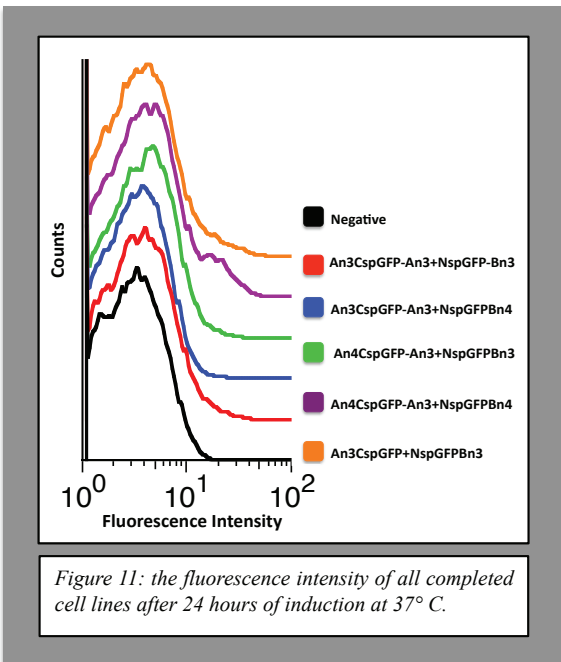
</div>

Experimental Design

To determine if split superpositive GFP can be used in a competition assay for inhibitors of protein-protein interactions, we devised and performed a competitive inhibition assay using combinations of An3, An4, Bn3 and Bn4 proteins. In this naming

system, A and B refer to the two different chains of the heterodimer, n refers to the presence of asparagine in the dimer interface portion of each heptad repeat, and 3 and 4 refers to the number of heptad repeats in each chain of the heterodimer.

Cells were coexpressed with an An heterodimer conjugated to the C terminal fragment of superpositive GFP as the bait protein, a Bn heterodimer conjugated to the N terminal fragment of superpositive GFP as the trap protein, and an An heterodimer used as the competitive inhibitor (refer to materials and methods). Cell fluorescence was evaluated using flow cytometry at variable time points after induction. Cell lines containing inhibitor heterodimers were compared against a non-recombinant, baseline fluorescent *E. coli* strain as a negative control and a cell line containing the same bait/trap heterodimer pair absent the inhibitor as a positive control.
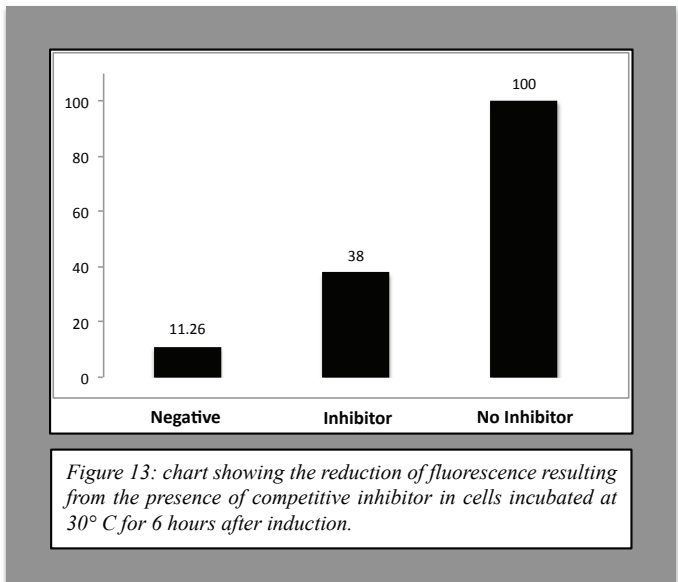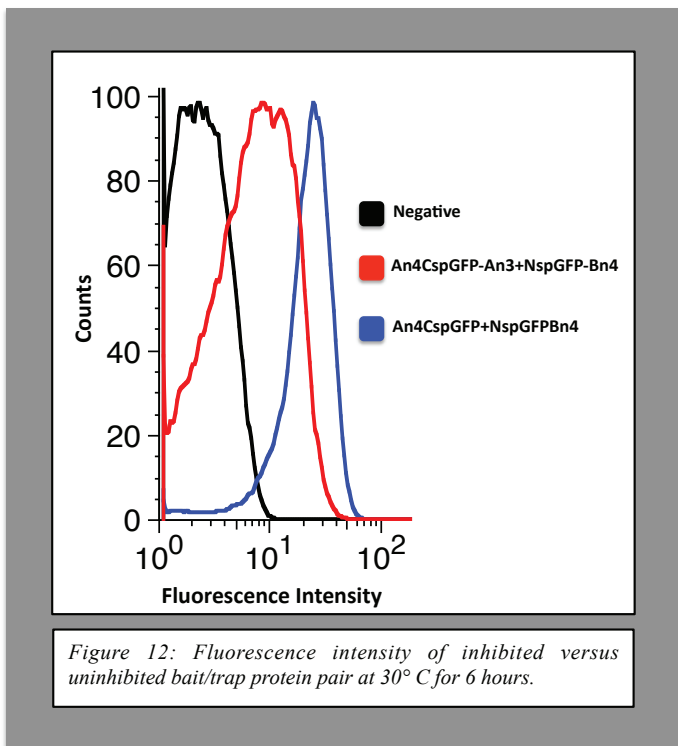
A second series of experiments was run using Pdar and Prb native proteins as the bait and trap protein pair, and Prb as inhibitor. These were coexpressed in a similar manner to the leucine zipper series, with either native Prb or mutant Prb D83N conjugated to CspGFP acting as the bait protein motif, either native Prb or mutant Prb D83N acting as an inhibitor of the bait/trap protein-protein interaction, and either native Pdar or mutant PdarN34D conjugated to NspGFP acting as the trap protein motif (see Materials and Methods).

Figure 11: the fluorescence intensity of all completed cell lines after 24 hours of induction at 37° C.

Experimental set one: Helices evaluated after 24 hours of induction

To test the limits of our assay, we first performed flow cytometry after 24 hours of induction (see Materials and Methods). At this time point, no significant differences between cell lines containing inhibitor and cell lines absent of inhibitor were found (see Fig 11). In an effort to thoroughly determine whether the results after 24 hours of induction were useful, several different parameters were evaluated for correlation to Δ Bait/Trap K_D minus Inhibitor/Trap K_D. First, mean fluorescence intensity of whole populations was evaluated. Next, a negative control cell line was used to gate fluorescence intensity that was occurring only above baseline *E. coli* fluorescence. These gated populations were evaluated for differences by mean fluorescence intensity above the gate and % of total cell population above the gate. None of these parameters demonstrated a statistical difference between the different

Figure 12: Fluorescence intensity of inhibited versus uninhibited bait/trap protein pair at 30° C for 6 hours.

Figure 13: chart showing the reduction of fluorescence resulting from the presence of competitive inhibitor in cells incubated at 30° C for 6 hours after induction.

18

cell lines.

Experimental set one: Helices evaluated after 6 hours of induction

Experimental set one refers to the first series of flow cytometry experiments performed using a mutant An4CspGFP protein. We evaluated three cell lines at a 6-hour
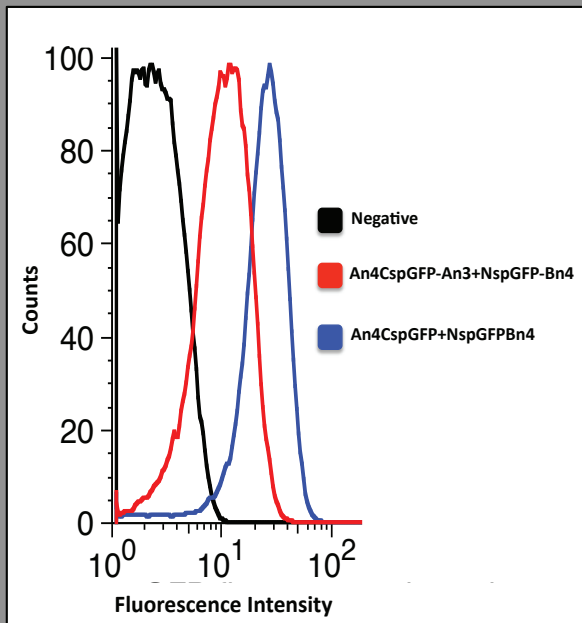

Figure 14: Fluorescence intensity of inhibited versus uninhibited bait/trap protein pair at 25° C for 6 hours.
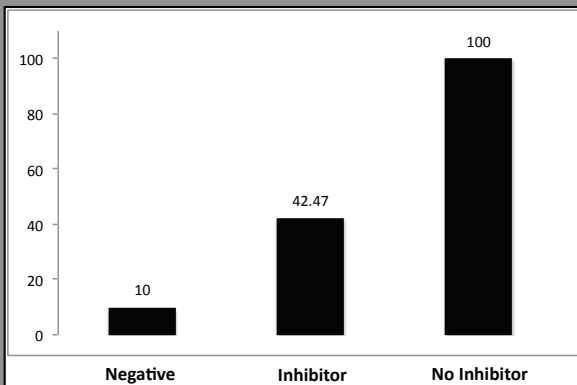
time point in the hopes that this time point would meet the criteria required for an appropriate time dependent recovery of fluorescence (refer to reporter proteins section of introduction). The first cell line was a positive control containing An4CspGFP and Bn4CspGFP. The second cell line contained An4CspGFP and Bn4CspGFP as well as a free An3 as an inhibitor. The final cell line was a negative control to provide baseline *E. coli* fluorescence for comparison. Cell lines were cultured at 25° C and 30° C after induction. Results of population fluorescence showed a decrease in recovered fluorescent signal in cultures containing

Figure 15: chart showing the reduction of fluorescence resulting from the presence of competitive inhibitor in cells incubated at 25° C for 6 hours after induction.

19

inhibitor protein (refer to figures 12 and 14). At 25° C, the culture containing the inhibitor experienced a 57.53% reduction in mean fluorescence intensity compared to the positive control (refer to figure 15). At 30° C, the culture containing the inhibitor experienced a 62% reduction in mean fluorescence intensity compared to the positive control (refer to figure 13).

Experimental set two: Helices evaluated after 6 hours of induction

Experimental set two refers to the series of flow cytometry experiments performed with a complete helical protein set where the mutation encountered in the An4CspGFP plasmid had been identified and corrected. The bacteria populations containing the leucine zipper heterodimer protein set were evaluated after 6 hours of induction based on the results of the previous section, which indicated that 6 hours of induction provided a more robust fluorescent signal than was obtained after 9 hours or 24 hours, which were the two other time points tested.

To evaluate the effectiveness of the inhibitor assay, the mean fluorescence of the different bacterial populations containing a distinct protein set were each plotted against the bait-trap K_D of the protein set minus the inhibitor-trap K_D (refer to figure S1 and S3 and table S1 and S2). Although these values do not correspond to any biologically or chemically real parameter, they provide a way to compare results based on how inhibited the reassembly of split-superpositive GFP is.

The results from this experiment unfortunately do not follow the trend that would be expected if this assay were to be useful in a rapid screening of potential inhibitors. Both the populations incubated at 25° C and 30° C after induction did not show a significant trend toward reduction of fluorescence with increasing inhibitor affinity.

The set evaluated with 25° C incubation temperature shows some indication of the desired trend, but an R^2 regression analysis for goodness of fit calculated in Excel provided a value of .455, which is well below an acceptable value for a biological model (refer to figure S2 and table S1).
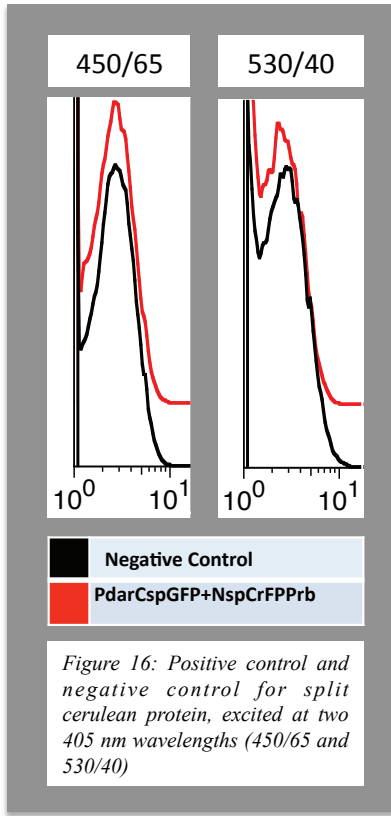
The set evaluated with 30° C incubation temperature, the slight trend shown in the 25° C set is not present (refer to figure S3 and table S2). In fact, the trend observed shows increasing fluorescence with inhibitor protein present. This can likely be explained by the tendency of split-superpositive GFP signal to become more erratic in cells incubated at higher temperature. One potential explanation for this phenomenon could be the lack of eukaryotic chaperone proteins in BL21 *E. coli.* Without the benefit of chaperone proteins, protein folding in *E. coli* tends to improve at lower temperatures, which improve the thermodynamics of protein folding.

Native proteins evaluated after 6 hours of induction

The results from the native protein set showed a potentially more useful trend than that of the leucine zipper heterodimer set. Despite a problem encountered with the positive control in this assay, there was a discernible difference in fluorescence signal between cell populations based on the affinity of the inhibitor protein that was expressed. (Refer to figures S5, S7, S8 and S10)

Additionally, both the cell populations incubated at 25° C and the populations incubated at 30° C exhibited the same trend in fluorescence. The cell populations incubated at 25° C after induction expressing the low affinity inhibitor (Prb) experienced 6.05 times the mean fluorescence intensity of populations containing high affinity inhibitor (PrbD83N) (Refer to figure S6). In cell populations incubated at 30° C after

21

induction, the strains expressing the low affinity inhibitor experienced 7.13 times the mean fluorescence intensity of strains expressing the high affinity inhibitor (refer to figure S9).

Orthogonally reporting cerulean fluorescent protein assay

The positive control for the orthogonal reporting assay (PdarCspGFP + NspGFPPrb without inhibitor) was excited at 405 nm for cerulean fluorescence. The sample containing the positive control did not exhibit fluorescence activity above that of the negative control (Refer to figure 16).

Figure 16: Positive control and negative control for split cerulean protein, excited at two 405 nm wavelengths (450/65 and 530/40)

Discussion

Goals of this project

The stated goal of this project was to evaluate a new system using the split superpositive GFP and orthogonal split cerulean fluorescent protein PCA as a tool to evaluate the ability of candidate inhibitor proteins to disrupt the activity of protein-protein interactions *in vivo* in a high throughput manner. This project has potential as a tool for scientists in the development of inhibitor proteins for therapeutic purposes. This system is meant mainly as a first step to identify candidate proteins as potential inhibitors, which could be better characterized using more quantitative methods after they are identified. Following these goals, it should be kept in mind that the results discussed below are not fully quantitative, nor are they characteristic evaluations of the interactions

involved. The purpose of this project is only to evaluate relative results based on recovery of fluorescence related to known high affinity and low affinity inhibitor reduction of the interactions of bait and trap proteins.

Cloning Problems

As is often the case with experimental research, problems can arise that interfere with the pursuit of valid results. Over the course of this project, a major problem that we experienced was misinterpretation of sequencing results. This manifested itself, for example, during the sequential process of cloning the plasmid constructs in preparation for the actual experiment. In this case, contamination by a pET plasmid without a second multiple cloning site (pET11a) led to an inability to introduce the NspGFP fragment conjugated leucine zipper heterodimer PCR constructs into the plasmid.

Additionally, after evaluating the results discussed in the experimental set one, it was discovered that the experiments were performed using incorrectly cloned plasmids due to misinterpretation of sequencing information. After the incorrectly cloned An4CspGFP plasmids were identified as containing a non-conservative single amino acid residue mutation (a mutation in which an amino acid residue mutation results in a change in the chemical properties between the original and the mutant residue), the helix protein set was completed, and the plasmid containing the non-conservative mutation was corrected by site directed mutagenesis. The results from these experiments were still included in this body of work because they were still important in the process of understanding and identifying the problems that had occurred, and contextualize later results achieved with properly sequenced constructs. The results in experimental set two come from recloned plasmids that contain the correct sequence.

It was discouraging to observe such poor fluorescent recovery in the population containing the positive control for the Pdar/Prb protein set. However, it should be noted that these results are most likely caused by a cloning error and probably do not reflect the values that would be obtained using a correctly cloned positive control plasmid.

Based on previous results obtained by other members of the McNaughton lab, which showed high fluorescent signal obtained with the same protein set that was used for the positive control in this experiment, a possible explanation for low fluorescence would be mutation of the promoter region of the plasmid used to express the PdarN34DCspGFP protein without inhibitor. If a damaged promoter resulted in lower expression levels of that protein, then its plausible that the fluorescence could be decreased as a result.

Summary of results with split superpositive GFP competitive inhibition

The results found using flow cytometry at 6 hours after induction show promising evidence of a sound proof of concept that split superpositive GFP can be used to detect competitive inhibition of protein-protein interactions. The data collected at 24 hours after induction supports the model of rate determinate saturation from fluorescence recovery on split superpositive GFP reassembly for demonstrating differences in fluorescence intensity based on interactions of conjugate interacting proteins. Specifically, it is likely that the cell populations show no statistical differences in mean fluorescence intensity because the reassembly of split-superpositive GFP has equilibrated, and saturation of the cells in the population has occurred, preventing an observable difference in the fluorescent signal of the cell lines. Because the formation of actively fluorescent NspGFP is functionally irreversible, recovery of fluorescence within a cell can be thought of as a

pseudo K_r problem. The hypothesis is that although observing the fluorescence recovery within a cell expressing a bait/trap interacting protein pair shows the results of N terminal fragment GFP fluorescence recovery accumulating over time, the system should be able to provide insight into the K_D of the interacting proteins conjugated to the two fragments of split superpositive GFP because the affinity of the proteins affects the kinetics of reassembly of the two split superpositive GFP fragments. Because the presence of a competitive inhibitor changes the effective K_D of the interacting proteins by lowering the concentration of substrate available for binding, it is reasonable to expect that the kinetics of reassembly of split superpositive GFP will decrease in the presence of an inhibitor.

The small size of the helical proteins used may have contributed to the problems that were experienced observing a trend following decrease of fluorescence recovery upon inhibition of bait/trap binding. The designed leucine zipper heterodimers used as bait/trap proteins in this assay were desirable for an initial proof of concept because of their simplicity and predictability of form (provided by a standard repeating motif and single secondary structure included in the protein), but this assay is only useful for application to questions involving naturally occurring protein-protein interactions if the assay shows similar results for proteins with much more complicated tertiary structures. It is possible that the results obtained with mutant An4CspGFP followed a trend of reduction of fluorescence upon inhibition due to disruption of the dimer interface by the mutation, and not as a result of the inhibitor assay functioning appropriately.

Fortunately, the experiments that were performed using Pdar and Prb did show the potential for this assay design to be used as a rapidly cloneable, high throughput method of identification of candidate inhibitor proteins. We are confident that if this

experiment were repeated using a recloned positive control, the results would continue the trend observed in the experiments we performed, with tiers of increasing fluorescence intensity following decreased affinity of inhibitor protein present.

The trend observed in differences in fluorescence between the different cell populations in the native protein split-superpositive GFP assay deviated from the expected results. We expected to find a trend correlating the difference between the K_D of the bait-trap pair and the K_D of the inhibitor-trap pair to the recovery of fluorescence within the cell populations. Instead, we found that fluorescence decreased independent of the affinity of the bait-trap pair, and was related only to the affinity of the inhibitor protein. Although these results were unexpected, they do not contraindicate the usefulness of this technique in fulfilling the stated goal of identifying candidate inhibitor proteins.

Results using split cerulean protein as an orthogonal reporter to the split superpositive GFP protein complementation assay

The evaluation of split-cerulean fluorescent protein for orthogonal reporting capabilities within the split-superpositive GFP system failed to return results demonstrating any detectable cerulean signal, even with the positive control. Because the positive control for this assay uses the same PdarD34NCspGFP-MCS2 pETDuet plasmid as the positive control used in the native protein split-superpositive GFP assay, its possible that these results are caused by expression problems resulting from damage in this plasmid that is not detectable in sequencing information of the multiple cloning site containing the protein sequence. Cerulean fluorescent signal was also entirely absent

from experimental strains containing a PrbNspGFP motif as an inhibitor protein in MCS2 of the pETDuet plasmid.

There are two possible explanations for the lack of cerulean signal in the experimental populations expressing a PdarCspGFP bait protein motif, PrbNspGFP trap/inhibitor protein motif, and PrbNCrFP trap protein motif within the same system. The first possible explanation is that since the nature of the experiment as supporting evidence in another project required co-expression of the bait protein and inhibitor protein motifs resulted in the immediate reassembly of CspGFP and NspGFP, leaving no free CspGFP protein to reassemble with NCrFP, which was expressed off of a separate plasmid. The second possible explanation is that there is a problem with the reassembly of CspGFP and NCrFP that results in a lack of cerulean signal when the two proteins are expressed in the same system. Due to the suspected damage to the plasmid used in the positive control, the first possibility cannot be ruled out.

The ability for NCrFP to act as an orthogonal reporter to the split-superpositive GFP protein complementation assay could be better evaluated with a system in which the CspGFP motif protein is not coexpressed with an N terminal fragment motif protein with which it can reassemble. For example, a system in which the competing bait and N terminal fragment motif proteins are coexpressed on a plasmid with two multiple cloning sites, while the CspGFP trap protein motif is expressed on a separate plasmid might be able to provide more insight into the orthogonal reporting capabilities of a split cerulean fluorescent protein/split GFP protein complementation assay.

Relevance

The eventual expectation of the split-superpositive GFP inhibitor-screening assay is that it could be applied to a question of actual biological relevance. For example, the methods outlined in this project could be applied to protein-protein interactions that occur erroneously in a manner that leads to disease. Gankyrin is an ankyrin repeat protein very similar in tertiary structure to Pdar. Gankyrins native ligand is the S6 subunit of the 26S regulatory particle of the proteasome. Gankyrin aids S6 in ubiquitin ligation targeting of the tumor suppressor proteins P53 and Rb for proteasomal degradation. Rb and P53 regulate the cell cycle and initiate DNA repair when mutations or DNA damage occur.

When gankyrin is overexpressed, P53 and Rb levels are lowered to the point where they can no longer prevent cells with DNA damage from exiting the G1/S phase of the cell cycle or initiate the repair of the damaged DNA. As a result cancerous cell lines survive and are capable of proliferation which leads to the formation of tumors. A gankyrin binding protein was recently developed by saturation mutagenesis of the protein Prb and shown using split superpositive GFP PCA to bind gankyrin with higher affinity than wild type Prb.

If the methods developed in this project were applied to gankyrin/S6 as a natively interacting protein pair and mutated gankyrin as an inhibitor, it could be determined whether or not the mutated Prb not only binds gankyrin, but also binds tightly enough to displace S6 and prevent the harmful effects of degradation of the P53 and Rb tumor suppressor proteins.

If the mutated Prb protein does bind tightly enough to gankyrin to displace S6, then the mutated protein could potentially be used as a therapeutic agent in hepatocellular

carcinoma cases where gankyrin is overexpressed, leading to tumor formation. Many

other diseases also result from the overexpression or mutation of proteins used in

regulatory protein-protein interactions. The process of developing inhibitors that can

target these proteins will be made much easier by the existence of a technique to rapidly

and accurately evaluate a designed macromolecular inhibitors ability to displace the
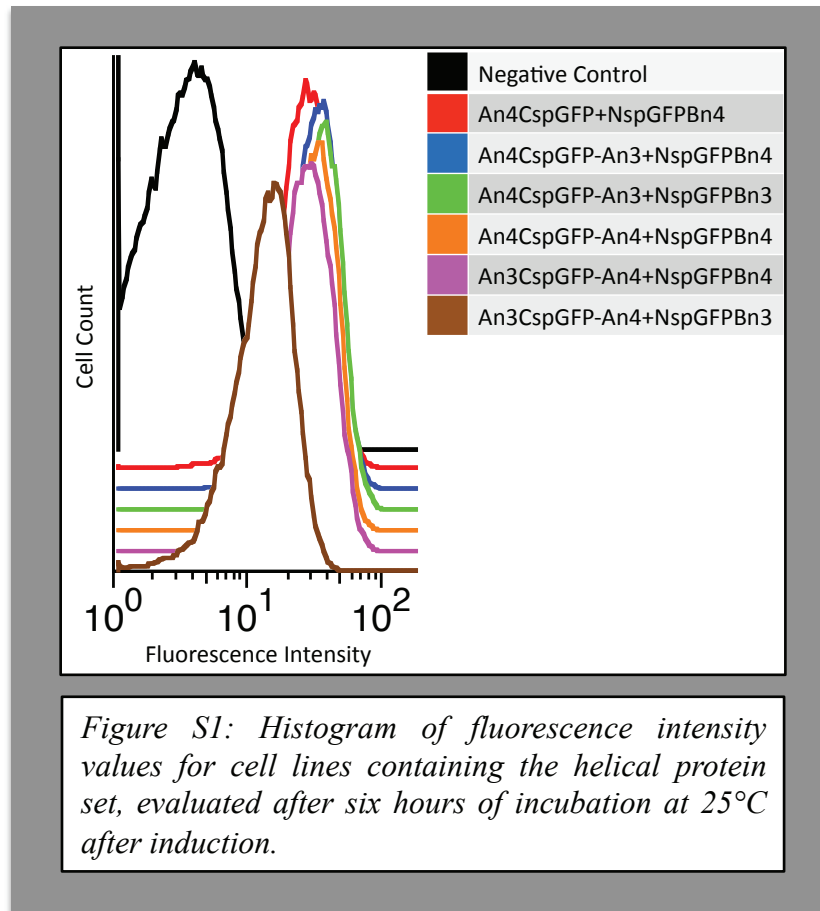
native substrate of the target protein.

Resources:

1) Karanicolas, J.; Corn, J. E.; Chen, I.; Joachimiak, L. A.; Dym, O.; Peck, S. H.; Albeck, S.; Unger, T.; Hu, W.; Liu, G.; Delbecq, S.; Montelione, G. T.; Spiegel, C. P.; Liu, D. R.; Baker, D. A *de Novo* Protein Binding Pair By Computational Design and Directed Evolution. *Molec. Cell* **2011,** *42* (2), 250-260.
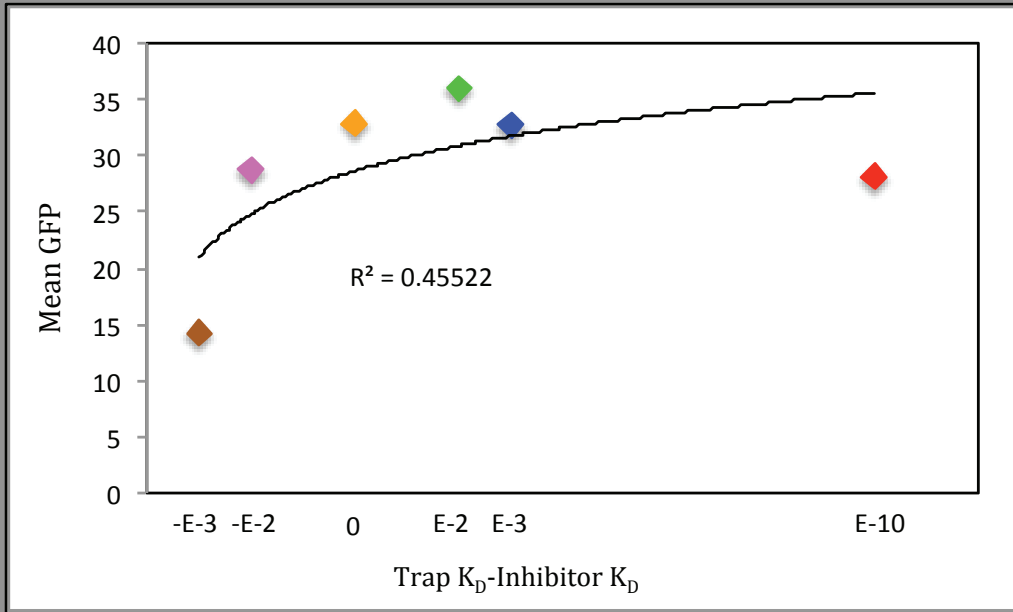
2) Thomas, F.; Boyle, A. L.; Burton, A. J.; Woolfson, D. N. A Set of *de Novo* Designed Parallel Heterodimeric Coiled Coils with Quantified Dissociation Constants in the Micromolar to Sub-nanomolar Regime. *J. Am. Chem. Soc.* **2013,** *135*, (13), 5161-5166.

3) Morell, M.; Ventura, S.; Avilés, F. X. Protein complementation assays: approaches for the in vivo analysis of protein interactions. *FEBS Lett.* **2009,** *583,* (11), 1684-1691.

4) Blakely, B. D.; Chapman, A. M.; McNaughton, B. R. Split-superpositive GFP Reassembly is a Fast, Efficient, and Robust Method for Detecting Protein-Protein Interactions *in Vivo. Mol. BioSyst.* **2012**, *8,* 2036-2040.
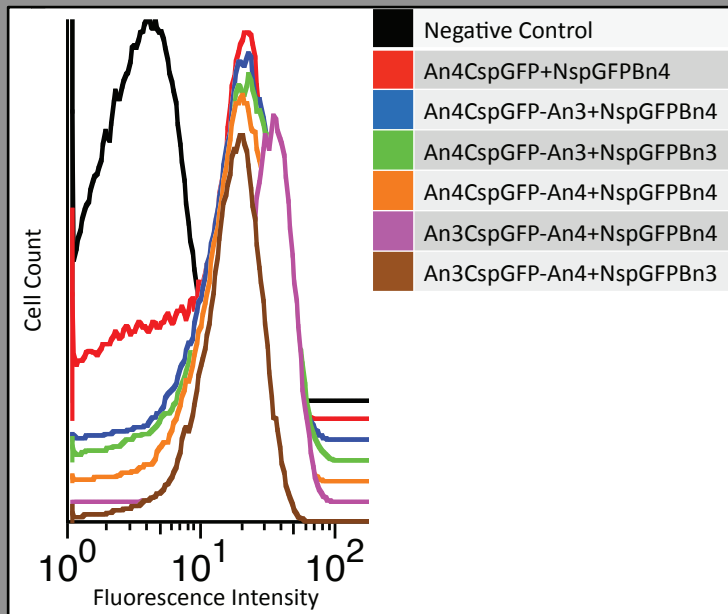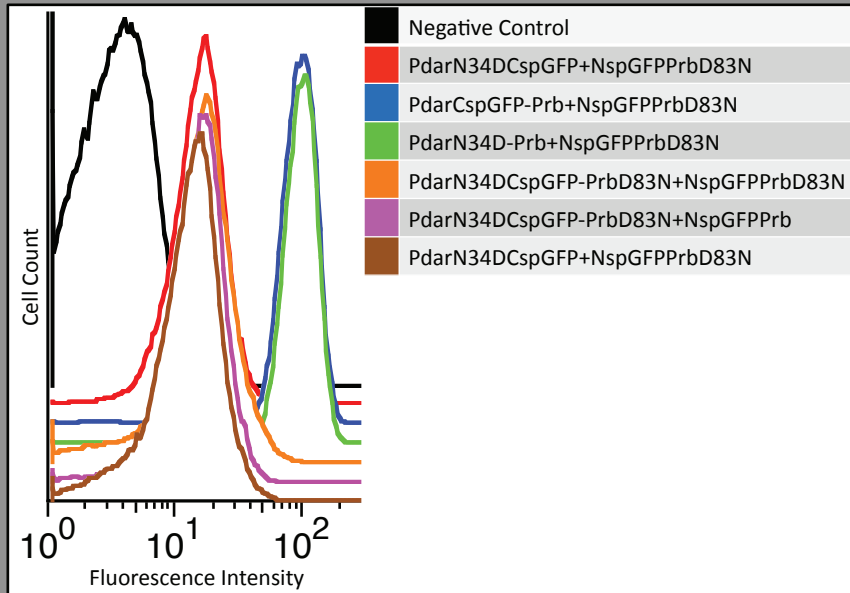
Figure S1: Histogram of fluorescence intensity values for cell lines containing the helical protein set, evaluated after six hours of incubation at 25°C after induction.

CspGFP	Inhibitor	NspGFP	Trap K_D-Inhibitor K_D	Mean Fluorescence
An4	N/A	Bn4	E-10	28.1
An4	An3	Bn4	E-3	32.7
An4	An3	Bn3	E-2	36
An4	An4	Bn4	0	32.9
An3	An4	Bn4	-E-2	28.7
An3	An4	Bn3	-E-3	14.2

Figure S2: Helices incubated for six hours at 25°C after induction. Mean fluorescence of each sample is plotted against traptrap KD-inhibitor KD. Table S1: Protein set for each cell line and the corresponding trap KD-inhibitor KD and mean population fluorescence values.

Figure S3: Histogram of fluorescence intensity values for cell lines containing the helical protein set, evaluated after six hours of incubation at 30°C after induction.

R² = 0.26533
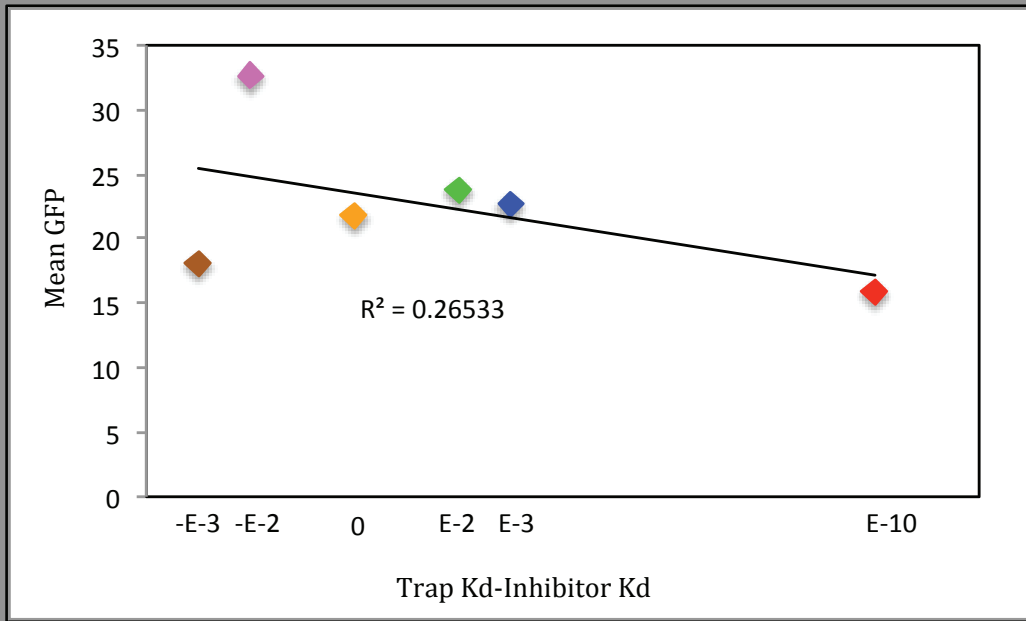
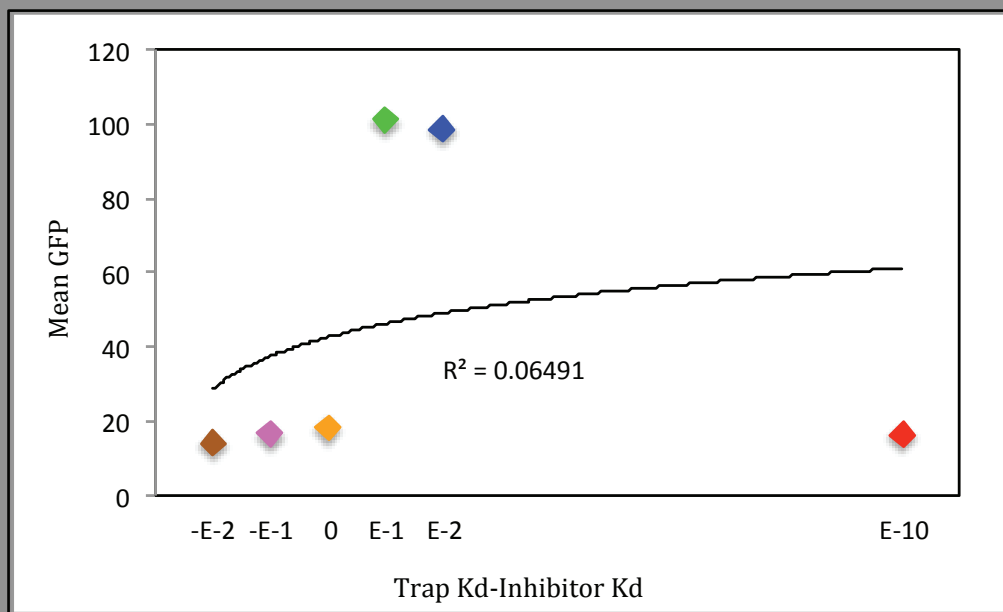| CspGFP	Inhibitor	NspGFP	Trap K_D-Inhibitor K_D	Mean Fluorescence
An4	N/A	Bn4	E-10	15.9
An4	An3	Bn4	E-3	22.7
An4	An3	Bn3	E-2	23.8
An4	An4	Bn4	0	21.8
An3	An4	Bn4	-E-2	32.6
An3	An4	Bn3	-E-3	18.1

Figure S4: Helices incubated for six hours at 30°C after induction. Mean fluorescence of each sample is plotted against trap KD-inhibitor KD.
Table S2: Protein set for each cell line and the corresponding trap KD-inhibitor KD and mean fluorescence of that population.

Figure S5: Histogram of fluorescence intensity values for cell lines containing the native protein set, evaluated after six hours of incubation at 25°C after induction.
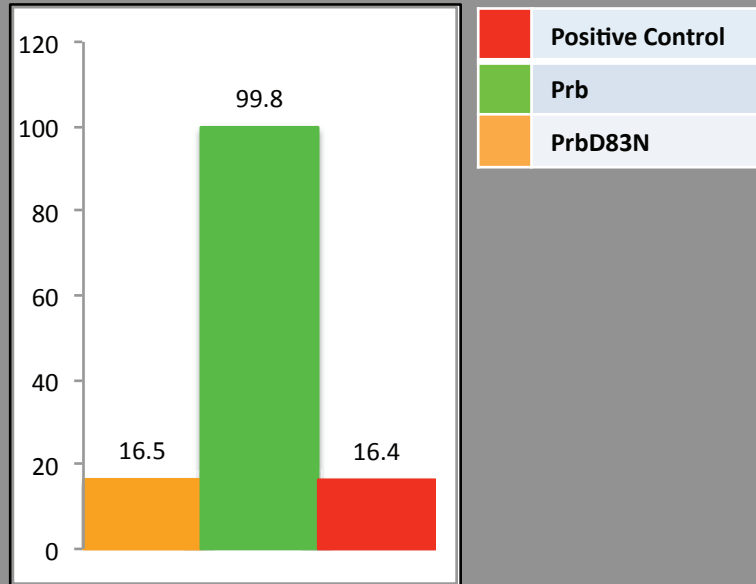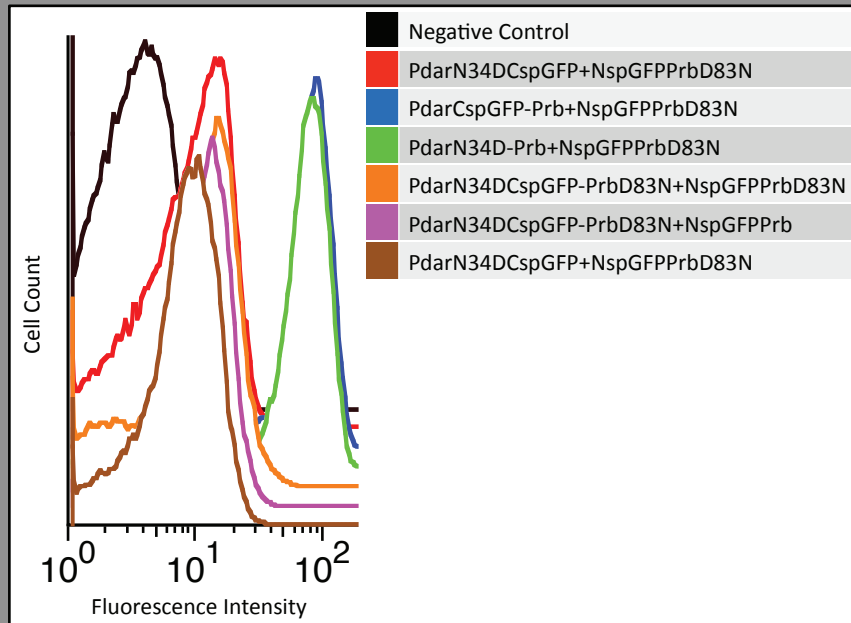
Figure S6: Relative mean fluorescence values of native protein populations incubated for six hours at 25°C by inhibitor molecule present.
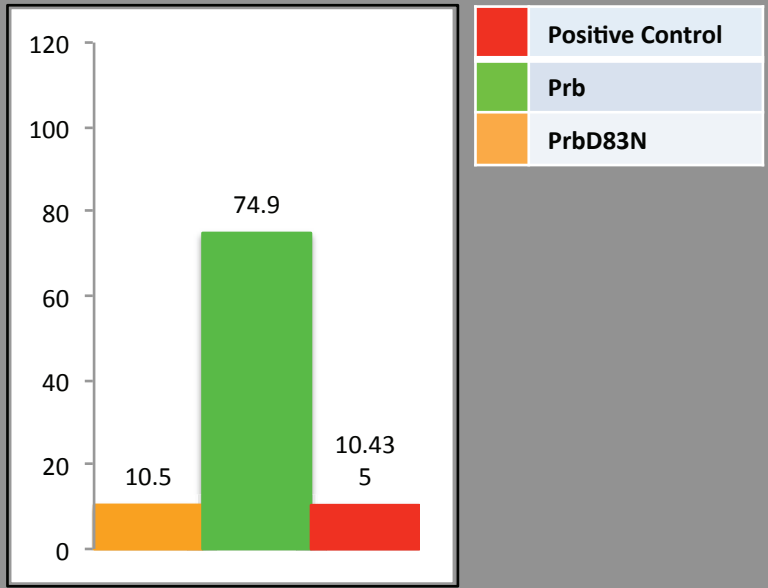
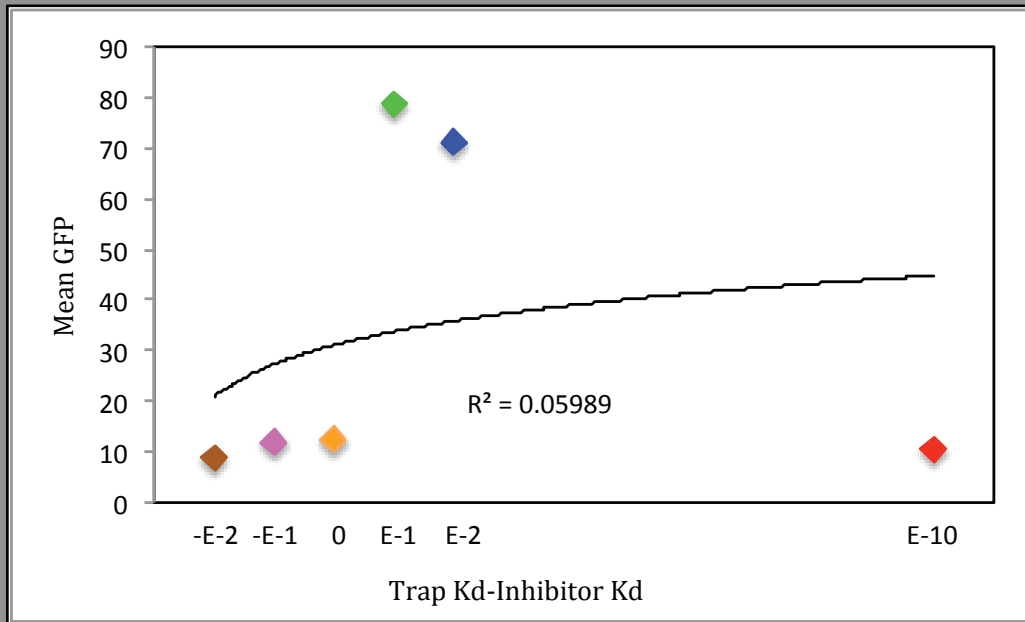| CspGFP	Inhibitor	NspGFP	Trap K_D-Inhibitor K_D	Mean Fluorescence
PdarN34D	N/A	PrbD83N	E-10	16.5
Pdar	Prb	PrbD83N	E-2	98.6
PdarN34D	Prb	PrbD83N	E-1	101
PdarN34D	PrbD83N	PrbD83N	0	18.4
PdarN34D	PrbD83N	Prb	-E-1	16.6
Pdar	PrbD83N	Prb	-E-2	14.2

Figure S7: Native proteins incubated for six hours at 25°C after induction. Mean fluorescence of each sample is plotted against trap KD-inhibitor KD. Table S3: Protein set for each cell line and the corresponding trap KD-inhibitor KD and mean fluorescence of that population.

Figure S8: Histogram of fluorescence intensity values for cell lines containing the native protein set, evaluated after six hours of incubation at 30°C after induction.

Figure S9: Relative mean fluorescence values of native protein populations incubated for six hours at 25°C by inhibitor molecule present.

CspGFP	Inhibitor	NspGFP	Trap K_D-Inhibitor K_D	Mean Fluorescence
PdarN34D	N/A	PrbD83N	E-10	10.5
Pdar	Prb	PrbD83N	E-2	71.1
PdarN34D	Prb	PrbD83N	E-1	78.7
PdarN34D	PrbD83N	PrbD83N	0	12.4
PdarN34D	PrbD83N	Prb	-E-1	12
Pdar	PrbD83N	Prb	-E-2	8.87

Figure S10: Native proteins incubated for six hours at 30°C after induction. Mean fluorescence of each sample is plotted against trap KD-inhibitor KD. Table S4: Protein set for each cell line and the corresponding trap KD-inhibitor KD and mean fluorescence of that population.