University of Ottawa
Chemistry and Biomolecular Sciences
Faculty of Science

Université d'Ottawa
Chimie et Sciences Biomoléculaires
Faculté de Sciences

A GFP-Based Sensor to Detect Transiently Expressed Proteins.

Matthew Eason

A thesis submitted in partial fulfillment of the requirements for the

Doctorate in Philosophy degree in Chemistry

Abstract

Green fluorescent protein (GFP) fusion tags are commonly used to study protein expression and cellular localization *in vivo*. But, GFP must undergo an autogenic post-translational modification, known as chromophore maturation, to become fluorescent, a process that can have a half-time longer than 30 minutes inside research model organisms. The timescale of chromophore maturation in GFP is thus slower than many key biological processes, limiting its usefulness in measuring those processes. In this thesis, we discuss the creation and engineering of a sensor for transiently expressed proteins (STEP) based on a fully matured but dim GFP. Upon specific binding of STEPtag, a small (15.5 kDa) protein to the sensor, full fluorescence is restored. Thus, by genetically fusing STEPtag to a protein of interest, it can be detected as soon as folding is complete, without any maturation delay. Through a combination of rational design and targeted directed evolution, we describe the improvement of the original sensor, gSTEP0, into an optimized version, gSTEP1. The sensor has been validated *in vitro* and in *E. coli* cells, and we have found that for gSTEP1, the fluorescence signal increases more than three-fold upon binding, with a K_d of 120 ± 30 nM and a k_{on} of 1.7 x 10^5 $M^{-1}s^{-1}$, allowing detection of the protein of interest on the second timescale. We have also created a yellow version of the biosensor, and provide preliminary attempts at developing orthogonal binding pairs, as well as red- and cyan-coloured STEPs, which could eventually be used in multiplex experiments. Our biosensor opens the door to the study of short-timescale processes in research model organisms, such as *Drosophila* and zebrafish embryogenesis, as well as in host-pathogen interactions, which we are currently investigating.

Acknowledgments

I would like to thank a number of people for their contributions that have allowed me to complete this work. For supporting the research, the researcher, or both, you all have my gratitude.

First and foremost, to my lovely wife, Emily Eason. You have been by my side throughout this entire process, and have never failed to support me while I chose the life of a starving student. Thank you for always believing in me, and for being a partner who both understands and completes me. I love you more than words can ever truly say.

To my partner in crime, Adam Damry, with whom I have spent so much time discussing ideas, problem solving, and simply joking around. I have always valued your insight and advice, and appreciate your willingness to lend a hand, even from the other side of the world. I wish you the best in everything you do, and I am confident that you will get it.

To my parents, Gord and Lynn Dee Eason, who have fostered curiosity, independence and a love of science and academics in me throughout my life. Thank you for providing me with all the warmth of a small-town upbringing, while ensuring I had all the opportunities for success that I needed.

To my colleagues on this project, Antonia Morris (née Pandelieva), Marc Mayer, Safwat Khan and Eduardo Ramirez, thank you so much for all of your assistance. I appreciate each of you lending your time and effort to this project, without which so many fascinating experiments would have been left undone.

To the number of highly skilled Lab Managers and Technicians who have provided me with training and expertise as I worked on this project, Andrew Ochalski, Dr. Christopher Clouthier, Dr. Shahrokh Ghobadloo and Dr. Sabina Sarvan. Thank you for all for being friendly and helpful, and for always making yourselves available. Your technical skills were invaluable in performing the experiments presented in this work.

To all the members of the Chica lab, past and present, thank you for always providing a lab environment that is both friendly and intellectually stimulating. It has been a pleasure to work alongside so many intelligent people, and I appreciate the time that I have spent with all of you.

And finally, to my supervisor, Dr. Roberto A. Chica, I thank you for your advice and direction through all the years that we have worked together. I appreciate that you have always fostered not just good science, but good scientists, by valuing communication and networking skills while still pushing for top-tier research. Looking back at the day that you walked me through making competent cells for hours longer than it should have taken, I see just how far I have come, and I strive to ensure that my work will be worthy of the coveted “C” for quality.

Table of Contents

List of Tables

List of Figures

List of Equations

List of Abbreviations

ANS – 1-Anilinonaphthalene-8-Sulfonate

BH3 – Bcl-2 Homology domain 3

BSA – Bovine Serum Albumin

cpGFP – Circularly Permuted Green Fluorescent Protein

dPSTR – Dynamic Protein Synthesis Translocation Reporter

EGTA – Ethylene Glycol-bis(β-aminoethyl ether)-N,N,N',N'-Tetraacetic Acid

FACS – Fluorescence-Activated Cell Sorting

FISH – Fluorescence *In Situ* Hybridization

FP – Fluorescent Protein

FRET – Förster Resonance Energy Transfer

GECI – Genetically Encoded Calcium Ion Indicator

GFP – Green Fluorescent Protein

IAA – 3-β-Indoleacrylic Acid

IPTG – Isopropyl β-D-1-Thiogalactopyranoside

ITC – Isothermal Titration Calorimetry

K_d – Dissociation Equilibrium Constant

LB – Luria-Bertani

MBP – Maltose-Binding Protein

MCS – Multiple Cloning Site

mRNA – Messenger Ribonucleic Acid

PMT – Photomultiplier Tube

POI – Protein of Interest

SDS-PAGE – Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

STEP – Sensor for Transiently Expressed Proteins

TAX – *Thermoascus aurantiacus* Xylanase 10A

TB – Terrific Broth

TIM – Triosephosphate Isomerase

Chapter 1: Introduction

1.1 Fluorescent Proteins

1.1.1 Fluorescence

Fluorescence is defined as the emission of a photon from a molecule, known as a fluorophore, caused by the relaxation of an excited singlet electronic state (S_1) back to the ground state of the molecule (S_0).[13] In order to reach the higher-energy excited state, the molecule absorbs a photon with sufficient energy to reach one of the vibrational states of the S_1 electronic state. Thus, a fluorophore will have both an excitation spectrum, or range of energies that it is capable of absorbing that will lead to fluorescence, as well as an emission spectrum, or range of energies at which it will emit. When considering fluorescence, these energies are often expressed as wavelengths, which are inversely proportional to the energy of the photon. The emission spectrum is typically at a longer wavelength (lower energy) than the excitation spectrum due to some conversion of energy into non-radiative processes. This phenomenon is known as the Stokes shift,[14] and is commonly caused by the rapid decay to the lowest vibrational level of the excited state, as well as the decay to higher vibrational levels of S_0, leading to excess vibrational energy being lost as heat (Figure 1.1).

The efficiency of a fluorophore is known as the brightness, and is defined as the product of the extinction coefficient (or molar absorption coefficient, ε) and the quantum yield (Φ).[15] The extinction coefficient represents the capacity of the fluorophore to absorb light of a given wavelength, and is defined as:

$$\varepsilon = \frac{A}{cl} \qquad \text{(Equation 1)}$$

Where A is the absorbance of a solution of the fluorophore at a given wavelength, c is its concentration, and l is the optical path length. This relationship is known as the Beer-Lambert law. The quantum yield

represents the probability that an excited fluorophore will emit a photon as fluorescence, and it is defined as:

$$\Phi = \frac{\#\ photons\ emitted}{\#\ photons\ absorbed} \qquad \text{(Equation 2)}$$

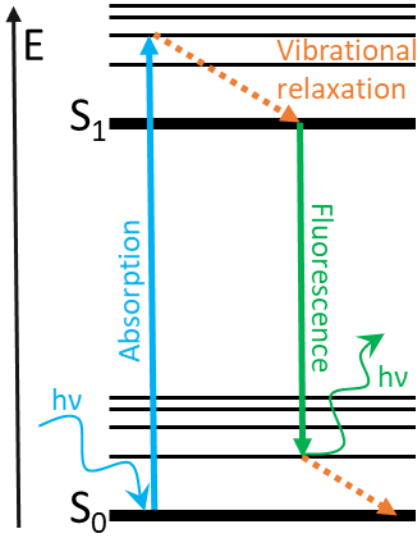


Figure 1.1 Fluorescence. A Jablonski diagram of a simple case of fluorescence. A photon of blue light is absorbed by a fluorophore, exciting an electron from the ground electronic state (S_0) to the first excited state (S_1). The excited electron rapidly decays to the lowest vibrational state of the excited state, from which it relaxes to the ground electronic state, accompanied by the emission of a (lower-energy) green photon. Adapted from Hochreiter *et al.*[6] under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

The quantum yield depends on the relative rates of radiative and non-radiative decay of the fluorophore, with a quantum yield close to unity implying that the non-radiative decay rates of the fluorophore are slow relative to the rate of fluorescence emission.[13] Thus, the fluorescence brightness of a molecule can be diminished, or quenched, by the presence of alternative relaxation pathways, or it can be increased by shielding it from these same pathways. These non-radiative decay pathways include internal conversion, where the excess electronic energy is converted to excess vibrational or rotational energy before being lost as heat, intersystem crossing where the excited singlet-state converts to a triplet state that slowly decays as phosphorescence, or quenching by transiently colliding or complexing with particular molecules, such as oxygen or halogens. Of particular relevance in

biochemistry, the presence of water, a polar, protic solvent, provides a number of possible mechanisms to quench fluorescence. As the excited state of the fluorophore typically has a larger dipole moment than the ground state, the solvent dipoles may reorient around it after excitation, lowering the energy of the excited state and therefore reducing the energy gap between the ground and excited state. This can red-shift the emission wavelength away from the detected range, increase the probability of internal conversion,[16] or stabilize the triplet form of the excited state, leading to intersystem crossing and the loss of fluorescence, as phosphorescence occurs on a much slower timescale.[17] Alternatively, hydrogen bonds between water and the fluorophore can provide pathways for charge transfer between the excited state and the solvent, which can also provide non-radiative decay pathways back to the ground state.[18]

The quenching effect of water can potentially be reversed by shielding the chromophore from solvent, for instance by binding it to a biomacromolecule. This is the case for a number of protein-binding dyes, such as 1-anilinonaphthalene-8-sulfonate (ANS) and Congo Red, among others, which increase in fluorescence upon non-covalent binding to proteins, and can therefore be used as markers to probe their structure and aid in their detection, for instance in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).[19] Because fluorescence spectroscopy is both sensitive and non-invasive, it is an excellent tool for studying proteins.

1.1.2 Fluorescent Protein Structure and Chromophore Maturation

The green fluorescent protein (GFP) was originally identified in *Aequorea victoria* by Shimomura *et al.*,[20, 21] and has the uncommon property of being an intrinsically fluorescent protein, i.e. it emits a fluorescent signal in the visible light spectrum without requiring an extrinsic cofactor or dye. The protein is comprised of an 11-stranded β-barrel that surrounds a central α-helix (Figure 1.2).[22, 23] This central helix contains the chromophore, a *p*-hydroxybenzylidene-imidazolidinone group that is

responsible for the fluorescence of the protein. The chromophore is formed by the autogenic cyclization of a tripeptide sequence, Ser65-Tyr66-Gly67 in GFP,[24] in a process known as maturation.

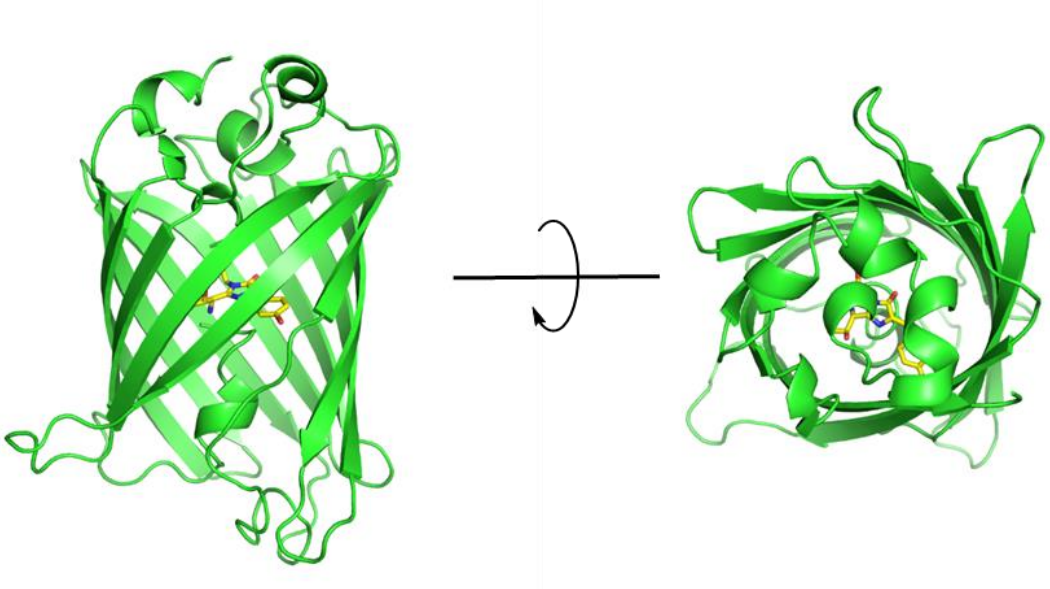


Figure 1.2 Structure of the green fluorescent protein. A cartoon representation of the crystal structure of the green fluorescent protein from *Aequorea victoria*, generated with PyMOL, based on the PDB entry 1EMB. Side and top views of the barrel are presented, and the chromophore is shown as yellow sticks.

In the generally accepted mechanism for GFP chromophore maturation, the amide nitrogen of Gly67 performs a nucleophilic attack on the carbonyl carbon of Ser65, forming a 5-membered ring (Figure 1.3).[25] This higher-energy intermediate is then conjugationally trapped by a dehydration step,[26] followed by the rate-limiting oxidation step, which requires the presence of molecular oxygen. The chromophore microenvironment created by the surrounding protein barrel has been shown to be crucial for maturation, as the central GFP helix that contains the chromophore tripeptide is kinked, disrupting the backbone hydrogen bond network and thereby lowering the energy barrier for cyclization, while also positioning the chromophore backbone in proximity to the catalytic Arg96 and Glu222 residues that catalyze the dehydration step.[26-28] Additionally, containment of the mature chromophore inside of the barrel is required for fluorescence, as the free chromophore in solvent has been shown to be nonfluorescent.[29] Computational modelling suggests that in solvent, the excited chromophore molecule can undergo radiationless deactivation through rotation around the exocyclic C-C double bond (cis-

trans photoisomerization), a process that is sterically and electrostatically inhibited by the protein barrel.[30] Thus, the tertiary structure of the protein contributes to its relatively uncommon function both by reducing the energy barrier of maturation and shielding the mature chromophore.

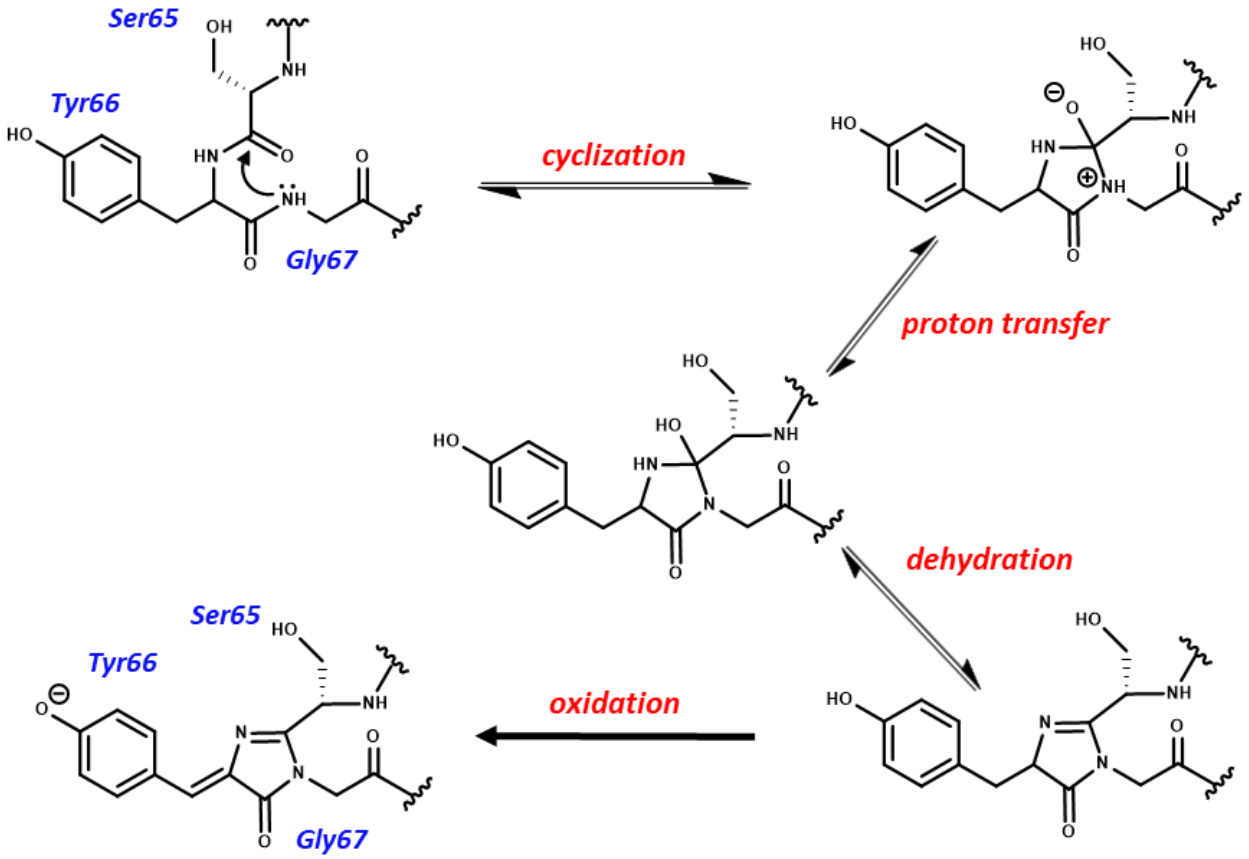


Figure 1.3 GFP chromophore maturation. A representation of the cyclization-dehydration-oxidation mechanism of chromophore maturation in GFP. In this mechanism, maturation begins with the cyclization of the peptide backbone between the Gly67 amide nitrogen and the Ser65 carbonyl carbon, followed by a dehydration step, and finally the rate-limiting oxidation of the Tyr66 $C_\alpha C_\beta$ bond. Figure adapted with permission from Grigorenko *et al.*[3] Copyright 2017 American Chemical Society.

This tertiary structure is not unique to the *Aequorea victoria* GFP, however, as there exist a number of homologues from other cnidarians,[31-34] as well as some bilaterians,[35] such as copepods[36, 37] and amphioxus.[38, 39] Within these homologues, as well as in some of the early mutagenesis experiments performed on GFP,[22, 25] mutations were discovered that can alter the excitation and emission wavelengths of the resulting chromophore, leading to different colours of fluorescent protein.

Currently, fluorescent proteins with ultramarine,[40] blue,[41] cyan,[42] green,[43] yellow,[44] orange,[45] red,[46] and near-infrared[47] spectra are available, spanning emission wavelengths from 440-685 nm.[1] Some of these colour changes occur as a result of mutations at position 66 (Figure 1.3), replacing the chromophore tyrosine with another aromatic amino acid to directly alter the conjugated system formed. For instance, replacing Tyr66 with phenylalanine or histidine in GFP results in a blue fluorescent protein, and replacing it with tryptophan results in a cyan fluorescent protein. These proteins are blue-shifted relative to GFP, as the aromatic side chains are less able to donate electrons to the chromophore system, relative to the electron-rich phenolate of a charged tyrosine.[48] In order to red-shift the spectrum, it follows that additional electron density must be donated to the system, stabilizing the excited state of the chromophore to reduce the gap between the ground and excited states. This was first accomplished with the creation of the yellow fluorescent protein,[22] where Thr203 was mutated to a tyrosine. Position 203 is adjacent to the chromophore, allowing its side chain to π-stack with the phenolate side of the conjugated system, leading to a roughly 15 nm red-shift in emission. Further red-shifting of the *Aequorea* GFP through mutagenesis proved difficult, however, such that the first reported red fluorescent protein, DsRed, was instead a homolog derived from *Discosoma* sp.[32] It was found that this red-shift occurs as the result of an additional dehydration step on the backbone of Gln66, further extending the conjugated system of the chromophore by two double bonds through the formation of an acylimine.[7] This requires a longer maturation time, and it has been found that the formation of green chromophore in DsRed competes with the formation of red chromophore, in a branched pathway model of maturation that leads to a mixed population of green and red chromophores.[49] Thus, this extended chromophore complicates maturation, but has nevertheless resulted in a number of widely-used red fluorescent proteins, particularly after extensive mutagenesis to monomerize the DsRed tetramer and to improve maturation and brightness.[46, 50, 51] While these examples are not an exhaustive list of all possible chromophore modifications or structures, they serve to demonstrate the potential for

engineering novel fluorescent proteins, as well as the wide variety of fluorescent proteins that are currently available to researchers.

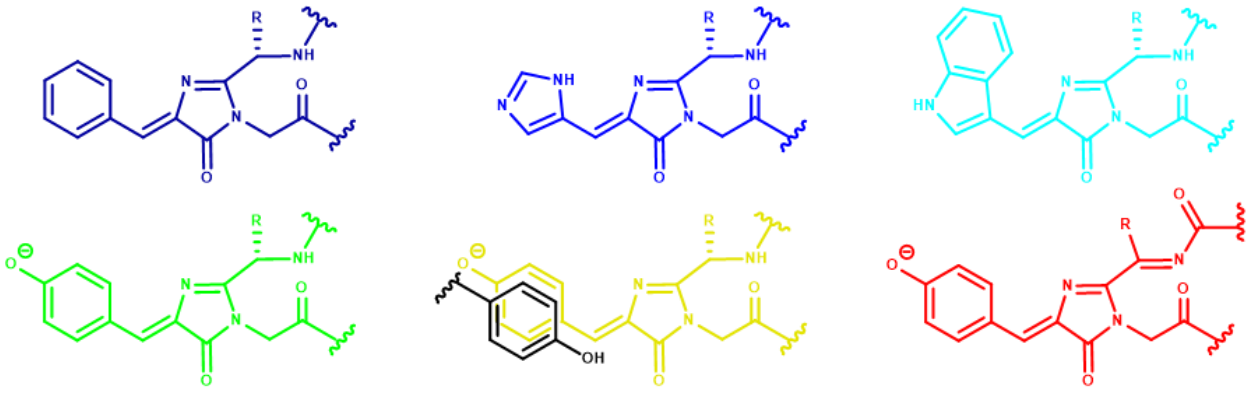


Figure 1.4 The fluorescent protein chromophore palette. A selection of chromophore structures used in fluorescent proteins of different colours. Mutating position 66 to other aromatic amino acids (Phe, His, Trp) leads to blue-shifted fluorescence relative to GFP, while adding π-stacking interactions or increasing conjugation through the formation of an acylimine leads to red-shifted fluorescence.

1.1.3 Fluorescent Protein Applications: Tags, Biosensors and the GCaMP Family of Calcium Ion Sensors

Because fluorescent proteins contain fluorophores comprised uniquely of proteinogenic amino acids, they can be fused with other proteins of interest using molecular biology techniques, and these fusions can be genetically encoded into research model organisms without needing any additional cofactors to become fluorescent. Thanks to the high sensitivity of fluorescence imaging equipment, these fusion proteins can then be used to study protein expression, localization, movement and turnover.[52, 53] By identifying DNA- and RNA-binding proteins, as well as proteins associated with particular subcellular locations, cells and tissues,[54, 55] GFP fusions can also provide localization information about any of these biomacromolecules or higher-order structures. The first examples of the use of GFP as a fusion tag were reported in 1994,[56, 57] and by 1995 dozens of fusions in multiple species and cell lines were already reported,[58] with a current search of the PubMed database for the MeSH term “green fluorescent proteins” returning more than 46 000 results. Although the original *Aequorea* GFP has fallen out of favour due to the development of improved versions such as the

enhanced green fluorescent protein (EGFP),[43, 59] a number of cyan, green, yellow and red fluorescent proteins are in common usage. Entire libraries of fluorescent protein fusions are now available, including the majority of the yeast proteome,[60] hundreds of proteins from mouse embryonic stem cells,[61] over 1000 proteins in human lung carcinoma cells,[62, 63] and 10 000 proteins from *Drosophila melanogaster*.[64] Overall, GFP and its mutants have become indispensable tools for biologists and biochemists, and these fusion tags continue to be relevant in fields of modern science, including cancer research,[65, 66] neurobiology,[67, 68] embryogenesis,[69, 70] virology,[71, 72] single-molecule and super-resolution imaging[73-75] and more.

In addition to these uses of fluorescent proteins as passive indicators, they have also been engineered into active indicators, where their fluorescence is altered by their environment. This turns the fluorescent protein into a biosensor, detecting an analyte, protein-protein interaction or enzymatic activity of interest, based on how they influence the measured signal of the fluorescent protein. This kind of modulation was originally performed by introducing Förster resonance energy transfer (FRET) between a blue and a green fluorescent protein, by fusing them together using a short linker containing a protease recognition site.[53, 76, 77] When the two proteins were kept in close proximity by the linker, FRET could be detected as green emission when exciting the blue fluorescent protein, due to the transfer of energy from the excited state of the blue chromophore to the green one. In the presence of active protease, the linker would be cleaved, separating the two proteins and reducing FRET. Thus, this system is a sensor for protease activity, quantified by measuring the ratio of blue to green fluorescence when exciting the blue chromophore.

Since then, hundreds of fluorescent protein-based biosensors have been developed for a wide range of targets, with detection based on changes in FRET, fluorescence lifetime, intensity, excitation and emission spectra, translocation of the fluorescent protein, and more.[78-80] One well-studied family of these biosensors are the genetically encoded calcium ion indicators (GECIs),[81] in particular those

based on GCaMP. First published in 2001, GCaMP is a biosensor that detects calcium ions based on a change in fluorescence intensity.[82] It consists of a circularly permuted enhanced GFP (cpGFP), in which the N- and C-termini of the protein have been moved from the top of the barrel to the side, next to the chromophore. This creates a pore in the barrel, exposing the chromophore to solvent and quenching fluorescence. In order to modulate the fluorescence intensity, the N-terminus is fused to the M13 fragment of myosin light chain kinase, and the C-terminus is fused to calmodulin. In the presence of calcium ions, calmodulin changes conformation and binds the M13 peptide, thereby covering the pore and restoring fluorescence. Crystal structures of the GCaMP2 mutant showed that this increase in fluorescence is due to the change in chromophore environment upon calcium ion binding (Figure 1.5).[11, 83] In the unbound state, the pK_a of the phenolate oxygen of the chromophore is increased, leading to protonation of the chromophore. This protonated chromophore absorbs light around 400 nm, and thus is not excited by the 488 nm laser generally used to excite GFP. When calcium ions bind to calmodulin, the change in conformation excludes solvent from the chromophore, stabilizing the anionic form of the chromophore, which is the productive form for the purposes of fluorescence spectroscopy experiments, absorbing around 495 nm. This manifests as an effective increase in the extinction coefficient when calcium ions bind, as a larger fraction of the population of chromophores is able to absorb the appropriate wavelength of light, while the quantum yield remains unchanged.[84, 85] This increase in brightness was around 4.5-fold over the baseline in the original GCaMP, with improved mutants later showing increases greater than 10-fold, improving the visibility of the sensor in living organisms.[86, 87] A variety of different colours of GCaMP-based calcium ion sensors have since been created using circularly permuted blue, cyan, yellow, orange and red versions,[84, 88-90] allowing these sensors to be used in systems where a green fluorescent indicator is already present. Interestingly, the cause of the brightness change of the different coloured sensors varies, as not all versions show the same decrease in pK_a when calcium ions bind. The blue and cyan sensors instead show an increase in quantum yield when bound, while a red version, named RCaMP, was found to show an increase in

both extinction coefficient and quantum yield. On the other hand, another red calcium ion sensor, R-GECO1, showed the expected increase in extinction coefficient when calcium ions were bound.[84] Crystal structures of these red sensors also showed large changes in the orientations of the circularly permuted fluorescent proteins and calmodulin, relative to the GCaMP crystal structures, implying that structural information is important to rationally improving these proteins, but also that a number of possible orientations can lead to productive sensors, so long as they can transmit changes due to calcium ions binding to the chromophore environment.

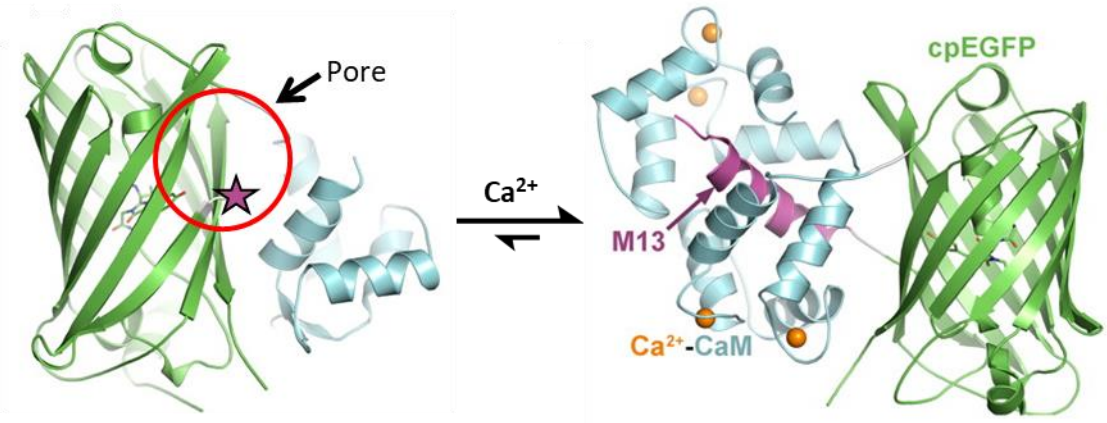


Figure 1.5 Crystal structures of GCaMP2. Crystal structures of the GCaMP mutant, GCaMP2, (left) in the calcium ion-free state and (right) in the calcium ion-bound state. The chromophore is shown as sticks, and calcium ions (Ca^{2+}) are shown as orange spheres. The domains are colour coded, with the cpGFP in green, calmodulin (CaM) in cyan, and the M13 peptide in purple. Note that the M13 peptide and the C-terminal half of calmodulin are not shown in the calcium ion-free state due to lack of electron density. A purple star shows where the M13 peptide should attach to the cpGFP. This figure was adapted with permission: This research was originally published in the Journal of Biological Chemistry. Akerboom *et al.*[11] © the American Society for Biochemistry and Molecular Biology

1.1.4 The Limitations Imposed by Maturation

Before it can be imaged, a fluorescent protein must first express, fold and mature, with the rate-limiting step generally being chromophore maturation.[52] Due to the additional steps involved in the maturation of the red chromophores, they tend to mature more slowly than their cyan, green and yellow counterparts. Thus, the half-time of fluorescent protein maturation varies from less than 10 minutes for the fastest-maturing cyan, green and yellow proteins,[44, 91, 92] to a range between 15 minutes and multiple

hours for most red and orange fluorescent proteins,[34, 46] to more than a day for certain red fluorescent timer proteins specifically selected for long maturation times.[93] Measured maturation rates can vary considerably based on the conditions at which they are measured, as the temperature kinetically affects the rate at which the chemical reactions occur, and the state of the chromophore when the measurement begins can also differ. Some measurements are performed by expressing fluorescent proteins in the absence of oxygen to halt maturation at the final oxidation step, either using *in vitro* synthesis methods,[94] or by growing cells anaerobically.[95, 96] Other methods use time-lapse microscopy to detect the increase in fluorescence after the induction of fluorescent protein expression,[92] thus measuring not only the final oxidation step, but the entire process of transcription, translation, protein folding and maturation. As an example of the potential variation in these measurements, GFP was originally determined to have a half-time of maturation of 2.9 hours,[25] based on experiments growing cells anaerobically at 22°C, but was later determined by the same group to have a half-time of 1.4 hours,[96] using a similar method with less anaerobic growth, as they had determined that letting the cells grow anaerobically for extended periods slowed the measured maturation time once oxygen was finally introduced. On the other hand, a more recent experiment using a cell-free expression system with oxygen scavenging at 37°C found a half-time of 0.62 hours.[94] Despite the difficulty of absolutely quantifying maturation rates, fluorescent proteins that mature rapidly have nevertheless been produced, and Balleza *et al.* have shown that most commonly-used fluorescent proteins mature in *E. coli* with half-times less than an hour at 37°C, generally reaching 90% of maximum fluorescence between 20 minutes and 5 hours after expression.[92]

As a result, most fluorescent proteins will be mature after a few hours, and many imaging experiments incorporate an incubation of multiple hours between protein expression and imaging, allowing fluorescence to develop before any measurements are taken.[69, 97-101] But this is not always feasible, for instance if the goal of the experiment is to track protein expression in real-time. In such a

case, using a reporter fluorescent protein with a long maturation time will reduce the temporal precision of the measurements, and has been shown to "smear" measurements of promoter dynamics, where the variability of individual maturation times in molecules of a slow-maturing fluorescent protein lead to signal averaging when it was transcribed in bursts from its promoter.[92] In general, using a fluorescent protein reporter to analyze any dynamic process that varies on the minute timescale post-expression could be limited by maturation of the chromophore. Little *et al.* demonstrated this effect when measuring the formation of the bicoid gradient in *Drosophila* embryos, using an EGFP-bicoid fusion protein expressed at the anterior of the embryo.[10] They compared images of the live embryos to images of the same embryos after fixation, and found that observed shape of the gradient had changed, despite there only being a three-minute delay between the two measurements (Figure 1.6). The anterior portion of the embryo, where the most recently translated proteins were, showed a relatively larger increase in fluorescence after fixation, compared to the average. They attributed this result to the presence of immature EGFP fusions, and demonstrated that the same effect could be seen when using a slower maturing red fluorescent protein tag, mRFP.[50] Most importantly, they also found that this effect could no longer be observed when using one of the fastest-maturing fluorescent proteins, Venus,[44] implying that it was dependent upon the maturation rate of the fluorescent protein used.

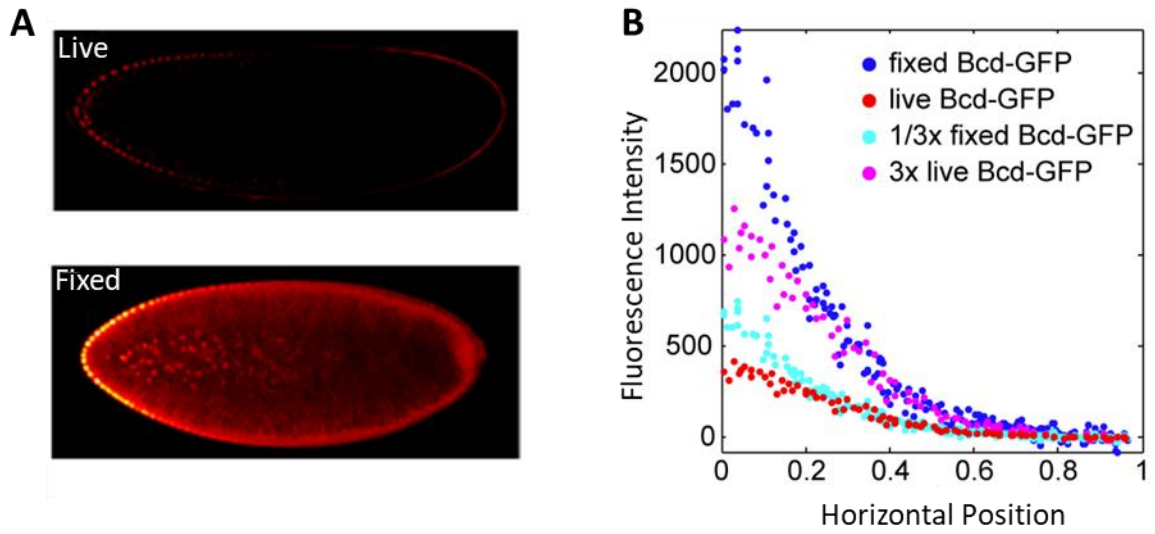


Figure 1.6 Maturation delay can influence measurements on the minute timescale. A, Confocal microscopy images of a bicoid-EGFP expressing *Drosophila* embryo. The live image was taken first, then the fixed image was taken under the same microscopy settings, with less than a 3 minute delay between the two images. **B,** Mean nuclear fluorescence intensities in the live and fixed embryos, as a function of position along the anterior-posterior axis. As the fixation process increases signal by roughly threefold, rescaled live and fixed intensity values are shown to highlight that the difference between the two is not solely due to the improved transparency of fixed material. Adapted from Little *et al.*,[10] under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

For the measurement of rapid processes using fluorescent proteins, it is therefore imperative to reduce the maturation time as much as possible, and Venus is generally the tag of choice for such experiments. It has been used to quantify protein expression with single-molecule sensitivity in live *E. coli* by Yu *et al.*, where they determined it to have a temporal sensitivity in the minute range.[102] This was sufficient for them to model the transcription and translation of the protein from their model promoter, a repressed *lac* promoter, down to the production of individual mRNA molecules. In another paper by Taniguchi *et al.*, the same group went on to perform high-throughput quantification of expression at the proteomic level, measuring the expression levels of over 1000 proteins in *E. coli*.[9] They then compared these protein quantities to the quantity of mRNA for the same gene, determined using fluorescence *in situ* hybridization (FISH). Intriguingly, they found that there was almost no correlation between the two levels (Figure 1.7), despite the central dogma dictating that all protein molecules are translated from mRNA, and the fact that other groups have identified correlations

between the two when using mass-spectrometry based methods to detect protein levels in cell populations.[103, 104] They explain that this lack of correlation occurs mainly due to the difference in lifetime between the two molecules, as mRNA is degraded much more rapidly than protein molecules, and thus that their fixed time point measurements do not properly integrate the overall levels of each over time. Although this effect was sufficient to explain the majority of the lack of correlation, their mathematical model still predicted some correlation between mRNA and protein quantity, so they hypothesized that there was also a level of transcriptional noise that further reduced the observed correlation to 0. Gedeon and Bokes, however, hypothesized that this lack of correlation could potentially be explained by the translation and maturation time of Venus.[12] When they incorporated this delay (approximated at 7.5 minutes) between protein translation and fluorescence signal into their own model, they found that it calculated a reduced correlation coefficient within the bounds determined by the experimental data, thus providing a source of the "transcriptional noise". They also demonstrated that, for a protein with such a delay between translation and fluorescence, the correlation between mRNA and protein abundance over a time course can be improved by shifting the protein levels "backward in time" by the amount of the delay, i.e. that for a protein that folds and matures instantly, there would be a directly observable correlation (Figure 1.7).

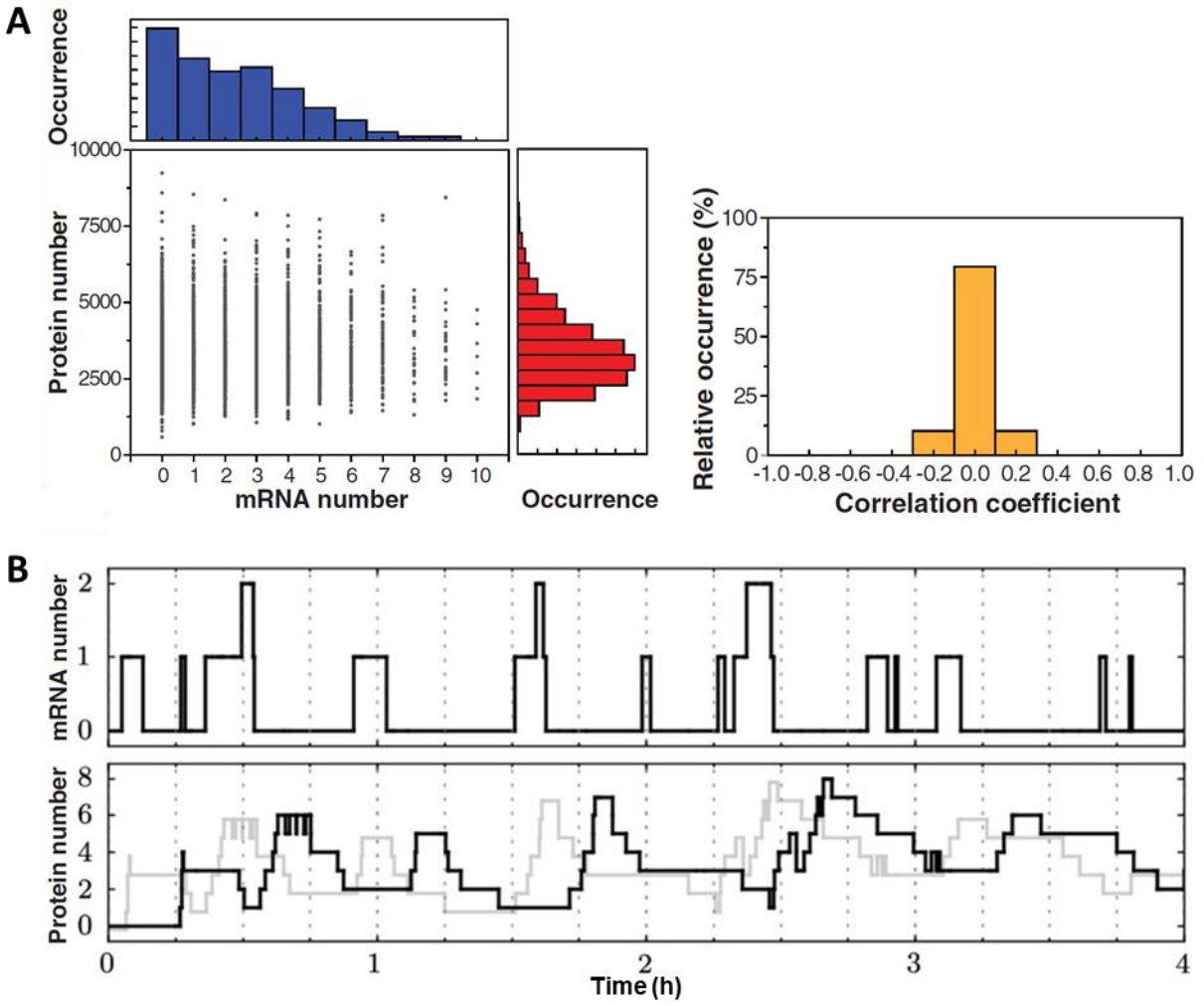


Figure 1.7 Maturation delay can lead to temporal decoupling of events. A, (left) Measured mRNA and protein levels for Venus-tagged TufA in *E. coli*, with each point representing a single cell. The correlation coefficient is r = 0.01 ± 0.03 (mean ± SD, n = 5447). (right) Correlation coefficients between mRNA and protein levels from 129 strains, each expressing a different Venus-tagged protein, demonstrating that the lack of correlation occurs generally. This figure is from Taniguchi *et al.*[9] Reprinted with permission from AAAS. **B,** Simulation of measured mRNA and protein levels over time using a model incorporating translation and maturation delay, for a hypothetical low-copy protein. For the protein number, the grey line represents the measured protein number for instantaneous translation and maturation, while the black line represents the measured protein number assuming a delay of 12 minutes. This delay is longer than that of Venus (approximated at 7.5 minutes), to emphasize the difference between the two states. Note that a correlation between mRNA and protein levels at the same time point is visible in the model without delay. Reprinted from Gedeon and Bokes,[12] with permission from Elsevier.

This presents a cautionary tale about using fluorescent protein fusion tags, or fluorescent protein-based biosensors that mature during the detection window of the experiment, including the FRET-based sensor described in the previous section, or bimolecular fluorescence complementation assays, which rely on the reconstitution of a fluorescent protein that was split into two immature halves.[105] As these proteins are not fluorescent until their chromophore is fully mature, any tagged proteins or interactions are "invisible" for multiple minutes after the fluorescent protein is formed, and thus will not be detected if the imaging experiment occurs within that time. On the other hand, for biosensors such as GCaMP, the fluorescent protein is expressed in a form that is able to mature in the absence of the target. As long as there is time for these biosensors to mature before the event of interest, then the problem of maturation can be avoided entirely. Thus, we seek to develop such a biosensor that can be used to detect any protein of interest on a rapid timescale, retaining the advantages of using fluorescent proteins as fusion tags, such as their ability to be genetically encoded and wide range of available chromophores, while expressing the actual fluorescent protein separately from the protein of interest, in a form that can mature before the target protein is expressed. To create this biosensor, we will exploit the GCaMP family of biosensors, replacing the calmodulin-M13 interaction with a peptide-protein interaction from a pair of Bcl-2 family proteins, Bim and Bcl-x_L. The peptide will be tagged to a circularly permuted GFP, while the protein will be used as a fusion tag attached to the protein of interest. This will allow the two halves of the sensor to be expressed separately, allowing the cpGFP to mature before the protein of interest is produced. Once the tagged protein of interest is expressed, binding of the peptide-protein pair will occlude the pore, stabilizing the phenolate form of the chromophore and increasing fluorescence. In theory, this should reduce the delay between expression and detection from the timescale of maturation (minutes), to that of diffusion (seconds).

1.2 The Bcl-2 family proteins: Bcl-x_L and Bim

There exist a wide variety of well-studied protein-protein interaction motifs that could be used in a biosensor, including coiled-coil heterodimers,[106, 107] antibody-antigen interactions,[108, 109] PDZ domain peptide ligands,[110-112] SH3 domain peptide ligands,[109, 113, 114] WW domain peptide ligands,[115, 116] and Bcl-2 type receptor/BH3 complexes.[117, 118] These motifs all have known specificities and affinities, and have structural data available to aid in the initial design of the biosensor scaffold. For our sensor, we selected the Bcl-2 type receptor/BH3 complexes due to their extremely tight binding, with K_d values in the low- to sub-nanomolar range,[119, 120] as well as their relatively small size.

The Bcl-2 type proteins are a family of pro- and anti-apoptotic proteins, defined in eukaryotes by their possession of at least one of four conserved sequence motifs known as Bcl-2 homology domains (BH1-4),[121] although some Bcl-2-like proteins exist in viruses that are structurally similar yet do not have any sequence similarity to these domains.[122] They are broadly divided into two groups: the pro-survival proteins containing three or four of the homology domains, such as Bcl-2 itself, Mcl-1 and Bcl-x_L, and the pro-apoptotic proteins, which are further divided into activator proteins that only contain the BH3 domain, such as Bim[123] and Noxa, and larger pore-forming proteins, such as Bax and Bak.[124] Hetero- and homo-oligomerization of these proteins govern their effect on the cell, with the pro-survival proteins binding to the pro-apoptotic proteins to sequester them, while homo-oligomers of the pro-apoptotic Bax and Bak proteins create pores in the outer membrane of the mitochondria, leading to the release of intermembrane space proteins that activate caspases in the cytoplasm and lead to apoptosis.[125] The smaller BH3-only pro-apoptotic proteins serve as regulators, either by binding competitively to the pro-survival proteins or by binding to Bax and Bak in order to promote oligomerization.[126]

Of these proteins, the mammalian pro-survival protein Bcl-x_L was the first to have its 3-D structure solved.[127] It was found to consist of eight α-helices, with helices α5 and α6 forming a central

hairpin surrounded by the others, as well as a C-terminal transmembrane region, which was removed for the original structural studies as it destabilized the protein in solution and the truncated protein was shown to retain its function *in vivo*. The BH domains are distributed around the structure, with BH1 encompassing the turn between α4 and α5, BH2 the turn between α7 and α8, BH3 being entirely located on α2 and BH4 being located on α1. Importantly, this structure contains a hydrophobic groove, formed predominantly between helices α3 and α4, that forms the binding site for the pro-apoptotic proteins (Figure 1.8).[4] This groove is formed by the BH1, BH2 and BH3 domains of the pro-survival protein, and differences in this cleft have been found to be at the root of the differences in specificity of the different pro-survival proteins. For instance, the wider groove in pro-survival Bcl-2 leads to a 10-fold weaker binding interaction with pro-apoptotic Bak and Bad proteins, compared to their affinity for Bcl-x_L.[124, 128] While the majority of the tertiary structure of pro-survival proteins such as Bcl-x_L is required to form the binding groove, studies of the pro-apoptotic proteins have found that only a portion of their BH3 domain is necessary to bind the groove. As a result, short peptides (between 16 and 26 amino acids in length) derived from the pro-apoptotic proteins have been shown to retain high affinity for the pro-survival proteins.[119, 129, 130] These peptides form an amphipathic α-helix, with four conserved hydrophobic residues pointing into hydrophobic pockets in the groove of the pro-survival protein, conferring binding affinity. The small size of these BH3 peptides makes them ideal as potential components of our sensor, and even the larger pro-survival proteins are less than 220 amino acids in length after the transmembrane domain is removed, making them similar in size to GFP and reasonable to use as a fusion tag.

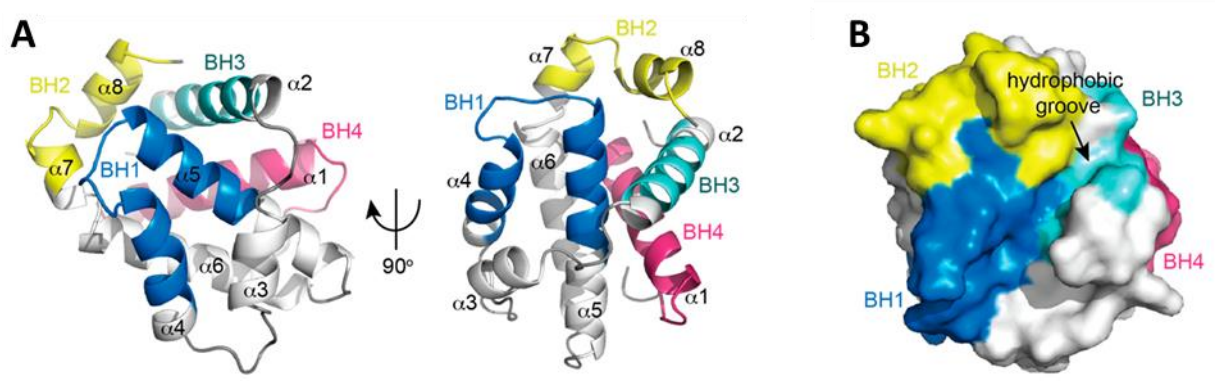


Figure 1.8 Annotated structure of Bcl-x_L, a pro-survival protein. A, Helical representation of apo-Bcl-x_L (PDB ID: 1MAZ), showing the Bcl-2 homology (BH) domains in different colours. The α5 and α6 helices form a central hairpin surrounded on either side by the other helices. **B,** Surface representation of apo-Bcl-x_L demonstrating the canonical hydrophobic binding groove created mainly by helices α3 and α4 with α5 forming the base. This groove is critical for mediating interactions with the BH3 domains of the pro-apoptotic proteins of the Bcl-2 family. Adapted from Lee and Fairlie,[4] under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

An important quality for a biosensor is that it have as small an effect as possible on the normal functionality of the cell. Thus, in order to use Bcl-2 family proteins as *in vivo* sensors, it is important to ensure that they will not trigger apoptosis. For this reason, we selected Bim as the pro-apoptotic BH3-only protein to use for our sensor, as a 26-amino acid peptide of Bim retains high affinity for the pro-survival proteins,[119] and structural methods to regulate its activity have been well studied. These methods start with the discovery of three different isoforms of the protein,[123] Bim_{EL}, Bim_L, and Bim_S, that show an inverse relationship between length and pro-apoptotic activity: Bim_{EL} is the longest sequence and showed the least cytotoxicity, while Bim_S is the shortest sequence, and was most effective at killing cells. This was later found to be due to sequestration of the longer Bim isoforms by binding to the microtubule-associated dynein motor complex,[131] and thus the Bim_{EL} and Bim_L isoforms would not be particularly good tags, as they would presumably lead to sequestration of their associated fusion proteins. Conveniently, it has been shown that the pro-apoptotic effect of Bim_S in epithelial cells can be abrogated by removing its C-terminal membrane binding domain, as localization to the mitochondrial membrane in necessary for its activity.[132] Thus, retaining only the BH3 domain of Bim

should limit its pro-apoptotic activity, while retaining its affinity to a pro-survival tag such as Bcl-x_L. As for Bcl-x_L, it has been shown to form a pro-apoptotic fragment when caspase cleavage removes the N-terminal BH4 homology domain,[133] but this activity also requires localization to the mitochondrial membrane, and can once again be eliminated by removal of the C-terminal membrane anchor sequence.[134] Therefore, eliminating the various targeting sequences from the Bim and Bcl-x_L proteins should allow them to serve as effective fusion tags, as they will no longer be sequestered to various locations in the cell, and their pro-apoptotic activities will be eliminated.

1.3 Overview of Principal Techniques and Data Analysis

1.3.1 Expression of Multiple Recombinant Proteins in *E. coli*

For the rapid production of high levels of recombinant protein in the laboratory, *E. coli* remains the organism of choice. It has been well studied, grows rapidly to a high culture density, and can easily be transformed with exogenous DNA to insert sequences of interest into the cells.[135] This exogenous DNA is prepared in the form of an expression plasmid, a circular molecule of DNA that minimally contains the gene of interest in a multiple cloning site, a promoter sequence that leads to transcription of the gene, a selectable marker that allows cells that contain the plasmid to be identified, and an origin of replication that allows the extrachromosomal plasmid to replicate independently from the chromosome, such that the plasmid is maintained through multiple generations of cellular division.

When only a single protein of interest is to be produced, as is generally the case when preparing recombinant proteins for *in vitro* tests using purified protein, most expression plasmids will be effective, varying mainly in the overall level of protein produced. But most of these plasmids are designed to express a single protein, so expressing multiple proteins, either simultaneously or sequentially, requires particular consideration of how the sequences will be cloned and how their expression will be controlled. The series of DUET vectors (Novagen) are the most commonly used

vectors for expressing multiple proteins, as each contains two multiple cloning sites, and four compatible vectors are available, allowing up to 8 proteins of interest to be expressed in a single cell. But all of these genes would be under the control of the same promoter, and thus would be expressed simultaneously. For our biosensor, we were interested in expressing two proteins sequentially, such that the cpGFP could mature before the protein tag was expressed, and so the DUET system did not suit our needs.

To allow sequential expression of our two proteins, we chose to place each of them under the control of a different promoter. At the time, there were no plasmids commercially available that contained two independent promoters, and so instead we chose to clone each gene into a separate plasmid, then transform the two plasmids into a single stock of bacteria. But cells containing multiple plasmids are known to suffer from plasmid incompatibility, i.e. the loss of all but one of the plasmids over generations in the absence of selective pressure.[8, 136] This occurs with modern expression plasmids because almost all of them use the same origin of replication, known as the ColE1 origin,[137] the pMB1 origin or the pBR322 origin,[138] based on the early plasmids in which this origin was discovered. This origin of replication encodes an RNA pre-primer, RNA II, which can hybridize with the DNA template strand. After processing by RNAseH to add a 3'-OH group, RNA II can be extended by DNA polymerase I, initiating replication fork assembly and allowing the polymerase Pol III to perform replication of the plasmid.[139, 140] This replication mechanism is regulated by the transcription of an antisense RNA molecule, RNA I, which is also encoded in the origin of replication of the plasmid, but under a much stronger promoter. RNA I is complementary to part of the 5' end of the RNA II pre-primer, and when the two are bound, the structure of RNA II is altered such that its 3' end cannot hybridize to the DNA template, preventing replication. This creates a negative feedback loop, where the more the plasmid is copied, the more replication is inhibited by the excess of RNA I. For a single type of plasmid in a cell, this serves to regulate the copy number of the plasmid, or number of plasmids

in each cell, preventing runaway plasmid replication and thereby limiting the metabolic burden on the cell. For multiple plasmids with the same origin of replication, incompatibility occurs from the combination of this negative feedback loop and the stochastic nature of the initiation of replication. As RNA II can replicate any of the plasmids present, replication will occur on a random plasmid. If at any point one of the plasmids replicates more than the others, such that it has more copies than the others, then replication of that plasmid becomes statistically more likely, while simultaneously continuing the negative feedback loop by synthesis of RNA I, reducing the overall likelihood of any replication event.

Once one plasmid gains an advantage in terms of relative copy number, it tends to maintain this advantage, such that more copies of that plasmid are present during cell division, and relatively few of the other plasmids. The plasmids segregate themselves randomly between the two daughter cells, so for the plasmids with few copies, it is possible that one of the daughter cells receives no copies, leading to segregational incompatibility (Figure 1.9). Over the course of multiple generations, this plasmid loss continues, as does the division of cells that have already lost plasmids, until eventually the culture consists mainly of cells containing a single plasmid. In order to avoid this plasmid loss, multiple selection methods can be used, where each plasmid contains a different selection factor. For instance, a number of antibiotic resistance genes are used in plasmids, including resistance for ampicillin, kanamycin and chloramphenicol.[141] If each plasmid used provides a different antibiotic resistance, growing bacteria in a cocktail of all resisted antibiotics will ensure that any cells that suffer plasmid loss will not be viable. But as the plasmid loss is still occurring, culture growth will be slower due to this population of nonviable cells.

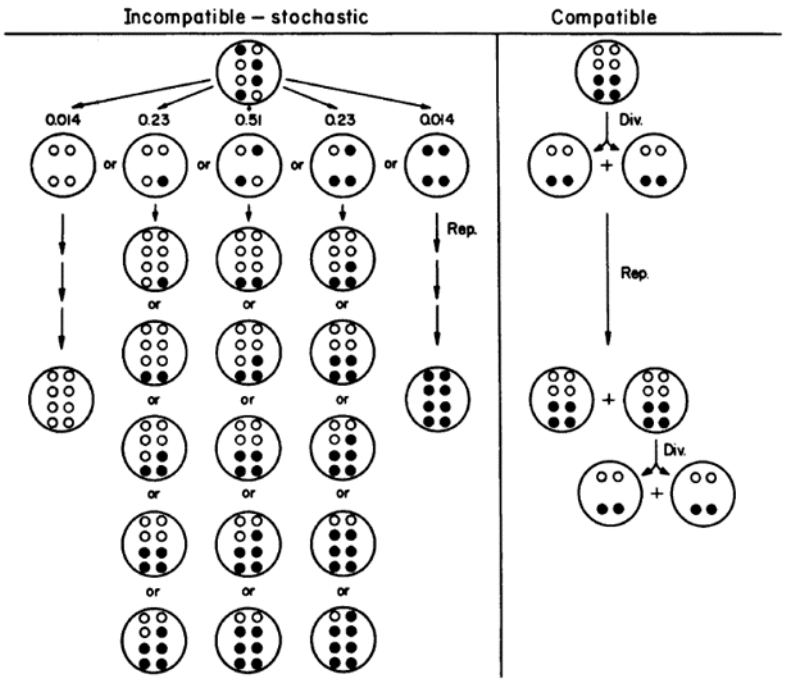


Figure 1.9 Example of plasmid segregational incompatibility. In this figure, small black and white circles represent two different plasmids inside a cell, with a total copy number of 8 plasmids. (left) The various possibilities of plasmid transmission to daughter cells after cell division, along with their probabilities (assuming a binomial distribution of plasmids), along with the possible populations of the different plasmids after plasmid replication. Note the possibility of plasmid loss due to the stochastic distribution of plasmids to the daughter cells, even when the initial populations are equal. (right) Plasmid transmission with two compatible plasmids, where each plasmid replicates independently to equilibrium after each cell division. Figure replicated with permission from Novick.[8] Copyright © 1987, American Society for Microbiology

To avoid plasmid incompatibility, compatible origins of replication can be used. Within the ColE1 family of origins of replication, mutants have been identified that control replication in the same fashion, but are still compatible, including p15A,[142] RSF1030,[143] and CloDF13.[144] These origins of replication have different sequences for the primer RNA II and the inhibiting RNA I, and thus are unable to interfere with the replication of plasmids containing the ColE1 origin of replication, and vice-versa.[145] There are also unrelated origins of replication, such as the pSC101 plasmid[146] that is regulated by monomer-dimer equilibrium of the RepA protein encoded by the plasmid.[140] While plasmids with these origins of replication generally have lower copy numbers, and are therefore less commonly used for protein expression, the ability to stably replicate alongside a ColE1 plasmid renders them useful for multiple protein expression.

In order to sequentially express multiple proteins, different promoters are required. While there are some constitutive promoters available, such as the pBla promoter used to provide β-lactamase activity in some plasmids,[147] they are less commonly used for recombinant protein expression, due to the increased metabolic burden on the cell resulting from the continuous expression of a protein. Instead, a number of promoters that can be induced by small molecules are commonly used, including 3-β-indoleacrylic acid (IAA)-inducible trp,[148, 149] isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible lac, which includes lacUV5,[150] tac,[151] and trc,[152] and arabinose-inducible araBAD.[153] These inducible promoters allow the time of protein expression to be controlled, as mRNA transcription is inhibited by the binding of a repressor protein to the promoter/operator, competing with RNA polymerase binding. This repressor can be allosterically modulated by binding of the inducer, causing a change in conformation that prevents the repressor from blocking the promoter, as in the trp and lac promoters,[149, 154] or changing it into an activator form, as in the araBAD promoter.[155] For these examples of inducible promoters, the repressor proteins are endogenously expressed in *E. coli*, but additional copies of the gene are often added to expression plasmids using the associated promoter, to ensure that sufficient quantities of repressor are created to repress the multiple copies of each plasmid present in the cell. By combining one constitutive promoter with an inducible promoter, or multiple inducible promoters with different inducers, different proteins can be induced at different times. Thus, by using two plasmids with compatible origins of replication, different selection markers, and different promoters, independent expression of two proteins in a single *E. coli* cell can be achieved.

1.3.2 Analysis of Equilibrium and Non-equilibrium Binding

The reversible binding between two molecules, A and B, to form the non-covalent complex AB, can be described using the following kinetic scheme, assuming 1:1 stoichiometry:[156]

$$A + B \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} AB \qquad \text{(Equation 3)}$$

In this case, k_{on} is the second order association rate constant, with units of M^{-1} s^{-1}, and k_{off} is the first order dissociation rate constant, with units of s^{-1}. When solutions of A and B are mixed, binding will occur to form AB, with a rate of:

$$Binding\ rate = k_{on}[A][B] \quad \text{(Equation 4)}$$

Where [A] and [B] are the molar concentrations of A and B. Once AB is formed, it will dissociate back into A and B with a rate of:

$$Dissociation\ rate = k_{off}[AB] \quad \text{(Equation 5)}$$

Where [AB] is the molar concentration of the AB complex. Association and dissociation will continue until the system reaches an equilibrium, where the rates of binding and dissociation are equal. The affinity of the two molecules for one another is defined by the relative concentrations of the bound and unbound molecules at equilibrium, which is also equal to the ratio of the two rate constants. These values are often expressed as the dissociation equilibrium constant, or K_d:

$$K_d = \frac{k_{off}}{k_{on}} = \frac{[A_{eq}][B_{eq}]}{[AB_{eq}]} \quad \text{(Equation 6)}$$

The K_d has units of M (for 1:1 stoichiometry), and represents the inverse of the affinity of the two molecules for each other: the lower the K_d, the more tightly they bind, and the more AB complex there will be at equilibrium. Note that the values $[A_{eq}]$ and $[B_{eq}]$ are the concentrations of A and B free in solution at equilibrium and are not equal to the concentrations of A and B initially mixed together, as some of each will be used to form the AB complex, such that:

$$[A_{eq}] = [A_0] - [AB_{eq}] \quad \text{(Equation 7)}$$

Where $[A_0]$ is the initial concentration of A in the solution. The same equation holds for B.

A simple method to determine the K_d of two molecules is to prepare a binding curve: a series of mixtures of A and B are allowed to reach equilibrium, with one concentration varying while the

other is fixed. By measuring the concentration of A_{eq}, B_{eq} or AB, a plot of AB versus the concentration of the varied molecule creates a binding curve that can be fit to a given theoretical model. In this thesis, AB concentration will be inferred by the increase in fluorescence when the two halves of the biosensor bind. In the ideal case, the concentration of the fixed molecule, for example A, will be much lower than the K_d. Under these conditions, the amount of B used to form the AB complex will be negligible across all concentrations of B, such that $[B_{eq}] \approx [B_0]$. This simplifies the interpretation of the binding curve, as $[A_0]$ and $[B_0]$ are known from the preparation of the experiment, and so an additional step of calculating $[B_{eq}]$ at each point is unnecessary. A range of concentrations of B are then tested that include concentrations both above and below the K_d. For concentrations of B much greater than the K_d, A will be saturated, and the concentration of AB will reach a plateau where effectively all of A is bound, and the addition of further B does not measurably change the concentration of AB. By substituting Equation 7 into Equation 6, we can show that:

$$\frac{[AB_{eq}]}{[A_0]} = \frac{[B_{eq}]}{K_d + [B_{eq}]} \qquad \text{(Equation 8)}$$

Where $[AB_{eq}]/[A_0]$ is the fraction of total A bound. The equation describes a hyperbola going from 0 bound A when no B is added, to the plateau at 1, when B is present in a concentration well above the K_d. A useful result of this equation for visual analysis of the binding curve is that when $[B_{eq}] = K_d$, the fraction of A bound is equal to 0.5. Thus, the concentration where the curve reaches half its maximum value is equal to the K_d. For a more accurate determination of the K_d, equation 8 can be fit to the data using a regression method, such as nonlinear least-squares.[157]

It is not always experimentally possible to keep the concentration of A well below the K_d, for example in the case of a tight-binding interaction with a nanomolar K_d, A might need to be in the same nanomolar range in order to accurately measure the signal produced by the AB complex. In this case, the concentration of B at equilibrium will no longer be equal to the initial concentration of B, a

phenomenon referred to as ligand depletion, and so [B_{eq}] will also need to be calculated while fitting the data.[5] Substituting Equation 7 for both A and B into Equation 6 gives the following relationship:

$$K_d = \frac{([A_0]-[AB_{eq}])([B_0]-[AB_{eq}])}{[AB_{eq}]} \quad \text{(Equation 9)}$$

Which can be rearranged into the following quadratic equation:

$$[AB_{eq}]^2 - (K_d + [A_0] + [B_0])[AB_{eq}] + [A_0][B_0] = 0 \quad \text{(Equation 10)}$$

By applying the quadratic formula, we arrive at an analytical solution for [AB_{eq}], of the form:

$$[AB_{eq}] = \frac{(K_d+[A_0]+[B_0]) - \sqrt{([K_d+[A_0]+[B_0]])^2-4[A_0][B_0]}}{2} \quad \text{(Equation 11)}$$

The other root of this equation, which takes the sum of all the terms in the numerator, returns a value of [AB_{eq}] greater than the total concentration of A or B, and therefore is not a physically possible solution. Dividing both sides by [A_0] gives the equation for fractional saturation, similar to Equation 8 for the case without ligand depletion, which can once again be fit to the measured data:

$$\frac{[AB_{eq}]}{[A_0]} = \frac{(K_d+[A_0]+[B_0]) - \sqrt{(K_d+[A_0]+[B_0])^2-4[A_0][B_0]}}{2A_0} \quad \text{(Equation 12)}$$

Another potential complexity when measuring equilibrium binding is the possibility of multiple binding events, such that:

$$A_n + nB \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} A_nB \quad \text{(Equation 13)}$$

In this example, n molecules of B are able to bind simultaneously to each molecule of A, changing the equilibrium equation to:

$$K_d = \frac{[A_{n\,eq}][B_{eq}]^n}{[A_nB_{eq}]} \quad \text{(Equation 14)}$$

Which rearranges to:

$$\frac{[A_n B_{eq}]}{[A_{n\,0}]} = \frac{[B_{eq}]^n}{K_d + [B_{eq}]^n} \qquad \text{(Equation 15)}$$

Equation 15 is known as the Hill equation,[158] of which equation 8 is simply a special case where n = 1. n is the Hill coefficient of sigmoidicity, or simply Hill coefficient, and graphically represents the deviation of the data from the previous hyperbolic curve, towards a sigmoidal shape. In theory the value of n should be equal to the number of binding sites available on A_n, although in practice this is not always the case, as n is generally less than the number of binding sites. It is instead broadly interpreted as the extent of cooperativity between each site, i.e. the amount by which affinity for B increases each time a molecule of B binds to A_n. In the case of perfect positive cooperativity, the binding of a single molecule of B to A_n leads to the immediate simultaneous binding of all possible molecules of B, which gives the binding scheme represented in Equation 13, with a Hill coefficient of n. As the Hill equation is simply a more general form of equation 8, it can be fit to data in the same fashion, simply by adding n to the variables being optimized.

The above equations all apply to systems that are at equilibrium, and suffice to determine the affinity of two molecules for one another, as expressed by the K_d. But no information can be gained about the individual rate constants of the forward and reverse reactions using equilibrium data, only their relative values. To obtain information about the individual rate constants, kinetic experiments must be used.[159] In a kinetic experiment, a system is perturbed, for instance by mixing two solutions of A and B together. The change in the system over time as it approaches the new equilibrium is then followed. When measuring the rate of change of the concentration of an AB complex, this rate is equal to the difference between its rate of formation and dissociation:[5]

$$\frac{d[AB]}{dt} = k_{on}[A][B] - k_{off}[AB] \qquad \text{(Equation 16)}$$

For the reversible binding of two molecules A and B that we have been considering, under pseudo-first order conditions where A is in large excess relative to B (and so $[A_{eq}] \approx [A_0]$) this equation can be combined with Equation 6 and integrated to find the exponential time course:

$$[AB] = \left([AB_0] - \left[AB_{eq}\right]\right)e^{-(k_{on}[A]+k_{off})t} + [AB_{eq}] \qquad \text{(Equation 17)}$$

Where AB_0 is the initial concentration of AB complex, usually 0 for an experiment begun by mixing A and B together. Because of the pseudo-first order assumption, the concentration of A in this equation is a constant, and a fit of equation 17 to kinetic data will determine an observed rate constant:

$$k_{obs} = k_{on}[A] + k_{off} \qquad \text{(Equation 18)}$$

Extracting k_{on} and k_{off} can then be done in multiple ways. Because Equation 18 is linear, a plot of k_{obs} versus [A] for multiple concentrations of A will have a slope of k_{on} and an intercept of k_{off}. Alternatively, if K_d has already been determined from an equilibrium binding curve, Equation 6 and Equation 18 form a system of two equations with two unknowns, which can then be solved to obtain values for the two rate constants.

As with the equilibrium binding assays, it is not always possible to prepare an experiment where the concentration of A will be constant over the course of the experiment. If A is not added in sufficient excess over B (generally accepted to be a minimum of ten-fold), then the pseudo-first order assumption does not hold, and Equation 18 will not be accurate, as it fails to account for the change in concentration of A over time. Instead, the full integrated rate equation can be used, calculated by Hulme and Trevethick to be:[5]

$$[AB] = \frac{xy(e^{(x-y)k_{on}t} - 1)}{(xe^{(x-y)k_{on}t} - y)} \qquad \text{(Equation 19)}$$

Where x is defined as $[AB_{eq}]$, calculated using Equation 11, and y is defined as $[A_0][B_0]/[AB_{eq}]$. If K_d has been previously determined using a binding curve, then $[AB_{eq}]$ can be calculated directly, and

Equation 19 can then be fit to the data in the same way as Equation 17 to obtain a value for k_{on}. k_{off} can then be calculated using Equation 6. In this manner, both rate constants can be obtained, and the timescale on which binding occurs can be determined. This is of particular relevance for our biosensor, as the goal is to have it interact on a sub-minute timescale, and simply measuring the equilibrium K_d cannot differentiate a system with slow binding and slow dissociation from a system with fast binding and fast dissociation.

1.3.3 Flow Cytometry and Fluorescence-activated Cell Sorting

Flow cytometry is a method to obtain information about individual cells in a culture, by using microfluidics to pass them individually in front of a light source.[160] Light scattered from these cells can then be detected and correlated to obtain information about cell size and shape, and fluorescence emitted by the cells can also be detected when appropriate excitation sources are used. The power of this method comes from the ability to measure characteristics of individual cells at a rapid rate, as modern flow cytometers can interrogate thousands or even tens of thousands of cells per second. This allows researchers to obtain sample sizes far greater than would be possible from standard microscopy techniques, improving the statistical accuracy of measurements, and increasing the chance of detecting rare events.

In order to focus a suspension of cells into a flow with a small enough cross-section that individual cells can be distinguished, flow cytometers make use of hydrodynamic focusing.[161] Inside the instrument, the cell suspension is injected into the middle of a stream of sheath fluid. This sheath fluid adopts a laminar flow profile, preventing mixing of the fluids and restricting the cell suspension, or core, to the center of the flow. The relative areas of the sheath and core fluids can be controlled by varying their relative flow rates, as the volumetric flow rate of a liquid is equal to the product of the cross-sectional area and the velocity, and the velocities of the sheath and core fluids are not independent due to friction between the two, characterized by their viscosities. This allows the diameter of the core

fluid to be kept small enough that cells will generally pass in single-file, and so be detected individually, without needing to use tubing with an internal diameter on the order of a single cell, which would easily clog.

Once the cells are arranged in single-file, they pass by a light source to be interrogated. In most flow cytometers, the light source used for illumination is a laser, as they provide intense light focused in a narrow beam suitable for single-cell illumination.[160] Because a single laser emits light at a very specific wavelength, many flow cytometers include multiple lasers, in order to excite different fluorophores. For example, Beckman-Coulter currently offers solid-state lasers emitting at 375, 405, 488, 561, 638 and 808 nm in their CytoFLEX S line of research flow cytometers, with up to four of these lasers available in a single instrument. The 488 nm laser is particularly common, as it efficiently excites fluorescein dye, as well as EGFP.[43]

When the cells pass by the light source, a number of detectors record the interaction between the two. Light that is scattered at small angles from the incident light (0.5-2.0°) is referred to as forward scatter. To detect this scattered light, a detector is placed on the opposite side of the fluid stream to the light source, and an opaque obscuration bar is placed in front of it, such that unscattered light arriving directly from the light source is not detected. The amount of forward scattered light is roughly proportional to the size of the particle, but this is not a direct correlation, and cannot generally be used to accurately determine cell size. It can also be affected by the wavelength of light used, as well as the difference in refractive index between the cell and the fluid, such that dead cells with permeable membranes tend to scatter less than live cells of equal size.[162] On the other hand, light that scatters at wide angles to the incident light (roughly 90°) is referred to as side scatter. This scatter generally arises from irregularities in the surface or cytoplasm of the cell, and can thus be used to differentiate certain types of cells from one another, or to exclude aberrant cells from a population. Fluorescence detection is also measured at a wide angle from the incident light, and makes use of wavelength specific filters

and mirrors to detect only light in a given range of wavelengths, corresponding as much as possible to the emission spectrum of the fluorophore (Figure 1.10).[163]

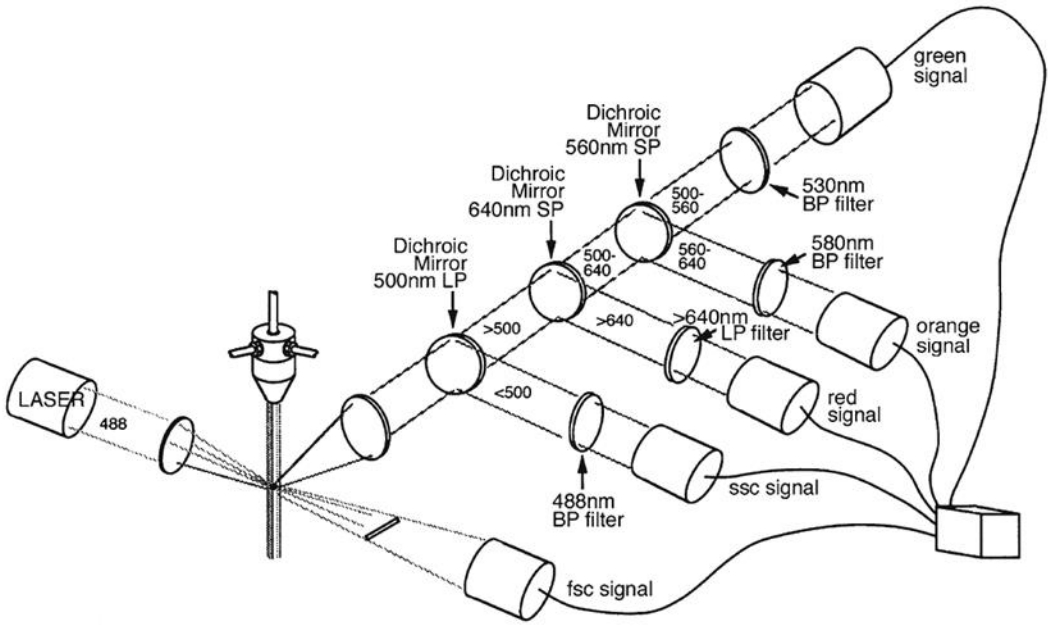


Figure 1.10 Detector configuration for flow cytometry. An example flow cytometry system with a single excitation laser interrogating a flow of cells, showing the detectors for forward scatter (fsc), side scatter (ssc) and fluorescence emission of different colours (red, orange and green). Numbers used represent the wavelengths of light, in nm, in a given beam. Filters and mirrors used are either bandpass filters (BP) that allow wavelengths around a certain value to pass, longpass filters (LP) that allow wavelengths greater than a certain value to pass, or shortpass filters (SP) that allow wavelengths shorter than a certain value to pass. (Reprinted with permission of John Wiley & Sons, Inc. Copyright 2001 from Givan, A. L. [2001] *Flow Cytometry: First Principles*, 2nd edition. Wiley-Liss, New York, NY.).[2]

For each of these signals, detectors measure the light pulse over the time that the cells spend in the beam, and record three descriptors: the height, or maximum intensity detected during the pulse, the width, or duration of the pulse, and the area, or total integrated intensity of the pulse.[164] All of these values are recorded for each cell, allowing the researcher to view the recorded data at the population level, usually in the form of a histogram, when visualizing a single parameter, or a density or contour plot, when visualizing two parameters.

Often, not all detected cells will be of interest for a particular experiment. For instance, some cells in a culture may be dead, or particles of dust in the micron and submicron range of diameters[165]

may be detected, as they will also scatter light. Because of this, each individual detection of a particle passing through the beam of the light source is referred to as an event, rather than a cell. To exclude undesired events from further analysis, software gates are used. A gate is a population of events selected based on some criteria, which can then be analyzed separately from the whole population. For instance, a control experiment with no cells present may provide a range of forward and side scatter values that correspond to dust particles in the buffer, and so a gate can be drawn that excludes this range, selecting only events that correspond to cells. A gate can also be drawn to exclude doublets, i.e. two cells stuck together, by making use of the fact that the height, width and area of the light pulse from an event can be tracked. Because the diameter of the core fluid is not much larger than the diameter of a single cell, doublets tend to orient along the axis of flow. This leads to an event that is roughly twice as long, and therefore has twice the width, but with the same height, and thus twice the area. By using a gate that excludes cells with high values of area relative to height, doublets can be excluded from the count, enriching the data in single-cell values.[166]

Once the populations of interest have been identified, they can then be analyzed. These analyses can be based on comparisons of the sizes of different populations of cells, for instance to determine what fraction of bacteria in a culture are viable,[167] contain antigens for a known antibody conjugated with a fluorescent dye, or otherwise have a phenotype that can be interrogated using a fluorophore.[168] In the particular case where it is desirable to collect the cells from a population of interest, in order to perform other experiments with them, fluorescence-activated cell sorting (FACS) can be performed. A FACS instrument functions on the same principles as a flow cytometer, but instead of sending all the cells to a waste container after interrogating them, it has the capacity to split off individual cells into another container for collection. An example of how this can be done is by droplet sorting, in which the stream is vibrated after interrogation, such that it splits into droplets small enough that each cell is enclosed in an individual drop. Because the flow rate of the cells is controlled, the time between

the interrogation of a cell by the light source and its enclosure in a droplet is known, allowing the droplets containing cells of interest to be determined. These droplets are then positively or negatively charged by the application of a voltage pulse, and as the stream of droplets passes by a charged deflection plate, the selected droplets are either attracted or repelled by the plate, sending them into a different collection container from the other, uncharged droplets, which pass directly into waste.[160]

FACS is an effective screen to identify and collect bacteria with improved phenotypes, so long as a method exists to correlate this phenotype with a fluorescent signal, and has been used in such a capacity for decades.[169] Its high-throughput nature, with the potential to screen millions of cells, makes it an excellent tool for the directed evolution of fluorescent proteins, i.e. the screening of large libraries of random or semi-random mutants to identify improved variants.[170] One of the most successful variants of the *Aequorea victoria* GFP, EGFP, was discovered with the assistance of FACS in 1996,[59] and FACS has also been part of the development of the well-known mFruits line of fluorescent proteins,[46] other red fluorescent proteins,[171, 172] FRET pairs,[173] and at least one single-molecule fluorescent protein-based biosensor.[174] Reciprocally, fluorescent proteins are also fine reporters for flow cytometry applications in a variety of cells,[175] with multiple colours available to study protein expression.[176] With good experimental design, more complex studies can also be performed, including, but not limited to: alternative gene splicing,[177] protein-protein interactions,[178] cell tracking,[179] and indirectly reporting on enzymatic activity.[97] Thus, fluorescent proteins and flow cytometry are tools that go hand-in-hand, each complementing the other in biological assays.

1.4 Thesis objectives

In Chapter 1, an overview of fluorescent proteins has been presented, along with a particular case where their usage is limited: when the detection window of the experiment coincides with the maturation of the chromophore. We have mentioned our goal of developing a two-part biosensor that would avoid this problem by expressing a circularly permuted fluorescent protein ahead of the protein

of interest, providing time for maturation to occur, then detecting a change of fluorescence in the presence of the protein of interest. In order to create this change in fluorescence, a protein-protein interaction between the protein of interest and the biosensor must be introduced, for which we selected the Bcl-2 family of proteins to use as fusion tags, based on their relatively small size and high affinity. Finally, we have introduced the techniques that will be used to characterize this biosensor, mainly equilibrium binding curves using fluorescence as a reporter on binding, as well as flow cytometry on cells expressing both halves of the sensor.

In Chapter 2, we will describe the initial development and testing of our biosensor, which we name a Sensor for Transiently Expressed Proteins, or STEP. We demonstrate that the interaction of a Bcl-x_L based tag with a Bim peptide that has been fused to a circularly permuted GFP causes an increase in green fluorescence, and thus that the biosensor does function both *in vitro* with purified proteins, and in *E. coli* cells. We also show that the tag can be fused to a mock protein of interest without greatly affecting the affinity or dynamic range of the sensor.

In Chapter 3, the process of improving the STEP will be described, including both rational design hypotheses tested as well as some high-throughput screening using FACS. The current best sensor that we have developed is named gSTEP1 and we present its characterization there, including equilibrium and kinetic assays, both *in vitro* and in *E. coli*.

Chapter 4 will expand on the available options for the STEP, encompassing attempts made to develop biosensors using different protein-protein interaction pairs, as well as different colours of fluorescent proteins. The overarching goal is to eventually allow multiplexing of these sensors, such that multiple proteins could be detected at once.

Finally, Chapter 5 will summarize the state of the STEP, and provide some direction for future engineering of this protein biosensor. Options for both rational design and directed evolution will be

discussed, followed by some potential applications that are being explored. A number of fluorescence-based protein detection systems have been developed, and their advantages and disadvantages relative to our biosensor will be explored.

Chapter 2: First STEPs: Design of a Fluorescent Sensor for Transiently Expressed Proteins

Matthew G. Eason, Antonia T. Pandelieva, Safwat T. Khan, Hernan G. Garcia & Roberto A. Chica
Manuscript in preparation

2.1 Statement of Contribution

Matthew G. Eason designed all protein sequences. Matthew G. Eason and Safwat T. Khan performed cloning experiments. Matthew G. Eason and Antonia T. Pandelieva performed characterization experiments and analyzed data. Matthew G. Eason performed flow cytometry. Roberto A. Chica and Hernan G. Garcia conceived the project. Matthew G. Eason wrote the chapter.

2.2 Introduction

Proteins are involved in a wide variety of tasks in the cell, including enzymes catalyzing metabolic processes, transcription factors regulating transcription, structural proteins comprising the cytoskeleton, and more. Thus, the study of proteins inside living cells requires robust methods that can specifically identify proteins of interest in this complex environment. One such method, which has been used with great success, is that of genetically fusing a fluorescent protein (FP), generally a homolog of the green fluorescent protein,[53] to a protein of interest. The use of these natively fluorescent proteins as fusion tags provides excellent spatial and temporal resolution of protein localization. Using fluorescence microscopy, the location of the tagged proteins can be tracked in real time, potentially at resolutions in the tens of nanometers when using super-resolution microscopy techniques and compatible FPs.[180-183]

In order to form the fluorescent chromophore, the FP tag must undergo an autogenic post-translational modification known as maturation. During this process, the tripeptide sequence of the chromophore undergoes cyclization, dehydration, and oxidation in order to form the conjugated system required for fluorescence.[3] As these reactions are catalyzed by the surrounding protein structure,[26] maturation does not begin until after the FP has folded, leading to a delay between protein folding and

fluorescence, commonly quantified as a half-time. *In vitro*, this maturation half-time can be as low as 2 minutes in engineered FPs,[44] however, for most green FPs it is closer to 30 minutes.[94, 96] Depending on the organism studied, this maturation delay can become even longer *in vivo*,[10, 184] which leads to a period of time immediately post-translation where an FP fusion tag is incapable of reporting on the protein of interest. This results in a loss of information when reporting on transiently-expressed proteins, such as the Hes proteins that drive the somitogenesis clock in zebrafish,[185] or Insulin-like growth factor-binding protein 1 (IBP-1), which is rapidly secreted from the cell.[186] Over 160 proteins with half-lives less than 4 minutes have been identified in yeast,[187] comprising roughly 4% of the tested proteins, and including transcription factors, mRNA splicing factors and cell cycle proteins. Thus, transiently-expressed proteins are involved in regulating the growth and development of organisms, yet little information on their expression and localization can be gained using conventional FP tags, as the majority of FP-tagged molecules will no longer be present in the cell by the time fluorescence would otherwise be detected. Maturation delay can also introduce error into experiments that compare temporal information from FP-tagged protein expression to other cellular processes, such as mRNA transcription.[12, 92, 188]

To address this gap in temporal information, we have created a sensor for transiently expressed proteins (STEP). The STEP is based on a circularly permuted, green fluorescent protein (cpGFP), which can fold and mature independently of the expression of the protein of interest (POI). This cpGFP is derived from the calcium ion sensor GCaMP3; the N- and C-termini are located next to the chromophore, resulting in a pore on the GFP surface.[86] Water enters the core of the protein via this pore and quenches fluorescence, resulting in a dim signal. A short peptide from the BH3 domain of Bim, a Bcl-2 family protein,[189] is genetically fused to the N-terminus of cpGFP, creating a construct we call gSTEP, for green-STEP. The Bim peptide enables the specific binding of a protein tag, which we call STEPtag, based on another Bcl-2 family protein, Bcl-x_L. The binding of STEPtag to gSTEP

plugs the pore in the cpGFP, eliminating quenching and restoring bright fluorescence (Figure 2.1). By pre-expressing gSTEP in the organism of interest, the chromophore can be given time to mature independently, such that it is ready to detect STEPtag fused to the POI the instant it is expressed. This sensor then operates on the timescale of protein-protein binding, rather than chromophore maturation. It should therefore enable detection of a tagged POI in seconds, with high spatiotemporal resolution, opening the door to the study of protein localization immediately after translation.

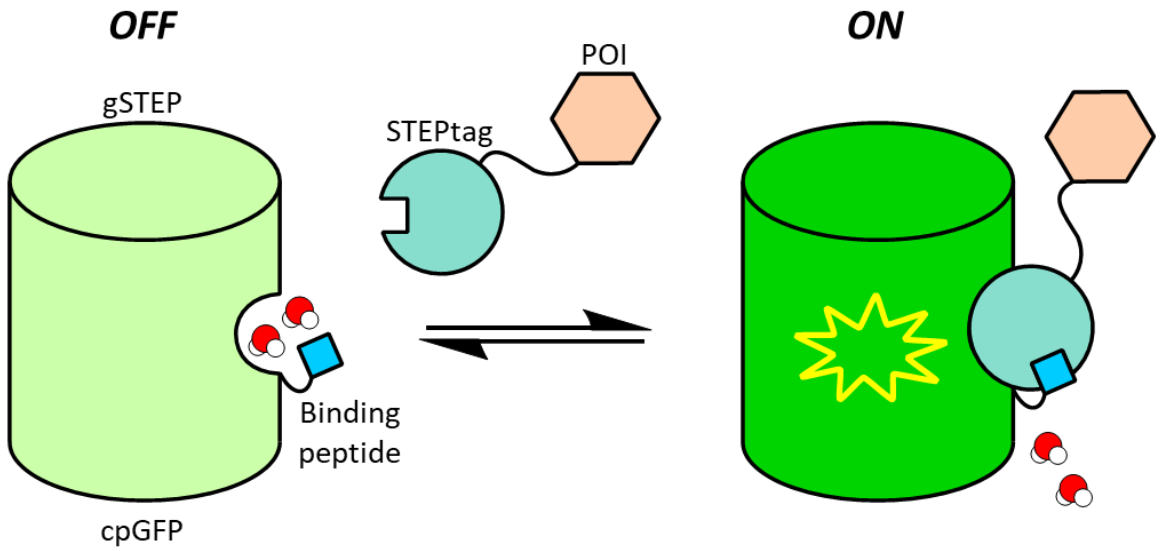


Figure 2.1 General schematic of the STEP. The STEP is comprised of a circularly permuted fluorescent protein, in this case a green fluorescent protein (cpGFP), to create a green STEP (gSTEP). A binding peptide is fused to the terminus of the fluorescent protein, next to the pore in the barrel. This peptide binds specifically to a binding partner, termed STEPtag, forming a binding pair. The pore created by circular permutation exposes the core of the cpGFP to solvent, quenching the chromophore and resulting in a dim fluorescent signal (OFF, above). When the binding partner binds to the binding peptide, the resulting complex covers the pore, restoring a bright fluorescent signal (ON, above). By genetically fusing STEPtag to a protein of interest (POI) and pre-expressing gSTEP, expression of the POI can be detected by following the increase in fluorescence over time.

2.3 Results

2.3.1 Preparing the Scaffold

As the GCaMP family of calcium ion sensors operates by varying the fluorescence intensity of a circularly permuted fluorescent protein in response to a protein conformational change that is

triggered by a binding event, in their case the binding of calcium ions, it was a natural choice to begin with one of their cpGFPs for the STEP. We selected the cpGFP from GCaMP3 to begin with, as that version of GCaMP is brighter, more stable and has a greater dynamic range than the previous generations.[86] We then selected the Bcl-2 family of binding proteins to serve as the binding pair,[125, 190] specifically because of their extremely tight binding interaction (< 1 nM K_d values),[119] but also because of the small size (26 amino acids) of the BH3 domain required to bind to the receptor protein.[189] This domain is comprised of a helix that is similar in size and structure to the M13 peptide used in GCaMP to guide the calmodulin domain to the pore, and we hypothesized that it would perform a similar function in the STEP, without being so large as to block the pore in the absence of binding partner. We chose the Bim BH3 domain as our binding peptide and Bcl-x_L as the binding protein, as their interaction has been characterized,[191] and a crystal structure of the bound state was available.[192] The Bim peptide of this crystal structure was then aligned to the location of the M13 peptide of the GCaMP2 crystal structure[11] using PyMOL (Figure 2.2), and it was estimated that Bcl-x_L binding could occur adjacent to the pore. Based on this inspection, we selected two potential sequences for the binding tag: either the soluble domain of Bcl-x_L,[193] which we named exSTEPtag (extended STEPtag), or an N-terminal truncation of this domain, spanning the α2 helix to the α8 helix, named STEPtag. The truncation at the N-terminal region to create STEPtag reduced the overall size of the tag by removing a 58 amino acid loop between the α1 and α2 helices that is unnecessary for protein binding,[127] as well as the α1 helix that can participate in the formation of a domain-swap dimer when the aforementioned loop is removed.[194, 195] Both options included a C-terminal truncation relative to the wild-type sequence that removed a hydrophobic, membrane-anchor domain in order to improve solubility and prevent the tag from affecting localization of the POI.[196, 197]

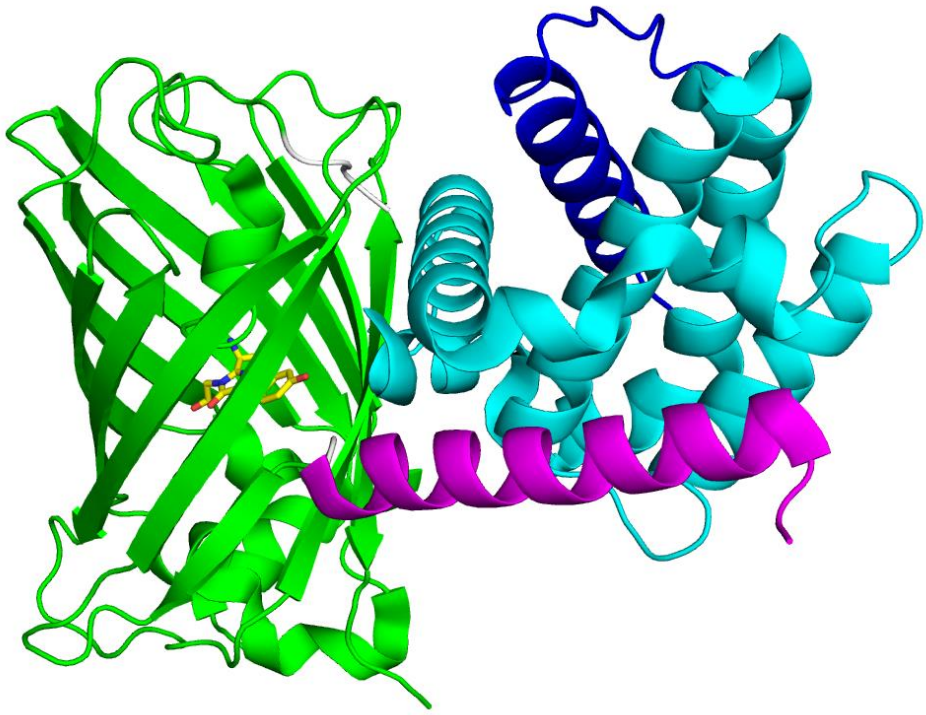


Figure 2.2 Hypothetical binding of STEPtag next to the pore of cpGFP in the STEP. A model of the bound STEP was created by overlaying the crystal structures of GCaMP2 (PDB: 3EK4) and Bim bound to Bcl-x_L (PDB: 1PQ1). The cpGFP domain of GCaMP2 is shown in green, with the chromophore shown as yellow sticks. Note the pore in the β-barrel between the two termini (white), which exposes the chromophore. The Bim peptide (magenta) was aligned to the N-terminus of the cpGFP, replacing the M13 peptide that was present. The portion of Bcl-x_L used to create STEPtag is shown in cyan, with the α1 helix included in exSTEPtag shown in blue. The loop between the α1 and α2 helices in exSTEPtag is not shown due to lack of electron density in the crystal structure.

Having selected the individual parts that make up the STEP, the next stage was to assemble them into a functional sensor. Firstly, to verify that the fully bound STEP would indeed have a fluorescent signal greater than that of the unbound STEP, we created a series of constructs with both the Bim peptide and either STEPtag or exSTEPtag fused to opposite termini of the cpGFP, thus simulating a fully bound sensor. Different lengths of flexible glycine-serine linkers were used to vary the conformational freedom available to the binding pair, as well as the inter-domain distance between them and the cpGFP barrel. These linkers ranged in length from zero amino acids, using only the leucine-glutamic acid and threonine-arginine linkers already present on the GCaMP-derived cpGFP sequence, up to 30 amino acids, to approximate untethered binding of the binding pair.[189] These

unimolecular constructs were expressed, and their fluorescence was compared to that of cpGFP (Table 2.1). We found that while every construct had a similar emission peak around 515 nm, the excitation spectra varied across different constructs, with some showing greater signal when excited at 395 nm, and others showing greater signal when excited at 495 nm. These two excitation wavelengths correspond to the protonated (neutral) and deprotonated (anionic) forms of the chromophore, respectively, with excited-state proton transfer occurring in the neutral form to produce a similar emission peak to that of the anionic one.[53] The sequences that showed improved signal from the 395 nm excitation peak were interesting, as in GCaMP2 this protonated chromophore signal decreases as a result of the binding event, and thus a bright signal at 395 nm would be indicative of exposed chromophore.[11] As cpGFP is expected to already have a solvent-exposed chromophore, the sequences with increased signal when exciting at 395 nm must therefore be further stabilizing the protonated form of the chromophore. Thus, depending on the sequence tested, both the protonated and deprotonated states of the chromophore could potentially be stabilized by the bound STEP. As the 495 nm excitation peak is more conducive to imaging with a standard 488 nm laser line, we chose to focus on the sequences with increased fluorescence at that wavelength. When looking at the 495 nm signal, it is clear that increasing the linker length is detrimental, with the maximum change in fluorescence (expressed as the difference in fluorescence normalized to the baseline signal from an independently expressed cpGFP, $\Delta F/F_0$) being seen in the two constructs with no added linkers, Bim-L0-cpGFP-L0-STEPtag, and exSTEPtag-L0-cpGFP-L0-Bim. This seems to indicate that the two binding partners are sufficient to interact with the chromophore, and that keeping them close to the pore should improve the sensor signal when fully bound. As the increase in fluorescence was greatest in the construct with the smaller STEPtag, we chose it as our scaffold for creating the two halves of the STEP, naming it uni-gSTEP0, for unimolecular green STEP, initial iteration.

Table 2.1 Screening the unimolecular sequences.

Sequence[a]	$\Delta F/F_0$[b] (λ_{ex} 395 nm)	$\Delta F/F_0$[b] (λ_{ex} 495 nm)
Bim-L0-cpGFP-L0-STEPtag[c] (uni-gSTEP0)	-0.2 ± 0.2	3.1 ± 0.8
Bim-L5-cpGFP-L5-STEPtag	-0.1	0.1
Bim-L10-cpGFP-L10-STEPtag	-0.1	0.2
Bim-L30-exSTEPtag-L5-cpGFP	-0.1	-0.1
Bim-L30-exSTEPtag-L10-cpGFP	-0.2	0.1
exSTEPtag-L0-cpGFP-L0-Bim	-0.6	1.0
exSTEPtag-L5-cpGFP-L5-Bim	-0.4	0.9
exSTEPtag-L10-cpGFP-L10-Bim	-0.4	0.6
cpGFP-L0-Bim-L30-exSTEPtag	3.8	-0.6
cpGFP-L5-Bim-L30-exSTEPtag	2.7	-0.2
cpGFP-L10-Bim-L30-exSTEPtag	2.6	-0.2

[a] Sequences are named based on their arrangement of the three sensor parts (cpGFP, Bim and STEPtag), from the N- to the C-terminus. Linkers between parts are denoted Lx, where x is the number of amino acids

[b] Difference between the peak signal of the construct and a baseline signal (F_0), normalized to the baseline signal. In this case, F_0 is the fluorescence of an equivalent concentration of cpGFP, purified separately

[c] Error shown is the standard deviation of three biological replicates

2.3.2 Creating the STEP

Having created uni-gSTEP0 and demonstrated that it was brighter than cpGFP, we then split this construct at the linker between the cpGFP barrel and STEPtag, to create the two parts of the sensor. As the linkers connecting the cpGFP to the binding peptide and binding protein in the GCaMP family of biosensors have been shown to be important to the sensor signal,[82, 198] we left the threonine-arginine linker from GCaMP3 attached to the cpGFP construct, creating gSTEP0. We found that we could detect the addition of saturating concentrations of STEPtag to gSTEP0, while exciting at 485 nm, with a $\Delta F/F_0$ of 1.4 ± 0.1 at the emission peak, signifying that the brightness of the sensor more than doubles

when STEPtag is present. The *in vitro* equilibrium binding of the sensor was also measured, with an observed K_d of 240 ± 50 nM (Figure 2.3).The peak emission of uni-gSTEP0 was found to be roughly twofold brighter than the peak emission of the saturated gSTEP0, implying that the observed binding interaction was not producing the maximum possible increase in signal. This result, coupled with a noticeable amount of background fluorescence of unbound gSTEP0, leads to a relatively low signal-to-noise ratio, expressed here as the $\Delta F/F_0$.

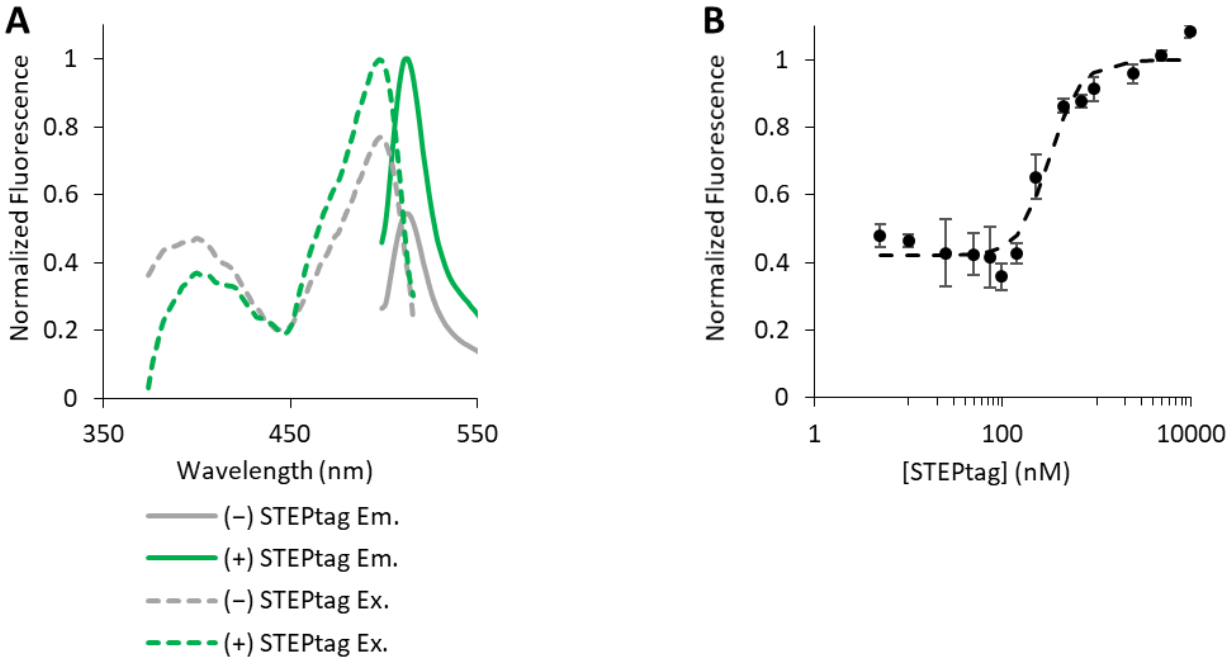


Figure 2.3 Characterizing gSTEP0 binding. A, The excitation and emission spectra of gSTEP0 (75 nM) alone and in the presence of 2500 nM STEPtag. Spectra are averages of three biological replicates of gSTEP0, and two biological replicates of STEPtag. **B,** A representative binding curve of gSTEP0 (75 nM) with a range of concentrations of free STEPtag. The dashed line indicates a fit of the Hill equation to the data. Error bars are the standard deviation of triplicate measurements.

2.3.3 Rational Improvement of $\Delta F/F_0$ by Enlarging the Pore

To improve the $\Delta F/F_0$, we considered two possible options: increasing the brightness of the bound state, by improving the interaction of the binding partner with the chromophore of gSTEP0, or decreasing the brightness of the unbound state, by further quenching the chromophore when the pore is open. We chose to begin with the second option, as we hypothesized that it could be accomplished simply by truncating the C-terminus of the sensor. The C-terminal linker sequence has been shown to

contribute to the high brightness of GCaMP5,[198] so we hypothesized that its presence in the STEP could be increasing the baseline fluorescence of our sensor, lowering the observed $\Delta F/F_0$. Thus, we generated C-terminal truncations of gSTEP0, removing the C-terminal linker sequence that was left over from GCaMP3, as well as a further 1-4 amino acids. These constructs were labeled gSTEP0-Tx, where x is the number of cpGFP amino acids truncated, and we found that gSTEP0-T1 -T2 and -T3 showed improved $\Delta F/F_0$, with the K_d of STEPtag binding remaining mostly unchanged (Table 2.2). gSTEP0-T4, on the other hand, was poorly soluble, remaining mostly in the pellet fraction during purification (Figure S2.1), and also showed less than 20% of the brightness of the other constructs. We chose to continue working with gSTEP0-T1, as it showed the greatest improvement in $\Delta F/F_0$, while maintaining a similar K_d relative to gSTEP0.

Table 2.2 Characterizing truncations of gSTEP0.

Sequence	K_d (nM)	$\Delta F/F_0$ (Ex. 485 nm)
gSTEP0[a]	240 ± 50	1.4 ± 0.1
gSTEP0-T1[a]	250 ± 60	2.2 ± 0.5
gSTEP0-T2	210	1.7
gSTEP0-T3	230	1.9
gSTEP0-T4	N/A[b]	N/A[b]

[a] Data obtained from three biological replicates, error shown is the standard deviation
[b] No binding curves were performed for gSTEP0-T4, due to the low signal obtained

2.3.4 Detection of STEP Fluorescence in Bacterial Cells

As our *in vitro* data showed that binding of STEPtag to gSTEP0 and its derivatives could be readily detected in purified protein samples, the next step was to test the sensor inside of a live cell. As a proof-of-concept, we turned to flow cytometry of *E. coli* cells to demonstrate that cells expressing gSTEP0-T1 could be differentiated from cells expressing both gSTEP0-T1 and STEPtag. To allow gSTEP0-T1 to be expressed and mature in advance of the measurements, it was subcloned into pSF-

OXB20. In this vector, the gene is placed under control of the OXB20 promoter, which is based on the RecA promoter.[199] This plasmid places the gene of interest under the control of a strong constitutive bacterial promoter, allowing gSTEP0-T1 to be present in the cell at all times. STEPtag was kept under control of the T7 promoter[200-202] by cloning it into pACYC-Duet-1, a vector whose antibiotic resistance and origin of replication is compatible with pSF-OXB20. By comparing a population of cells containing just gSTEP0-T1 with a population containing both vectors, we can see a distinct increase in fluorescence when STEPtag is present (Figure 2.4), with the mean signal increasing twofold, i.e. an *in cellulo* $\Delta F/F_0$ of 1. Interestingly, a control experiment where the pACYC vector containing STEPtag remained uninduced showed similar levels of fluorescence to the experiment where it was induced (Figure S2.9). We therefore hypothesize that the leaky expression from the T7 system[135] produces sufficient STEPtag during the overnight growth of the cultures to bind gSTEP0-T1, causing the observed shift in fluorescence regardless of the overexpression of STEPtag. Attempts to repeat the experiment with shorter growth times did not show sufficient levels of green fluorescence from gSTEP0-T1, as the level of expression from pSF-OXB20 did not seem to be sufficiently high. Nevertheless, while this expression system did not allow us to visualize an individual population of cells before and after expression of STEPtag, we were able to differentiate populations containing STEPtag and gSTEP0-T1 from those that only contained gSTEP0-T1.

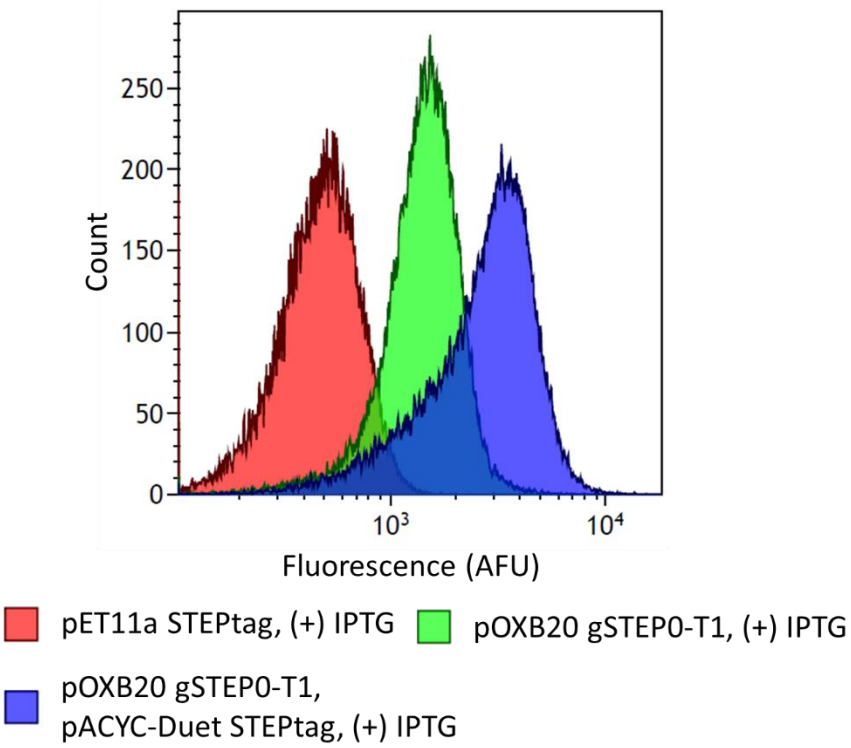


Figure 2.4 Bacterial flow cytometry of cells expressing the STEP. Flow cytometry histogram of STEP fluorescence in *E. coli* cells containing plasmids expressing either STEPtag alone, or gSTEP0-T1 and STEPtag. Fluorescence was excited at 488 nm, and emission was measured using a 525BP40 filter. STEPtag expression was induced by the addition of 1mM IPTG [(+) IPTG], 90 minutes before the cells were harvested for cytometry. The pET11a STEPtag negative control was expressed in BL21 (DE3) Gold cells, while the others were expressed in Tuner (DE3) cells.

2.3.5 *In vitro* Tagging of a Protein

In order to fulfill its intended purpose, the STEP needs to detect a protein of interest. Having demonstrated that gSTEP0-T1 can detect STEPtag alone, it remained to show that STEPtag could be fused to a POI without interfering with the sensor. As a mock POI, we chose to use *Thermoascus aurantiacus* xylanase 10A (TAX),[203, 204] as this enzyme forms a highly stable TIM-barrel[205] of moderate size (312 amino acids). It should thus be large enough to demonstrate whether tagged proteins can sterically hinder the binding interaction, while being stable enough to avoid any negative influence on the folding of the binding partner. While we expect transiently-expressed proteins to be less stable than TAX, the *in vitro* optimization of the STEPtag fusion required a POI that would be stable for the duration of the purification process. The active site glutamates, E131 and E237,[206] were mutated to

alanines, to prevent any enzymatic activity from interfering with the binding assay. Flexible glycine-serine linkers of different lengths were used to fuse STEPtag to TAX, from the four amino acid GSSG to the thirteen amino acid GSSGGGGSGGGGS. We also tested constructs with STEPtag fused at either the N- or C-termini of TAX (Table 2.3). All constructs are labelled from N- to C- terminus, with the number indicating the length of the linker. We found that the four, eight and thirteen amino acid linkers performed poorly, showing much greater K_d values for tags at both termini, with the shorter linkers also showing poor $\Delta F/F_0$. The fusions with linkers of ten amino acids, on the other hand, all performed comparably to STEPtag on its own, whether TAX was fused to the N- or C-terminus of STEPtag, and in fact showed slightly lower K_d values than were measured previously. We also verified that the sensor was specific to the binding partner, by demonstrating that the addition of untagged TAX or BSA to gSTEP0-T1 showed no increase in signal (Figure S2.8).

Table 2.3 Characterization of a mock POI-STEPtag fusion binding to gSTEP0-T1.

Binding Partner	K_d (nM)	$\Delta F/F_0$ (Ex. 485 nm)
TAX-L13-STEPtag[a]	350 ± 10	3.6 ± 0.9
TAX-L10-STEPtag	151	2.3
STEPtag-L10-TAX[a]	170 ± 30	1.9 ± 0.1
STEPtag-L8-TAX	310	1
STEPtag-L4-TAX	360	0.9

[a] Data obtained from two biological replicates, error shown is the standard deviation

2.4 Discussion

Our results show that we have created a fluorescent biosensor that can detect the protein-protein interaction between gSTEP0 and STEPtag *in vitro*, and that STEPtag can still be detected when it is used as a fusion tag at either the N- or C-terminus of a protein. The STEP can therefore in principle

detect any protein of interest that can be expressed as a fusion with STEPtag, by measuring the increase in fluorescent signal of gSTEP0 when the fusion protein is present.

By beginning the development of the sensor with a variety of unimolecular constructs, we were able to rapidly assess whether having STEPtag adjacent to the pore of the cpGFP affects brightness, and could also test a variety of different linker lengths and binding peptide positions without needing to optimize the peptide-protein binding interaction. We found that keeping the binding pair of the Bim peptide and the STEPtag close to the pore was important to increasing the brightness, as the constructs with no added linkers outperformed all others when comparing the signal from the anionic chromophore, excited at 485 nm. At the other end of the scale, the 30 amino acid linkers provide some interesting results. In the case of the cpGFP-Lx-Bim-L30-STEPtag constructs, which were designed with a long, flexible linker between Bim and the STEPtag to mimic the conformational freedom available to free STEPtag in solution, we see that there is an increase in signal from the neutral chromophore instead. This implies that the chromophore environment is different in these constructs relative to the others tested, increasing the pK_a of the chromophore phenol group. As the anionic form of the chromophore is preferable for imaging, these constructs were not selected for further development of the STEP.

The change in fluorescent signal of gSTEP0 upon binding to STEPtag follows the same trends as the difference in signal between cpGFP and uni-gSTEP0, as desired. As seen in the excitation spectra (Figure 2.3), some signal from the neutral chromophore is lost, while the signal from the anionic chromophore increases, similarly to how GCaMP2 changes in fluorescence upon calcium ion binding.[11] The current change in fluorescence upon binding of gSTEP0-T1 ($\Delta F/F_0 = 2.2 \pm 0.5$) approaches that of the original GCaMP calcium ion sensor, which demonstrated a $\Delta F/F_0$ of approximately 3.5.[82] As can be seen in the flow cytometry data (Figure 2.4), this is sufficient to use the STEP to discriminate an average cell containing no STEPtag from an average cell containing STEPtag,

although there is some overlap between the populations. While a kinetic analysis tracking fluorescence over time should still show the increase in fluorescence as STEPtag is produced, this overlap in brightness could lead to difficulties in the analysis of end-point data, as the brightest cells without STEPtag will show similar signal to the average cells with it present. Thus, engineering the STEP to improve the $\Delta F/F_0$ would improve its usefulness as a sensor. Through a combination of linker modification, site-directed mutagenesis and directed evolution, GCaMP has been engineered to obtain $\Delta F/F_0$ values greater than 35 *in vitro*,[87, 207] so we are optimistic that we will be able to improve this characteristic of the STEP.

As for the measured K_d of the STEP, 250 ± 60 nM is more than a hundredfold greater than the theoretical K_d of the Bim peptide and Bcl-x_L protein, possibly due to interference of the cpGFP domain with the ideal binding mode of the binding pair, or perhaps due to the truncations made to the sequence when creating the STEPtag. Although we hoped to avoid any formation of domain-swapped dimer by removing the α1 helix, it does contain the BH4 domain of the protein, which has been found to stabilize native Bcl-x_L by making hydrophobic contacts with the surrounding helices.[4] Thus, the fold of STEPtag might differ from the ideal Bcl-x_L structure, reducing the affinity of STEPtag for Bim. We did attempt to express full-length Bcl-x_L including both the α1 helix and the loop between the α1 and α2 helices, in the form of His-EK-Bcl-x_L,[208] as well as MBP-Bcl-x_L[191] to test this hypothesis, but in our hands these sequences did not express sufficient quantities of soluble protein to set up binding curves. Despite the lower than expected affinity of the STEP, its K_d remains at a feasible concentration for cellular proteins. For instance, protein concentrations in *S. cerevisiae* have been found to range from less than 50 to greater than 10^6 molecules per cell.[209] For an average cell volume of 42 μm^3,[210, 211] this leads to molar concentrations between 2 nM and 40 μM, with the median concentration around 100 nM. Thus, the affinity of the STEP should be sufficient to detect a variety of proteins of interest, as a similar range of concentrations is seen in other organisms as well.[212, 213] As a future improvement, a lower K_d would

be desirable to lower the limit of detection, in particular as we are interested in visualizing proteins shortly after translation begins, so it is likely that they will not yet have reached their maximum concentration. We plan to be able to improve the K_d through modification of the Bim-STEPtag binding pair, possibly by introducing synthetic versions of the Bim peptide with the potential for tighter binding.[120]

When considering the equilibrium binding of STEPtag to gSTEP0 and gSTEP0-T1, it is interesting to note that the Hill coefficient of the binding curves is around 2 (Figure S2.5, S2.6, S2.7), implying that there is some form of cooperative binding occurring[158] despite the fact that the expected stoichiometry of STEPtag binding to the gSTEPs should be that of the Bim-Bcl-x_L heterodimer, i.e. 1:1.[124] This result could potentially be due to a change in binding stoichiometry, if multiple molecules of STEPtag can bind to a single molecule of gSTEP, or it could be due to dimerization of the gSTEP itself, such that the binding of STEPtag to one gSTEP increases the affinity of the other. It has been reported that GFP forms a weak dimer, with a K_d in the 100 μM range,[214] while two possibilities for self-association of Bcl-xL have been reported, a domain-swap dimer that forms in the presence of detergents or high pH,[215] as well as an interaction between Bcl-xL and its own BH3 domain, with an IC50 of ≈640 nM.[216] As none of these possibilities seem likely to affect our binding curves, we have yet to specifically identify the cause of the observed cooperativity. It should be noted that the K_d values reported here still assume 1:1 stoichiometry, i.e. that only one molecule of STEPtag binds each molecule of gSTEP. While this apparent cooperativity lowers the dynamic range of the sensor for the purposes of quantifying STEPtag (and thus POI) concentration based on the fluorescent signal, it does not affect its use as a binary sensor. In order to detect transiently expressed proteins, it suffices that the sensor increases in signal once the tagged POI is expressed.

The STEP described here is a starting point to build towards an effective sensor for transiently expressed proteins in model organisms. By basing the STEP on previous FP-based biosensors, we are

harnessing the high spatial and temporal resolution of fluorescence measurements, as well as the utility and convenience of having a fully genetically encoded sensor, requiring no exogenous molecules to produce the desired signal. These characteristics have led to a rapid proliferation of genetically encoded fluorescent biosensors, with a recent review tabulating over 750 different variants,[78] but the examples of biosensors that detect other proteins tend to be engineered towards the detection of a particular protein of interest, rather than that of a fusion tag that can be used more generally. That being said, there exist at least two fluorescent biosensors for protein-protein interactions similar to the STEP, the dimerization-dependent fluorescent proteins,[217, 218] and the recently published Flashbody.[219] The dimerization-dependent fluorescent proteins were designed to probe for protein-protein interactions, where they use fluorescent protein-derived barrels as fusion tags for the two proteins whose binding is to be interrogated. One of the barrels, A, is fluorogenic, having low brightness initially, which increases 10- to 60-fold upon binding to the other barrel, B. Because this system was designed to identify proteins that already interact, the K_d was deliberately increased to the micromolar range to prevent the tags from creating false-positive interactions. As such, these dimerization-dependent fluorescent proteins tackle a different problem than the STEP, as the barrels cannot easily detect one another in the absence of additional forces to improve binding. The Flashbody, on the other hand, functions very similarly to the STEP, using a cpGFP fused to an affinity tag to detect binding by an increase in fluorescence. In this case, the Flashbody tags cpGFP with the variable region fragments of an antibody. This antibody binds to a peptide target, in this case the BGPC7 peptide, a 7 amino acid peptide that acts as the fusion tag to the protein of interest. The Flashbody shows a $\Delta F/F_0$ of roughly 3, as well as a K_d of 300 nM, similar to the STEP. Interestingly, as their antibody-peptide interaction has a theoretical K_d of 50 nM, they also see an increase in K_d relative to the original binding pair, matching what we see with the Bim-STEPtag interaction. Our selected binding pair has a fifty-fold lower theoretical K_d, and thus while the current Flashbody characteristics are comparable to those of gSTEP0-T1, we expect to be able to further improve the characteristics of our sensor. Improvements to the K_d and $\Delta F/F_0$ will open the door

to performing live-cell imaging of the STEP, which will allow us to reach our desired goal of obtaining rapid kinetic data for the study of transiently-expressed proteins.

2.5 Materials and Methods

2.5.1 Construct Preparation

Sequences for the 11 unimolecular constructs were codon optimized for *E. coli*, synthesized and cloned into pJ414 vectors by DNA2.0 (now ATUM). gSTEP0 and STEPtag sequences were obtained by PCR amplification of the appropriate regions of the uni-gSTEP0 sequence, using synthetic oligonucleotides (Eurofins) to add an N-terminal His tag to STEPtag, as well as NdeI and BamHI restriction sites flanking the gene. These sequences were then subcloned into pET11a vectors (Novagen), using the introduced NdeI and BamHI restriction sites. All reagents used for subcloning were obtained from New England Biolabs. gSTEP0 truncations were created in the same manner, using gSTEP0 as a template. A pET11a vector containing the gene encoding inactivated TAX was a generous gift from Stephen L. Mayo. STEPtag-TAX fusions were created by first using PCR to generate STEPtag sequences with the appropriate N- or C-terminal linkers, where the DNA sequence of the linker contains an XhoI restriction site. These sequences were then subcloned into pET11a vectors, in the same manner as above. In parallel, TAX constructs with the same linkers were also prepared, and could then be subcloned into the pET11a-STEPtag vectors using either NdeI/XhoI restriction sites (for N-terminal TAX), or XhoI/BamHI restriction sites (for C-terminal TAX). Constructs for flow cytometry were subcloned into either MCS1 of pACYC-Duet-1 (Novagen), or pSF-OXB20 (Oxford Genetics), using either NcoI/BamHI or XhoI/BamHI restriction sites, respectively.

2.5.2 Protein Expression and Purification

All vectors for *in vitro* use were transformed into chemically competent *E. coli* BL21 (DE3) Gold (Agilent). Expression was performed by culturing cells in Luria-Bertani (LB) broth,

supplemented with 100 µg/mL ampicillin. Cells were grown at 37°C with shaking to an OD600 of 0.6-0.8, at which point protein expression was induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were then incubated with shaking overnight at 16°C, harvested by centrifugation and lysed using an EmulsiFlex-B15 cell disruptor (Avestin). Proteins were purified by immobilized metal affinity chromatography using Profinity Nickel-charged IMAC resin (Bio-Rad), in a gravity flow column, according to the manufacturer's protocol. Buffer exchange into 20 mM sodium phosphate containing 50 mM NaCl (pH 7.4) and sample concentration was performed using Amicon Ultra-15 centrifugal filters with a 3K molecular weight cut-off (Millipore) for STEPtag, and Microsep Advance centrifugal filters with a 10K molecular weight cut-off (Pall) for all other proteins.

2.5.3 Fluorescence Assays

Proteins were quantified by either the linearized Bradford Assay[220] using the Quick-Start Protein Assay Kit (Bio-Rad), or by measuring absorbance at 280 nm in a cuvette using a SpectraMax Plus[384] microplate spectrophotometer (Molecular Devices). Extinction coefficients for the proteins were estimated using the ProtParam online tool.[221] All measurements were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4), at room temperature. Fluorescence spectra of all proteins were recorded on an Infinite M1000 microplate reader (Tecan). All fluorescence measurements were performed using 75 nM of fluorescent protein, and each sample well was prepared in triplicate. To calculate the K_d, the fluorescence signal when exciting at 485 nm and emitting at 515 nm was used to fit the Hill equation, accounting for ligand depletion,[5] across the range of STEPtag concentrations used. $\Delta F/F_0$ values for all the binding curves were calculated using the Hill equation fit to determine F_{max} and F_{min}, where F_{max} is the maximum signal of the fit, and F_{min} is the minimum signal of the fit.

2.5.4 Flow Cytometry

Constructs used were transformed into chemically competent *E. coli* Tuner™ (DE3) Competent Cells (Novagen), or BL21 (DE3) Gold, in the case of the STEPtag control. Cells were cultured in Terrific Broth (TB) (Fisher) supplemented with 100 µg/mL ampicillin (for cells containing pSF-OXB20 gSTEP0-T1 or pET11a STEPtag) and 34 µg/mL chloramphenicol (for cells containing pACYC STEPtag). Cells were grown with shaking at 37°C for 12 hours, then induced with 1 mM IPTG for 90 minutes. Cells were then harvested by centrifugation, washed twice in filtered 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4), and diluted to an approximate concentration of 10^6 cells/mL with filtered 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). All cells were passed through a 40 µm Falcon Cell Strainer (Fisher) immediately before beginning flow cytometry. Flow cytometry measurements were performed using a Gallios flow cytometer (Beckman Coulter), set to detect 50 000 events per run. Fluorescence was detected by exciting with a 488 nm laser, and detecting emission with a 525BP40 filter. Data analysis was performed using the Kaluza software package (Beckman Coulter).

2.6 Supplementary Information

2.6.1 Amino Acid Sequences

Bim-L0-cpGFP-L0-STEPtag (uni-gSTEP0)

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAY
HYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI
LGHKLEYNTRREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDG
VNWGRIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAA
ESRK*

Bim-L5-cpGFP-L5-STEPtag

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRGGSGGLENVYIKADKQKNGIKANFKIRHNIEDGG
VQLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTG
GSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFK
EDGNILGHKLEYNTRGGSGGREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFE
QVVNELFRDGVNWGRIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTF
VELYGNNAAAESRK*

Bim-L10-cpGFP-L10-STEPtag

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRGGSGGGGSGGLENVYIKADKQKNGIKANFKIRHN
IEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELY
KGGTGGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPT
LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELK
GIDFKEDGNILGHKLEYNTRGGSGGGGSGGREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHI
TPGTAYQSFEQVVNELFRDGVNWGRIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLNDHLEPW
IQENGGWDTFVELYGNNAAAESRK*

Bim-L30-exSTEPtag-L5-cpGFP

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSMSQS
NRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEGTESEMETPSAINGNPSWHLADSPAVNGATGH
SSSLDAREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNWG
RIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAAESRK
GGSGGLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLS
KDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPILVELDGDVNGHKFSVS
GEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERT
IFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNTR*

Bim-L30-exSTEPtag-L10-cpGFP

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRGGGGSGGGGSGGGGSGGGGSGGGGSMSQS
NRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEGTESEMETPSAINGNPSWHLADSPAVNGATGH
SSSLDAREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNWG
RIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAAESRK
GGSGGGGSGGLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSV
QSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPILVELDGDVNGH
KFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGY
IQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNTR*

exSTEPtag-L0-cpGFP-L0-Bim

MSQSNRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEGTESEMETPSAINGNPSWHLADSPAVNG
ATGHSSSLDAREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDG
VNWGRIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAA
ESRKLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSK
DPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSG
EGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTI
FFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNTRDLRPEIRIAQELRRIGDEF
NETYTRRHHHHHH*

exSTEPtag-L5-cpGFP-L5-Bim

MSQSNRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEGTESEMETPSAINGNPSWHLADSPAVNG
ATGHSSSLDAREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDG
VNWGRIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAA
ESRKGGSGGLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSVQ
SKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPILVELDGDVNGHK
FSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYI
QERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNTRGGSGGDLRPEIRIA
QELRRIGDEFNETYTRRHHHHHH*

exSTEPtag-L10-cpGFP-L10-Bim

MSQSNRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEGTESEMETPSAINGNPSWHLADSPAVNG
ATGHSSSLDAREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDG
VNWGRIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAA
ESRKGGSGGGGSGGLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNH
YLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPILVELDGD
VNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAM
PEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNTRGGSGGGGSG
GDLRPEIRIAQELRRIGDEFNETYTRRHHHHHH*

cpGFP-L0-Bim-L30-exSTEPtag

MLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPN
EKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGE
GDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFK
DDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNTRDLRPEIRIAQELRRIGDEFNET
YTRRGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSMSQSNRELVVDFLSYKLSQKGYSWSQFSDVEEN
RTEAPEGTESEMETPSAINGNPSWHLADSPAVNGATGHSSSLDAREVIPMAAVKQALREAGDEFELR
YRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNWGRIVAFFSFGGALCVESVDKEMQVLVSRIA
AWMATYLNDHLEPWIQENGGWDTFVELYGNNAAAESRKHHHHHH*

cpGFP-L5-Bim-L30-exSTEPtag

MLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPN
EKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGE
GDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFK
DDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNTRGGSGGDLRPEIRIAQELRRIGD
EFNETYTRRGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSMSQSNRELVVDFLSYKLSQKGYSWSQFS
DVEENRTEAPEGTESEMETPSAINGNPSWHLADSPAVNGATGHSSSLDAREVIPMAAVKQALREAGD
EFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNWGRIVAFFSFGGALCVESVDKEMQVL
VSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAAESRKHHHHHH*

cpGFP-L10-Bim-L30-exSTEPtag

MLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPN
EKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGE
GDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFK
DDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNTRGGSGGGGSGGDLRPEIRIAQEL
RRIGDEFNETYTRRGGGGSGGGGSGGGGSGGGGSGGGGSGGGGSMSQSNRELVVDFLSYKLSQKGYS
WSQFSDVEENRTEAPEGTESEMETPSAINGNPSWHLADSPAVNGATGHSSSLDAREVIPMAAVKQAL
REAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNWGRIVAFFSFGGALCVESVDK
EMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAAESRKHHHHHH*

cpGFP (used as baseline control for unimolecular scans)

MHHHHHHENLYFQGLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNH
YLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPILVELDGD
VNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSA
MPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYN*

gSTEP0

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAY
HYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI
LGHKLEYNTR*

STEPtag

MHHHHHHREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNW
GRIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAAESR
K*

gSTEP0-T1

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAY
HYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI
LGHKLEY*

gSTEP0-T2

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAY
HYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI
LGHKLE*

gSTEP0-T3

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAY
HYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI
LGHKL*

gSTEP0-T4

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAY
HYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI
LGHK*

TAX-L13-STEPtag

MAEAAQSVDQLIKARGKVYFGVATDQNRLTTGKNAAIIQADFGQVTPENSMKWDATEPSQGNFNFAG
ADYLVNWAQQNGKLIRGHTLVWHSQLPSWVSSITDKNTLTNVMKNHITTLMTRYKGKIRAWDVVNAA
FNEDGSLRQTVFLNVIGEDYIPIAFQTARAADPNAKLYINDYNLDSASYPKTQAIVNRVKQWRAAGV
PIDGIGSQTHLSAGQGAGVLQALPLLASAGTPEVAITALDVAGASPTDYVNVVNACLNVQSCVGITV
WGVADPDSWRASTTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRGGSSGGGGSGGGGSREVIPMAAVK
QALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNWGRIVAFFSFGGALCVES
VDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAAESRKHHHHHH*

TAX-L10-STEPtag

MAEAAQSVDQLIKARGKVYFGVATDQNRLTTGKNAAIIQADFGQVTPENSMKWDATEPSQGNFNFAG
ADYLVNWAQQNGKLIRGHTLVWHSQLPSWVSSITDKNTLTNVMKNHITTLMTRYKGKIRAWDVVNAA
FNEDGSLRQTVFLNVIGEDYIPIAFQTARAADPNAKLYINDYNLDSASYPKTQAIVNRVKQWRAAGV
PIDGIGSQTHLSAGQGAGVLQALPLLASAGTPEVAITALDVAGASPTDYVNVVNACLNVQSCVGITV
WGVADPDSWRASTTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRGGSSGGGGSGGREVIPMAAVKQAL
REAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNWGRIVAFFSFGGALCVESVDK
EMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAAESRKHHHHHH*

STEPtag-L10-TAX

MHHHHHHREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNW
GRIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAAESR
KGSSGGGGSGGMAEAAQSVDQLIKARGKVYFGVATDQNRLTTGKNAAIIQADFGQVTPENSMKWDAT
EPSQGNFNFAGADYLVNWAQQNGKLIRGHTLVWHSQLPSWVSSITDKNTLTNVMKNHITTLMTRYKG
KIRAWDVVNAAFNEDGSLRQTVFLNVIGEDYIPIAFQTARAADPNAKLYINDYNLDSASYPKTQAIV
NRVKQWRAAGVPIDGIGSQTHLSAGQGAGVLQALPLLASAGTPEVAITALDVAGASPTDYVNVVNAC
LNVQSCVGITVWGVADPDSWRASTTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRG*

STEPtag-L8-TAX

MHHHHHHREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNW
GRIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAAESR
KGSSGGSGGMAEAAQSVDQLIKARGKVYFGVATDQNRLTTGKNAAIIQADFGQVTPENSMKWDATEP
SQGNFNFAGADYLVNWAQQNGKLIRGHTLVWHSQLPSWVSSITDKNTLTNVMKNHITTLMTRYKGKI
RAWDVVNAAFNEDGSLRQTVFLNVIGEDYIPIAFQTARAADPNAKLYINDYNLDSASYPKTQAIVNR
VKQWRAAGVPIDGIGSQTHLSAGQGAGVLQALPLLASAGTPEVAITALDVAGASPTDYVNVVNACLN
VQSCVGITVWGVADPDSWRASTTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRG*

STEPtag-L4-TAX

MHHHHHHREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNW
GRIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAAESR
KGSSGMAEAAQSVDQLIKARGKVYFGVATDQNRLTTGKNAAIIQADFGQVTPENSMKWDATEPSQGN
FNFAGADYLVNWAQQNGKLIRGHTLVWHSQLPSWVSSITDKNTLTNVMKNHITTLMTRYKGKIRAWD
VVNAAFNEDGSLRQTVFLNVIGEDYIPIAFQTARAADPNAKLYINDYNLDSASYPKTQAIVNRVKQW
RAAGVPIDGIGSQTHLSAGQGAGVLQALPLLASAGTPEVAITALDVAGASPTDYVNVVNACLNVQSC
VGITVWGVADPDSWRASTTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRG*

2.6.2 Sample SDS-PAGE Gels

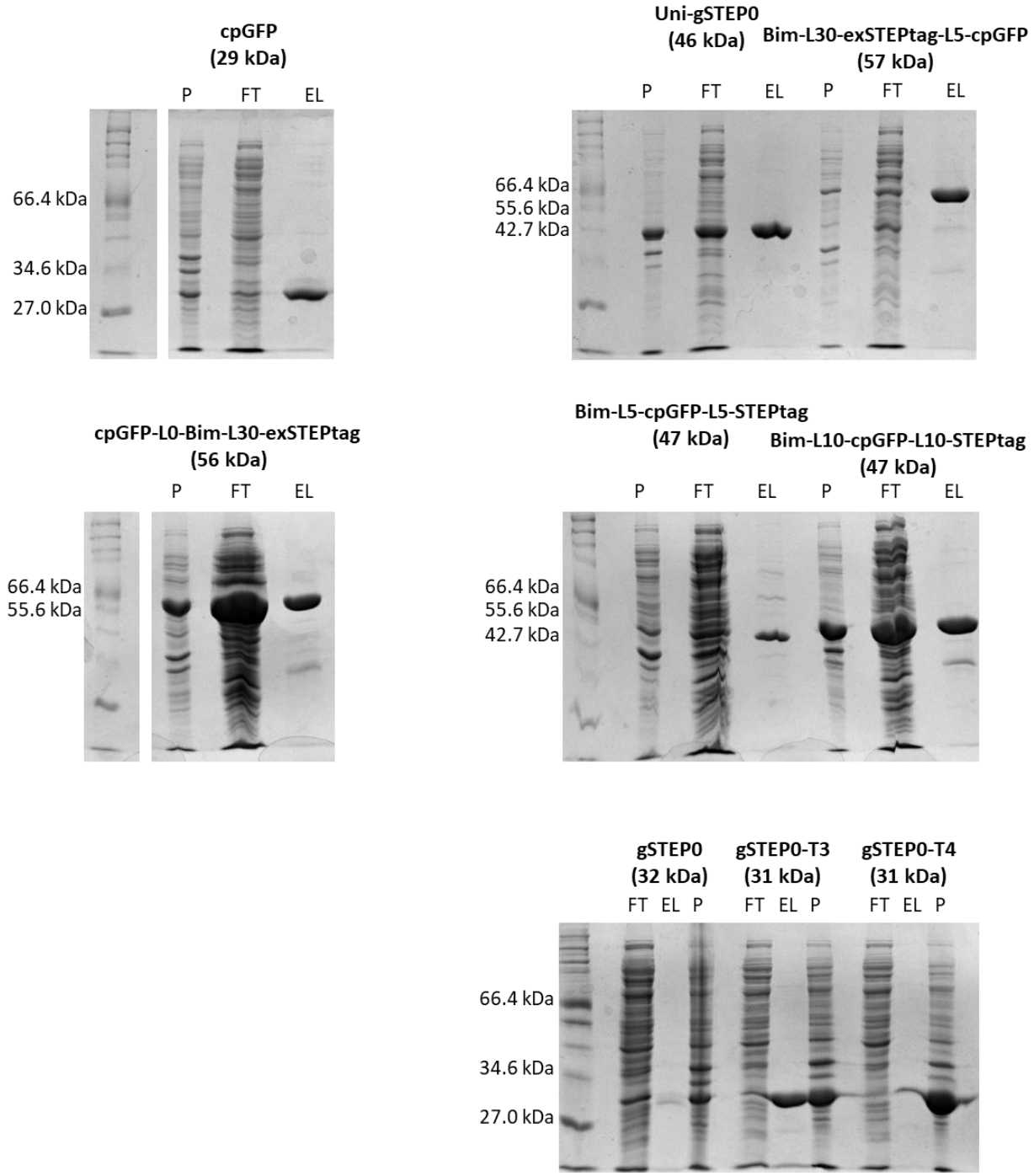


Figure S2.1 Sample SDS-PAGE gels of purification steps. Samples taken during protein purification were mixed 1:1 by volume with SDS-PAGE loading dye, heated to 95°C for 5 minutes, then loaded into 10% acrylamide SDS-PAGE gels. The ladder used is the P7702 Unstained Protein Marker, Broad Range (NEB). Electrophoresis was performed at 120 volts until the dye front had migrated to the end of the gel. Abbreviations are as follows: P – pellet, FT – flowthrough, EL – eluted (purified)

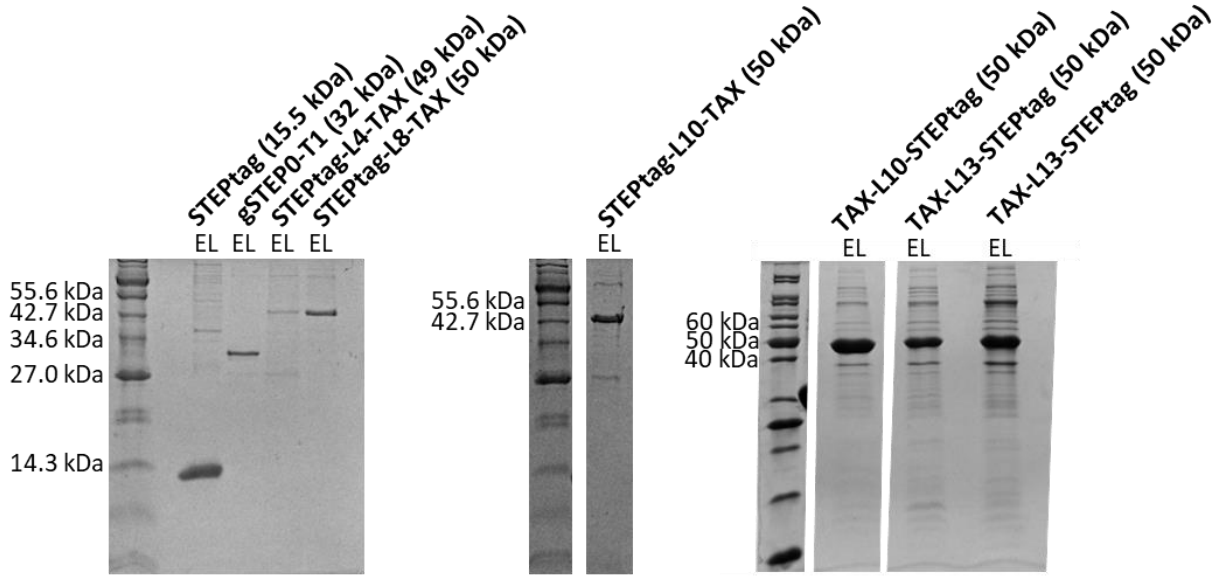


Figure S2.2 Sample SDS-PAGE gels of purified proteins. Samples taken after protein purification were mixed 1:1 by volume with SDS-PAGE loading dye, heated to 95°C for 5 minutes, then loaded into 15% acrylamide SDS-PAGE gels, to better resolve STEPtag. The ladder used is the P7702 Unstained Protein Marker, Broad Range (NEB) for the left and center gels, or P7717 Unstained Protein Standard, Broad Range for the rightmost gel (NEB). Electrophoresis was performed at 120 volts until the dye front had migrated to the end of the gel. Abbreviations are as follows: EL – eluted (purified)

2.6.3 Mass Spectrometry Analysis of STEPtag

As STEPtag migrates lower than the 14.3 kDa band of the SDS-PAGE ladder (Figure S2.2), we wanted to verify that the protein was not undergoing any form of post-translational degradation. Mass spectrometry of an FPLC purified sample of STEPtag was performed by the University of Ottawa's Proteomics Resource Center. The calculated molecular weight was 15 503.81 Da (see mass spectrum below), which is very close to the expected molecular weight of 15 504.4 Da, therefore we conclude that no degradation is occurring.

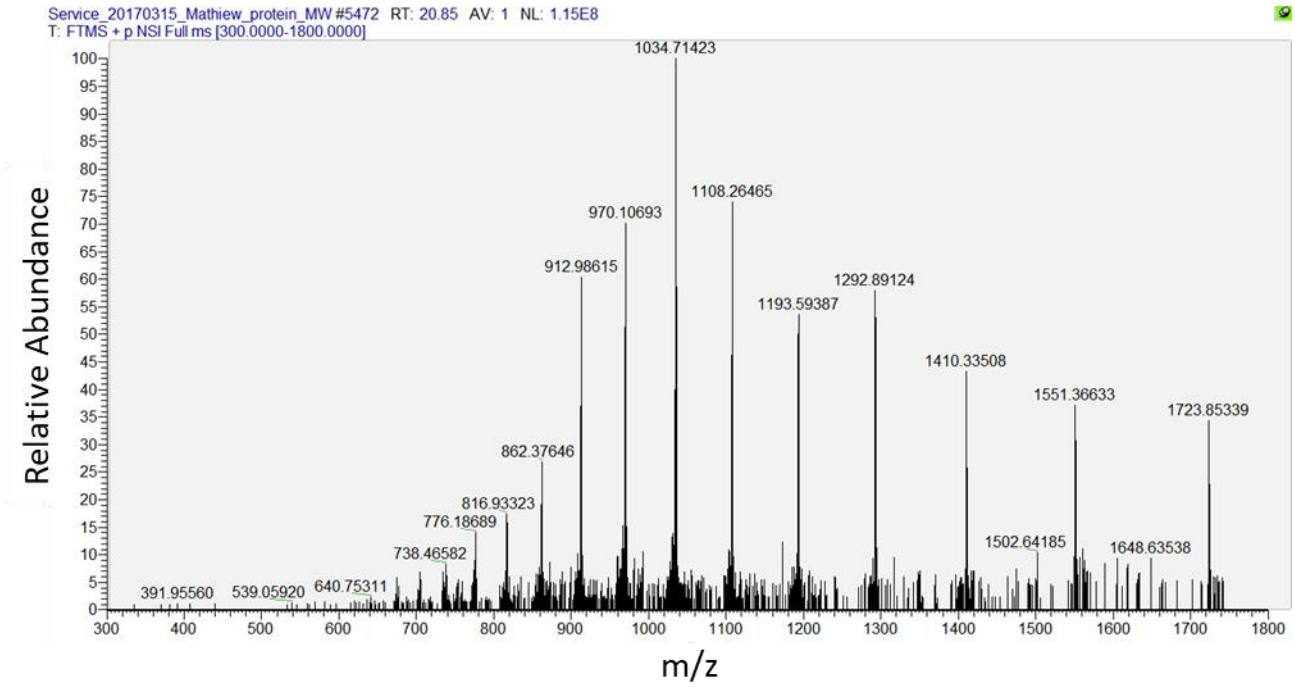


Figure S2.3 Mass spectrum of STEPtag. The instrument used was an LTQ-velos-Orbitrap Elite (Thermo Scientific), and the spectrum was acquired using an Orbitrap analyzer, with a resolution of 60K (at 200 m/z).

2.6.4 Raw Fluorescence Data

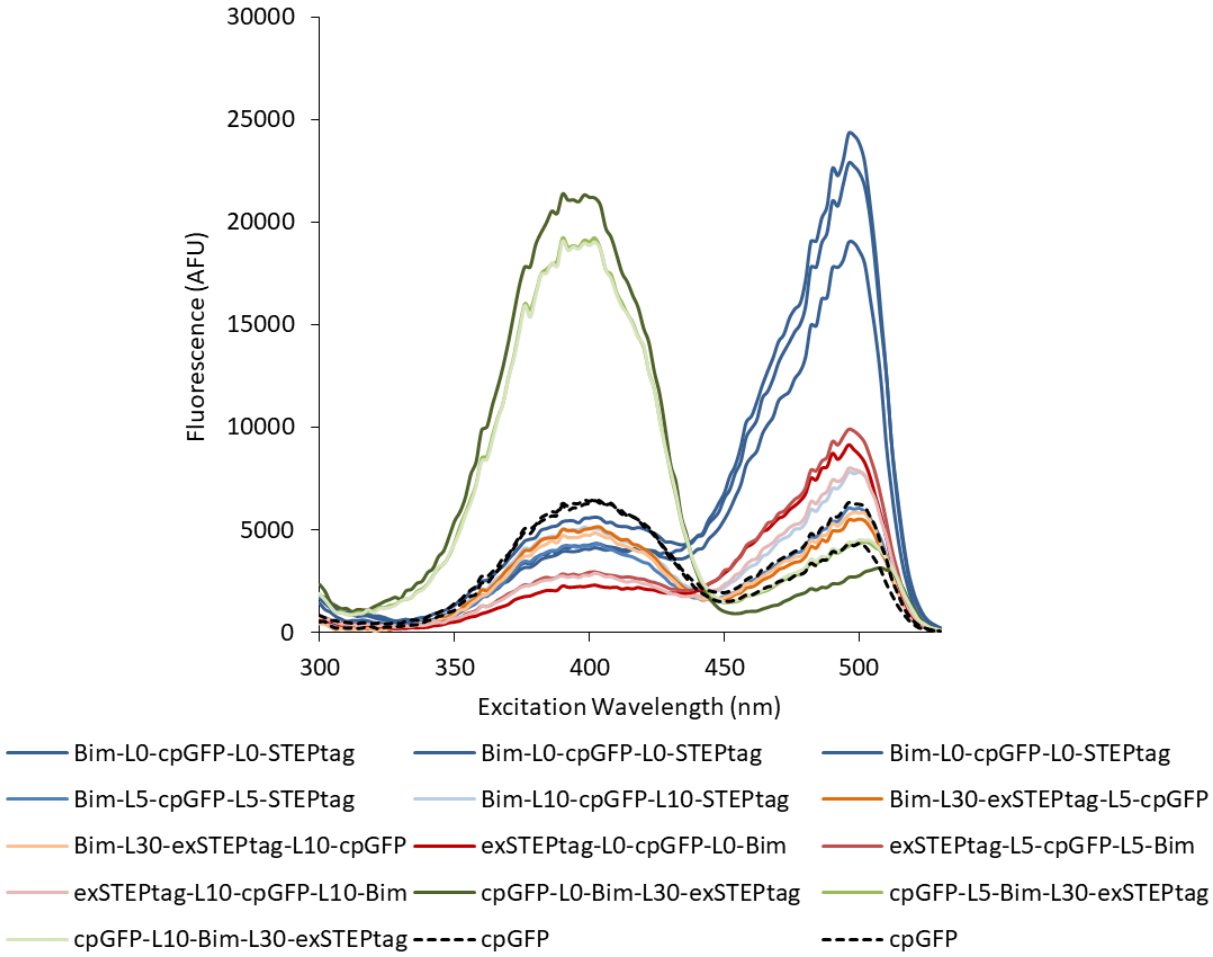


Figure S2.4 Excitation scans of the unimolecular constructs. Scans of 75 nM of each protein were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). The emission wavelength used was 550 nm. Multiple biological replicates of Bim-L0-cpGFP-L0-STEPtag (uni-gSTEP0) and cpGFP are included.

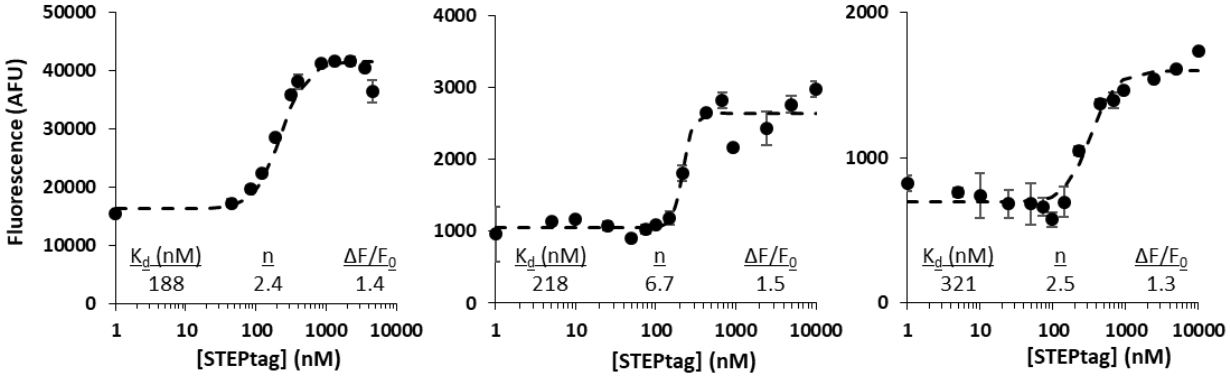


Figure S2.5 Binding curves of gSTEP0 with STEPtag. Scans of 75 nM of biological replicates of purified gSTEP0, mixed with varying concentrations of biological replicates of purified STEPtag, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). The fluorescence signal of the left panel is the emission between 520 and 540 nm, when exciting at 488 nm, as an approximation of the signal through a green filter set (530/20 nm) when exciting with a 488nm laser, while the center and right panels are the peak emission signal when exciting at 485 nm. The dashed lines are the results of a fit of the Hill equation to the data, and K_d, $\Delta F/F_0$ and the value of n calculated from the Hill equation fit are reported for each curve. Error bars are the standard deviation of triplicate measurements.

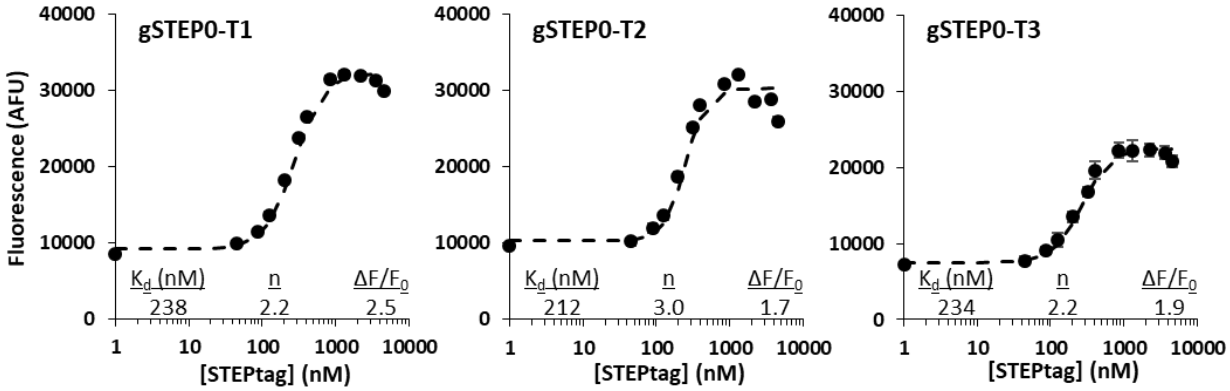


Figure S2.6 Binding curves of gSTEP0 truncations. Scans of 75 nM of purified gSTEP0 truncations, mixed with varying concentrations of STEPtag, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). These experiments were performed with the same purification of STEPtag. The fluorescence signal is the emission between 520 and 540 nm, when exciting at 488 nm. The dashed lines are the results of a fit of the Hill equation to the data, and K_d, $\Delta F/F_0$ and the value of n calculated from the Hill equation fit are reported for each curve. Error bars are the standard deviation of triplicate measurements. These curves were performed under the same conditions as the left panel of Figure S2.5, using the same STEPtag.

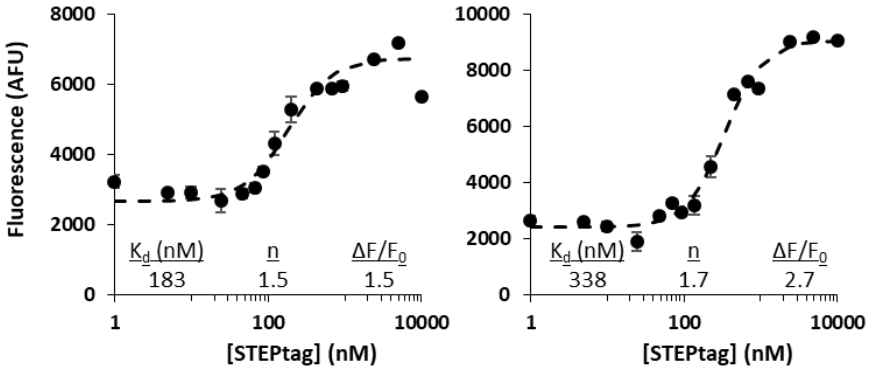


Figure S2.7 Replicate binding curves of gSTEP0-T1. Scans of 75 nM of biological replicates of purified gSTEP0-T1, mixed with varying concentrations of purified STEPtag, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). The fluorescence signal is the emission at 515 nm, when exciting at 485nm. The dashed lines are the results of a fit of the Hill equation to the data, and K_d, $\Delta F/F_0$ and the value of n calculated from the Hill equation fit are reported for each curve. Error bars are the standard deviation of triplicate measurements.

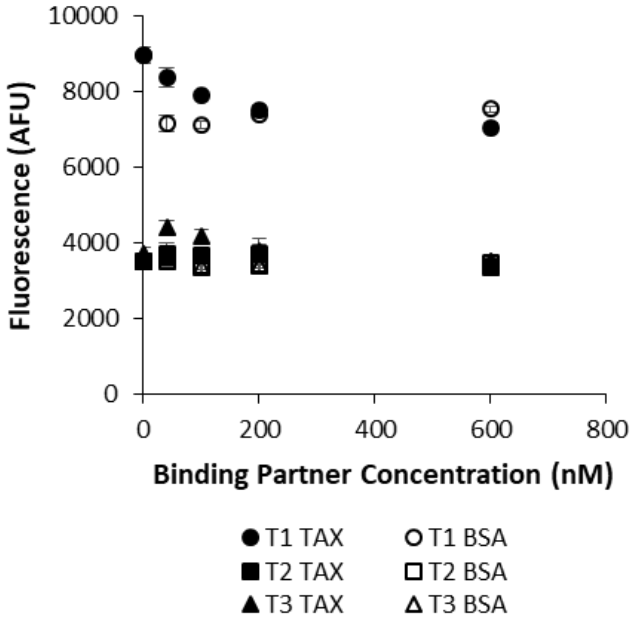


Figure S2.8 The fluorescence of gSTEP0 truncations are unaffected by TAX and BSA. 75 nM of purified gSTEP0 truncations, mixed with varying concentrations of purified TAX or BSA, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). The fluorescence signal is the emission between 520 and 540 nm, when exciting at 488 nm. Error bars are the standard deviation of triplicate measurements.

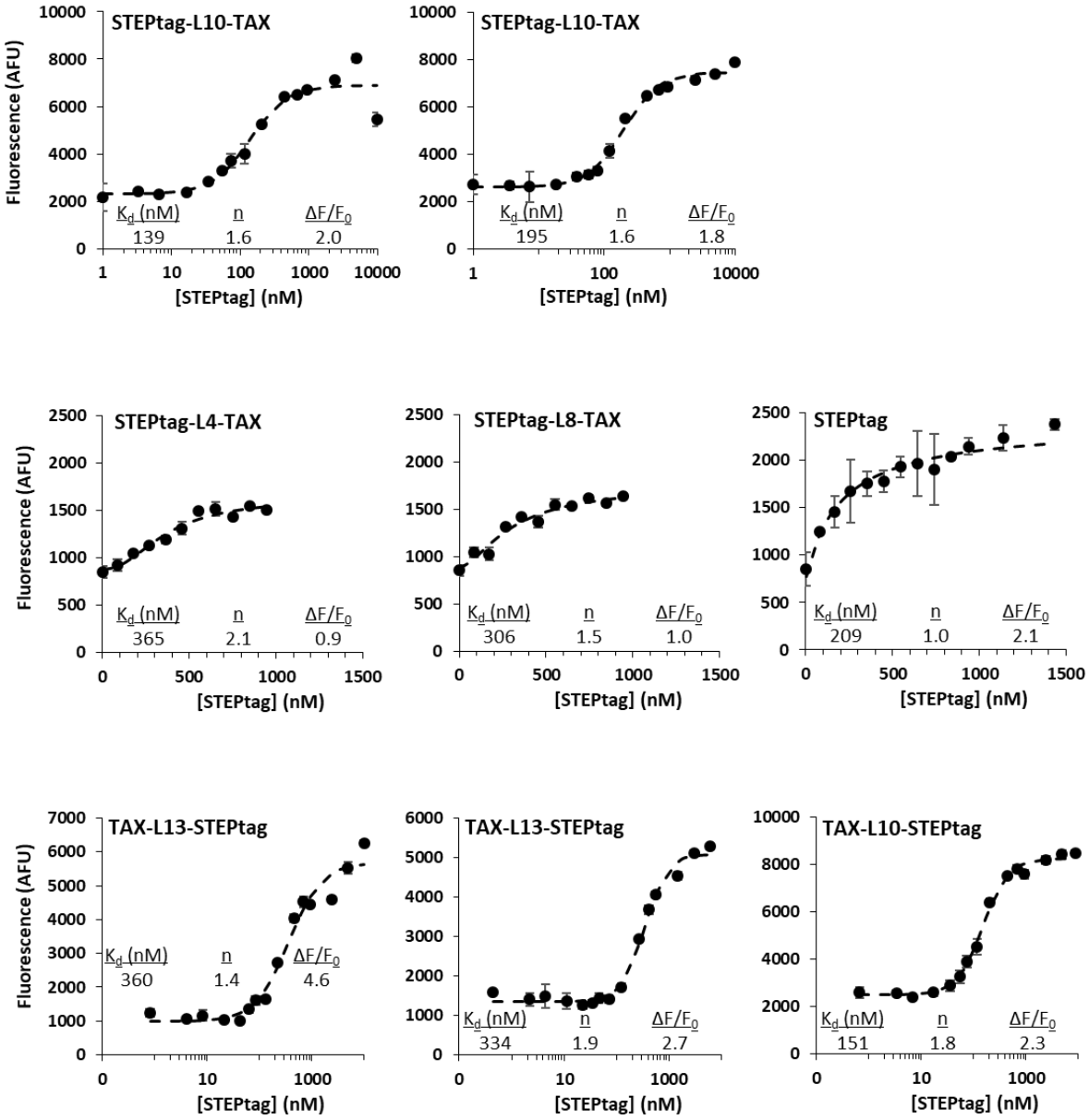
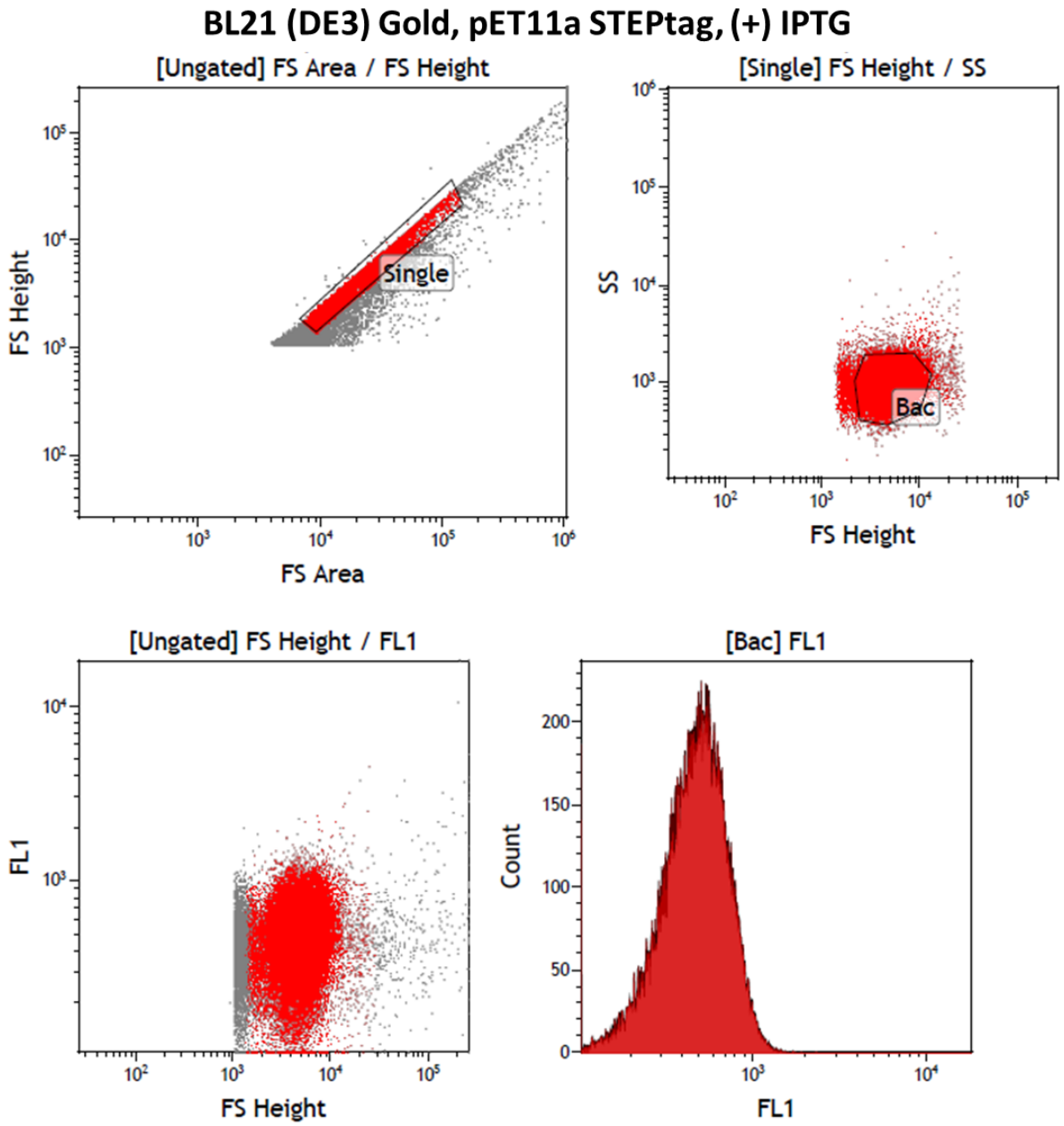
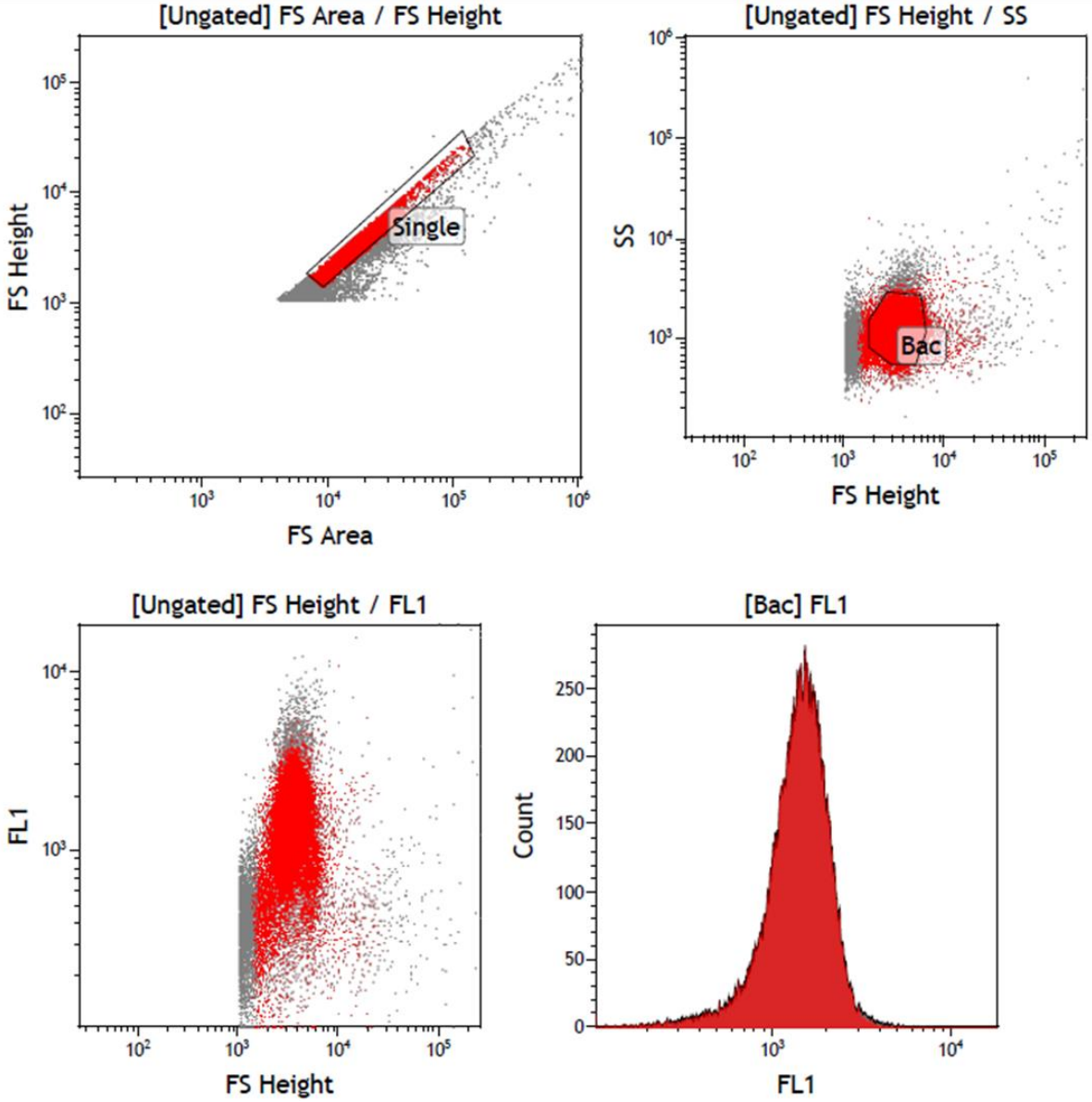


Figure S2.9 Binding curves of gSTEP0-T1 with TAX-STEPtag fusions. Scans of 75 nM of purified gSTEP0-T1, mixed with varying concentrations of TAX-STEPtag fusions, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). The fluorescence signal is the emission at 515 nm, when exciting at 485nm. The dashed lines are the results of a fit of the Hill equation to the data, and K_d, $\Delta F/F_0$ and the value of n calculated from the Hill equation fit are reported for each curve. Error bars are the standard deviation of triplicate measurements. Note that the STEPtag-L4 and L8-TAX curves are on a linear concentration scale. A representative STEPtag curve has been included beside them to compare.

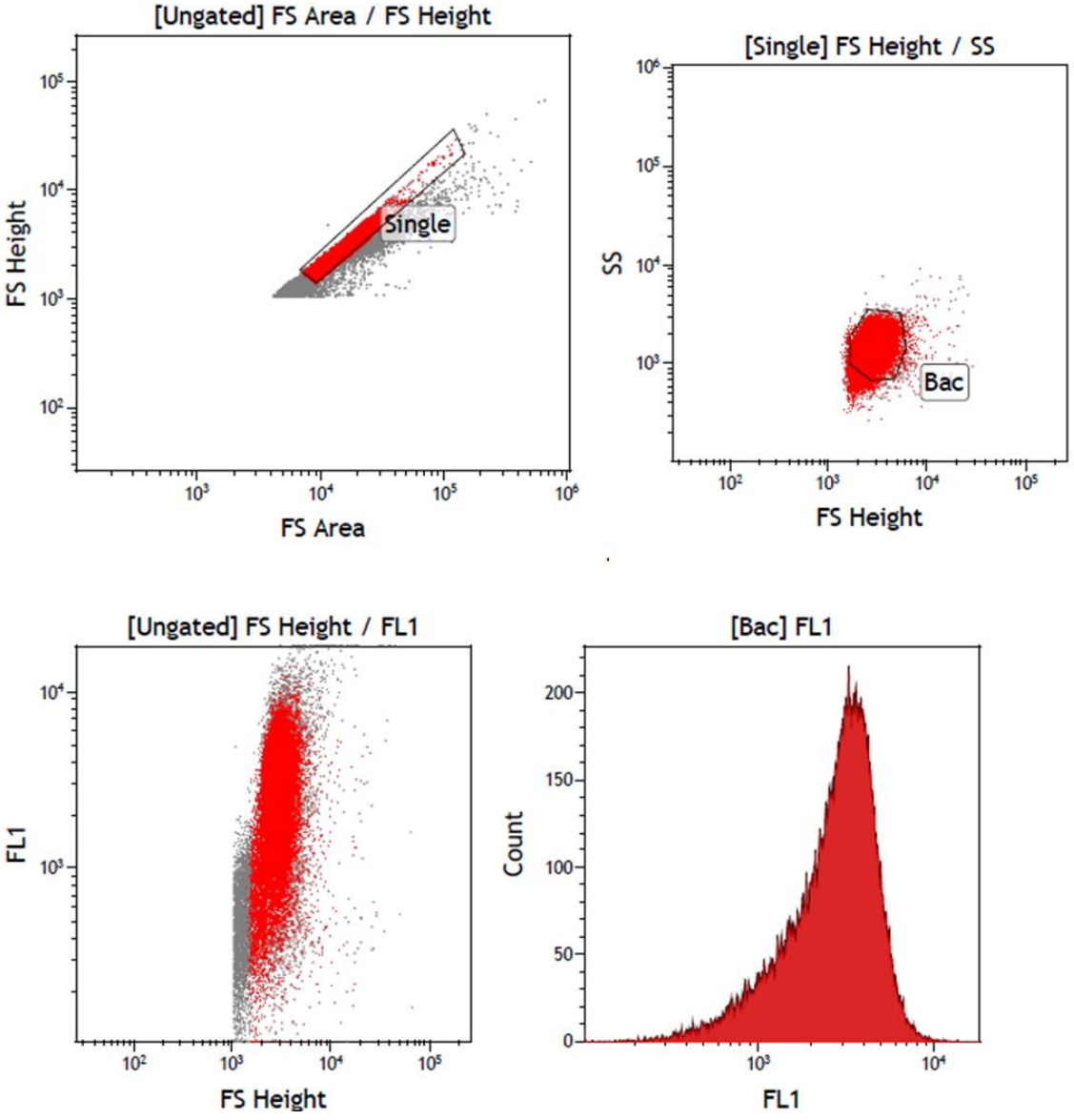
2.6.5 Additional Flow Cytometry Data



Tuner (DE3), pSF-OXB20 gSTEP0-T1, (+) IPTG



Tuner (DE3), pSF-OXB20 gSTEP0-T1,
pACYC-Duet STEPtag, (+) IPTG



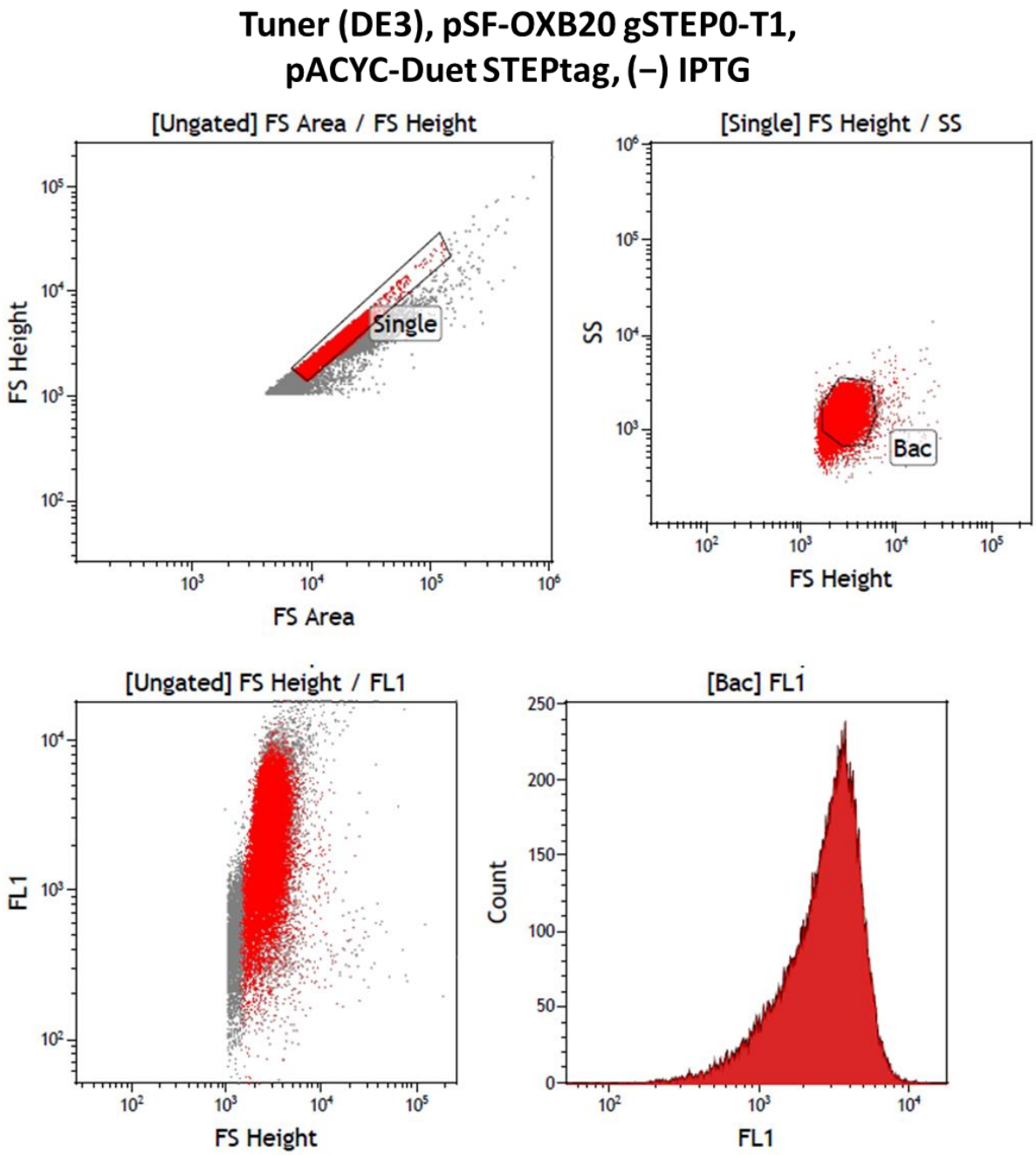
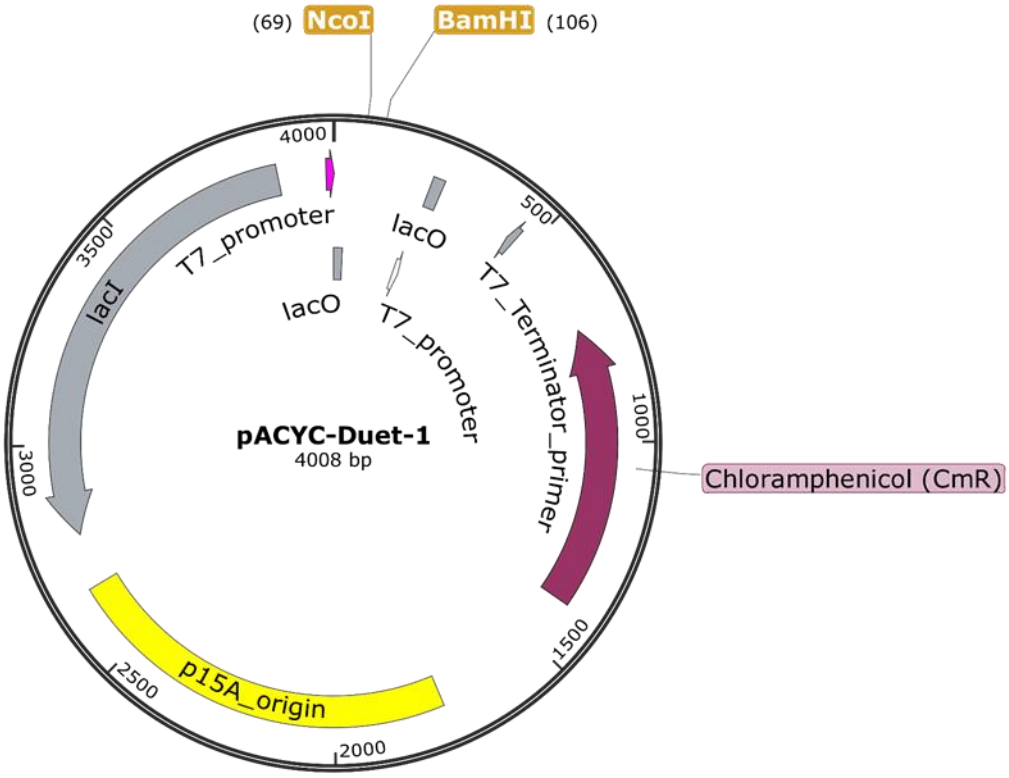
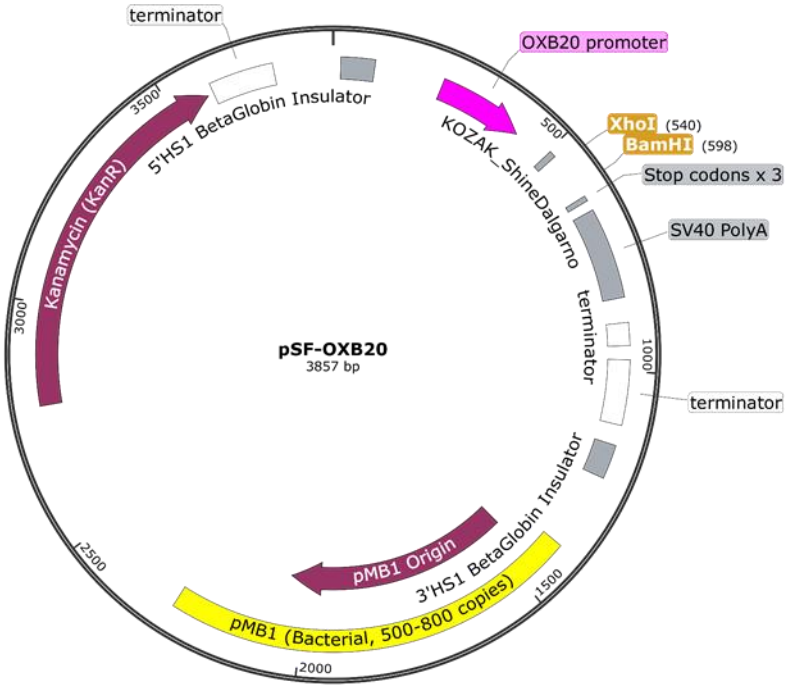


Figure S2.10 Flow cytometry data. Cell strain, plasmids and presence of IPTG are all indicated in the figure titles. Dot plots of forward scatter (FS) area versus height for doublet discrimination, forward scatter versus side scatter height (SS), for selection of well-formed cells, and forward scatter versus green fluorescence height (FL1) for the flow cytometry runs shown in Figure 2.4. Events with a large forward scatter area to forward scatter height ratio are more likely to be doublets, i.e. multiple cells passing through the detector simultaneously, so they were removed with the "Single" gate (red), which also removes the population of small, nonfluorescent particles, which we hypothesize to be dust. The "Bac" gate from the forward scatter versus side scatter plot was used for the fluorescence histogram, to screen out outliers with anomalous shapes or sizes.

2.6.6 Vector Maps





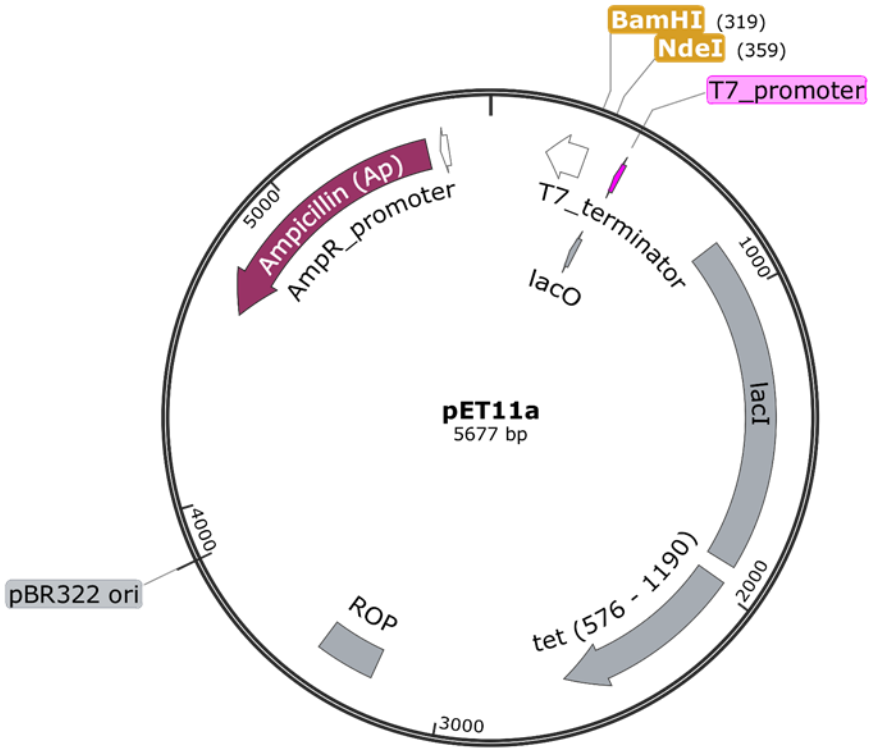


Figure S2.11 Vector maps of plasmids used in this chapter. Vector maps were generated using SnapGene Viewer. Promoters used are highlighted in magenta, and restriction sites used for cloning are highlighted in brown. Note that pSF-OXB20 and pACYC-Duet-1 have different antibiotic resistances (Kanamycin and Chloramphenicol), as well as compatible origins of replication (pMB1 and p15A), such that they can both be stably maintained within the same cells.

Chapter 3: STEPping Forwards: Developing the STEP into a Sensor for Real-Time Imaging

Matthew G. Eason, Antonia T. Pandelieva, Marc M. Meyer & Roberto A. Chica

Manuscript in preparation

3.1 Statement of Contribution

Matthew G. Eason designed all protein sequences, all cloning and expression strategies and subcloned sequences. Matthew G. Eason and Antonia T. Pandelieva performed *in vitro* characterization experiments and analyzed data. Matthew G. Eason and Roberto A. Chica designed the two-stage FACS sort. Antonia T. Pandelieva performed FACS experiments. Matthew G. Eason performed stopped-flow and flow cytometry experiments. Marc M. Meyer performed the plate-based kinetic assay. Matthew G. Eason wrote the chapter.

3.2 Introduction

In the previous chapter, we created our Sensor for Transiently Expressed Proteins (STEP), and demonstrated that it showed a measurable increase in fluorescence when gSTEP0 is in the presence of a protein of interest bound to STEPtag. Yet these *in vitro* demonstrations were only the first steps towards the desired use of the sensor: detecting transiently expressed proteins in model organisms. With that in mind, we moved forward with three major goals: improvement of the signal-to-noise ratio and K_d, *in cellulo* testing and kinetic characterization.

Firstly, our best version of the sensor, gSTEP0-T1, was found to have a K_d of 210 nM, much greater than the theoretical K_d of the parent binding partners (0.8 nM), as well as a $\Delta F/F_0$ of 2.0, lower than that of the GCaMP sensors on which it is based (ranging from 3.5 to over 10). We hoped to improve both of these characteristics, first by rational design, then by developing a high-throughput screening method to perform directed evolution. Our rational design hypotheses focused on influencing the affinity of the binding partners, thus lowering the K_d, and were twofold: first, that the specific sequence of the binding peptide could influence the binding interaction, and second, that changing the

proximity of the binding peptide to the cpGFP domain could relieve potential clashes between the STEPtag and the cpGFP barrel that would destabilize the bound state, and thus negatively affect binding. To verify the effect of changing the binding peptide, we tested alternative sequences of the Bim peptide, testing a homologous sequence as well as a selection of high-affinity designed mutants. To change the proximity of the binding peptide and the cpGFP domain in gSTEP0-T1, flexible glycine-serine linkers were added between the two sequences.

Secondly, as the original goal of the STEP was to use it in an *in vivo* context to track transiently expressed proteins inside of research model animals, we moved towards that goal by demonstrating that the STEP signal can be detected inside of live *E. coli* cells. Our previous attempt at flow cytometry using pSF-OXB and pACYC vectors (Chapter 2) was difficult to work with, required multiple vectors, and seemed to have high levels of background STEPtag, so we tested different combinations of vectors and promoters in order to determine an efficient method to produce both halves of the STEP inside the same cell, either simultaneously or sequentially. Some of these constructs were then used to perform high-throughput screening of STEP variants, using fluorescence-activated cell sorting (FACS), while others were used for *in cellulo* kinetic characterization.

For this third goal of kinetic characterization, we were interested in two parameters: the timescale at which the gSTEP-STEPtag binding interaction occurs, and the time required to detect STEPtag after its expression in a cell. For the STEP to be a useful alternative to standard fluorescent protein tags, it needs to operate on a faster timescale than the tens of minutes required for GFP maturation, and ideally even faster than the few minutes required for Venus, the fastest-maturing FP. To directly measure the binding kinetics of the STEP, we used a stopped-flow spectrophotometer, providing rapid mixing and fluorescence measurements on the sub-second timescale. To measure STEP signal post-STEPtag induction, we used our *in cellulo* expression system, and compared the time required to detect STEP signal to the time required to detect EGFP signal, after the induction of both.

3.3 Results

3.3.1 Rational Improvement of gSTEP0-T1

Our first rational improvement to gSTEP0-T1 involved changing the binding peptide. The Bim peptide present in gSTEP0-T1 uses the mouse amino acid sequence, which has been show to bind with sub-nanomolar affinity to human Bcl-$x_{L,}$[119] the template for the design of STEPtag. We hypothesized that using the human Bim peptide could improve the affinity of the sensor, as that peptide-protein binding interaction has been co-evolved, and Dutta *et al.*[120] have shown the K_i for the two proteins to be below the limit of detection of their assay (< 0.1 nM). In that same work, Dutta *et al.* also designed two Bim mutants, named XXA1 and XXA4, which demonstrate improved specificity for Bcl-x_L and maintain high affinity. Although XXA1 and XXA4 were measured to have K_i values of 0.09 nM and 0.2 nM, respectively, two mutants of XXA1 were found to have K_i values that were below the limit of detection, and thus could potentially have higher affinity than the wild-type sequence. We chose to test these two mutant peptides, named G2gE and Y4eK, as well as the original XXA1 and XXA4 sequences (Table 3.1). Interestingly, the human Bim peptide improved both the affinity and the signal of the STEP, lowering the K_d from 210 nM to 160 nM and increasing the $\Delta F/F_0$ from 2.0 to 3.3. As for the designed peptides, the XXA1 and XXA4 peptides performed roughly as expected, with no real change in K_d when using XXA1, and an increased K_d with XXA4, matching the trends seen in the original paper. The G2gE peptide was the best performing of the mutants, with a K_d of 180 nM and a $\Delta F/F_0$ of 2.3, but the human Bim construct still outperformed it in both metrics tested.

Table 3.1 Characterizing gSTEP0-T1 constructs with different binding peptides.

Sequence	K_d (nM)	$\Delta F/F_0$
gSTEP0-T1-hBim[a]	160 ± 40	3.3 ± 0.1
gSTEP0-T1-XXA1	190	2.3
gSTEP0-T1-XXA4[a]	280 ± 40	1.3 ± 0.1
gSTEP0-T1-G2gE[b]	180 ± 30	2.3 ± 0.2
gSTEP0-T1-Y4eK[a]	240 ± 40	1.4 ± 0.3

[a] Data obtained from three biological replicates, error shown is the standard deviation
[b] Data obtained from two biological replicates, error shown is the standard deviation

While we were varying the Bim peptide sequence, we also investigated our second hypothesis, that potential steric clashes between gSTEP0-T1 and STEPtag could be relieved by increasing the distance between the peptide and the cpGFP domain, using flexible linker sequences. As the original screen of the unimolecular STEP had shown poor results with linker sequences of 5 amino acids and above, we restricted ourselves to linkers of 1-5 amino acids, balancing the competing interests of creating space for the binding interaction to occur and maintaining proximity to the chromophore for signal modulation. These mutants were named gSTEP0-T1-Lx, where x is the number of amino acids in the linker. We found that the sequences with 1-3 amino acids added did not outperform the parent protein (Table 3.2), and in fact the single amino acid linker seems to prevent the STEP from being affected by STEPtag. The 4 and 5 amino acid linkers, on the other hand, do demonstrate improved affinity, with K_d values of 130 and 120 nM, respectively. We chose to continue working with the 4 amino acid linker, as gSTEP0-T1-L4 had more reproducible results across multiple replicates, as well as generally higher overall brightness, compared to gSTEP0-T1-L5 (Figure S3.3).

Table 3.2 Characterizing gSTEP0-T1 mutants with increased linker length between the Bim peptide and cpGFP.

Sequence	K_d (nM)	$\Delta F/F_0$
gSTEP0-T1-L1[a]	N/A	N/A
gSTEP0-T1-L2[b]	210 ± 30	1.0 ± 0.2
gSTEP0-T1-L3[b]	180 ± 60	1.5 ± 0.7
gSTEP0-T1-L4[c]	130 ± 10	2 ± 1
gSTEP0-T1-L5[c]	120 ± 70	3 ± 2

[a] No change in the fluorescence of gSTEP0-T1-L1 was observed upon addition of STEPtag
[b] Data obtained from three biological replicates, error shown is the standard deviation
[c] Data obtained from five biological replicates, error shown is the standard deviation

Having identified improved mutants from both rational design hypotheses, we incorporated all beneficial mutations into the same scaffold, creating a gSTEP with both an extended linker between the peptide and the barrel, as well as the human Bim peptide, rather than the mouse peptide. We named this construct gSTEP0-T1-hBim-L4. Unfortunately, it did not show additive improvements from the two parents, with a K_d of 160 ± 5 nM and a $\Delta F/F_0$ of 0.8 ± 0.1, when exciting the anionic form of the chromophore at 485 nm. On the other hand, we noticed that this mutant displayed a considerably larger change in fluorescence when the neutral form of the chromophore was excited, at 410 nm (Figure 3.1), displaying a $\Delta F/F_0$ of 2.9 ± 0.1, as well as a slightly improved K_d of 117 ± 3 nM. As we do not expect the excitation wavelength to change the binding kinetics of the sensor, it seems that this improvement in apparent K_d comes from an improved fit of the Hill equation to the data (Figure S3.4). Regardless, we were intrigued by this result, as exciting gSTEP0-T1-hBim-L4 at 410 nm did show the desired combination of low K_d (as seen in the L4 mutant) and high $\Delta F/F_0$ (as seen in the human Bim peptide mutant). Replicating this combination for the anionic chromophore would result in the best version of the gSTEP yet.

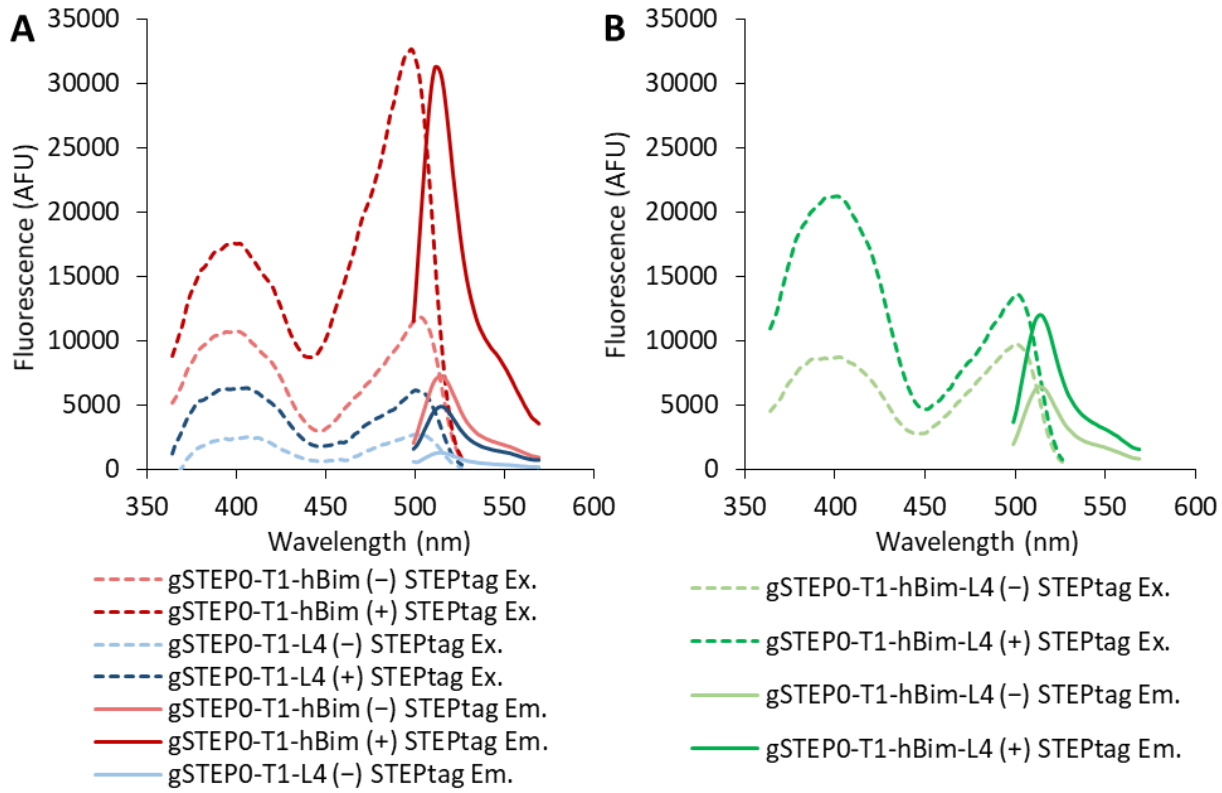


Figure 3.1 Fluorescence spectra of improved gSTEP0-T1 mutants. A, The excitation (Ex.) and emission (Em.) spectra of the most improved mutants obtained from each of the linker extension and peptide swapping experiments, each in the presence (+) or absence (−) of STEPtag. **B,** The excitation (Ex.) and emission (Em.) spectra for the combined mutant gSTEP0-T1-hBim, containing both the human Bim peptide and the four amino acid linker. Each spectrum was gathered using 75 nM of gSTEP0-T1 mutant, either alone or in the presence of 5000 nM STEPtag.

3.3.2 High-Throughput Screening of STEP Mutants Using FACS

While performing the rational improvements of the STEP, we realized that a high-throughput screening assay would be helpful for further engineering efforts, as it would allow us to supplement the small, directed design hypotheses with the screening of large combinatorial libraries, as well as libraries of random mutants for the purposes of directed evolution. We selected FACS as our high-throughput screening method of choice, as it can feasibly screen libraries of millions of mutants, and directly reads the fluorescent signal of the STEP. In order to properly screen the capacity of the STEP to detect a protein of interest, we needed to develop a bacterial expression system that would allow us to measure the fluorescence of both the bound and unbound states of STEP mutants. As an ideal STEP

has not just a bright fluorescent signal in the on state, but also a dim signal in the off state, we started by designing a two-step screening assay that could interrogate both of these states sequentially (Figure 3.2). In this assay, both halves of the sensor (gSTEP and STEPtag) are cloned into the same pET-Duet expression vector. This vector contains two multiple cloning sites, each of which is preceded by a T7 promoter, such that the two proteins are co-expressed when IPTG is added. The first step of the screening assay is a positive sort of these cells, where both halves of the sensor are expressed and the cells with the highest brightness are kept. These hits then have the STEPtag gene removed by a single restriction digestion, the vector containing just the fluorescent gSTEP mutant is recircularized by ligation and the DNA library is retransformed and rescreened. In the second screening step, a negative sort is performed, selecting cells with low fluorescence in the absence of STEPtag, in order to identify mutants with low background fluorescence and thus a high $\Delta F/F_0$.

To test this two-step screening method, we performed combinatorial saturation mutagenesis on the introduced GSGS linker in gSTEP0-T1-hBim-L4, as it seemed likely that these added amino acids in close proximity to the pore could be contributing to a change in the chromophore pK_a, leading to the large population of the neutral chromophore and to change upon STEPtag binding being seen in the neutral chromophore peak rather than the anionic one. Additionally, the sequences of the linkers in GCaMPs have previously been shown to contribute to the modulation of fluorescence of those sensors,[82, 89] so it seemed possible that an improved sequence could be found. This combinatorial saturation library consisted of all combinations of the 20 natural amino acids at the 4 positions in the added GSGS linker of gSTEP0-T1-hBim-L4, for a theoretical library size of 20^4 or 160 000 protein sequences. Such a library would be challenging to screen with lower-throughput methods, such as 96-well plate-based screening, but was of a reasonable size for FACS sorting.

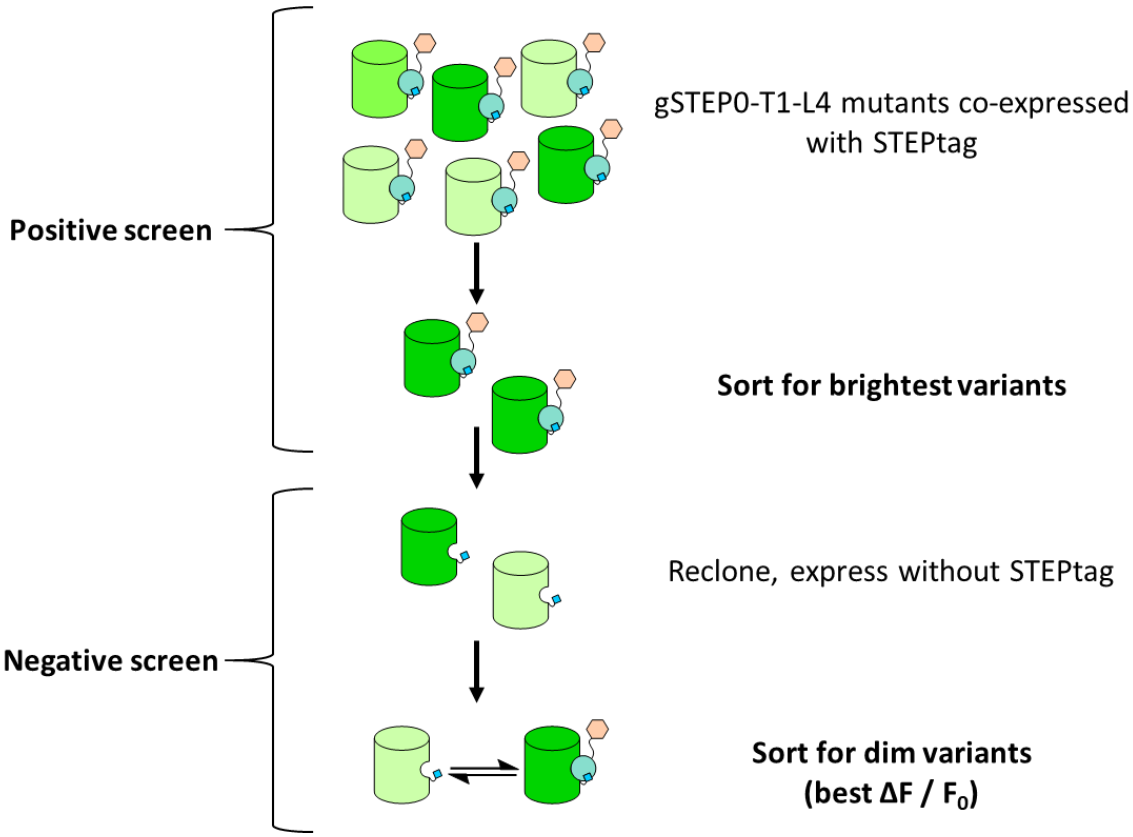


Figure 3.2 Design of a two-step FACS screen for STEP mutants with improved $\Delta F/F_0$. To identify sensors with improved signal-to-noise ratios, two consecutive FACS sorts are performed. In the first (Positive screen), the entire mutant library is co-expressed with STEPtag, and cells containing the brightest mutants are retained. These mutants are then recloned to remove STEPtag from the expression vector, and transformed into new cells. These cells, containing gSTEP mutants known to be bright in the presence of STEPtag, are then sorted again (Negative screen), this time keeping cells with dim signal. Cells that pass through both sorts should contain sensor mutants that are bright in the presence of STEPtag, and dim in its absence, resulting in an increased $\Delta F/F_0$.

In the positive sort, we tested cells for fluorescence when exciting both at 405 and 488 nm, as the parent protein, gSTEP0-T1-hBim-L4, was shown to have an improved $\Delta F/F_0$ when exciting the neutral chromophore at around 405 nm (Figure 3.1), and we desired a sensor with improved signal when exciting the anionic chromophore. To select mutants where this shift had occurred, we selected only cells with high fluorescence when excited at 488 nm, and low fluorescence when excited at 405 nm (Figure 3.3). These hits then had the STEPtag removed, and the vectors were retransformed to perform the negative sort. In this case, we selected three different populations in the second screen to determine if our negative screen was performing as intended: low, moderate, and high fluorescence

cells were each selected from the sort in the absence of STEPtag. The low population contained cells with fluorescence levels equal to those of cells expressing no fluorescent protein at all. In theory, this population would contain ideal sensors, where the off state could not be distinguished from background, while the on state when STEPtag is added remains bright, as seen in the positive sort. But this population would also contain cells that did not successfully express the sensor, and so libraries selected from the population would be expected to have a high level of false positives. The high population contained cells with fluorescence levels equal to those of the brightest proteins in the positive sort, and so would be expected to contain proteins with mutations that generally increase the fluorescence of the sensor, without necessarily improving the $\Delta F/F_0$. Finally, the medium population contained cells with fluorescence levels between the other two, and so should still have measurable baseline fluorescence, but be brighter in the on state (Figure S3.15).

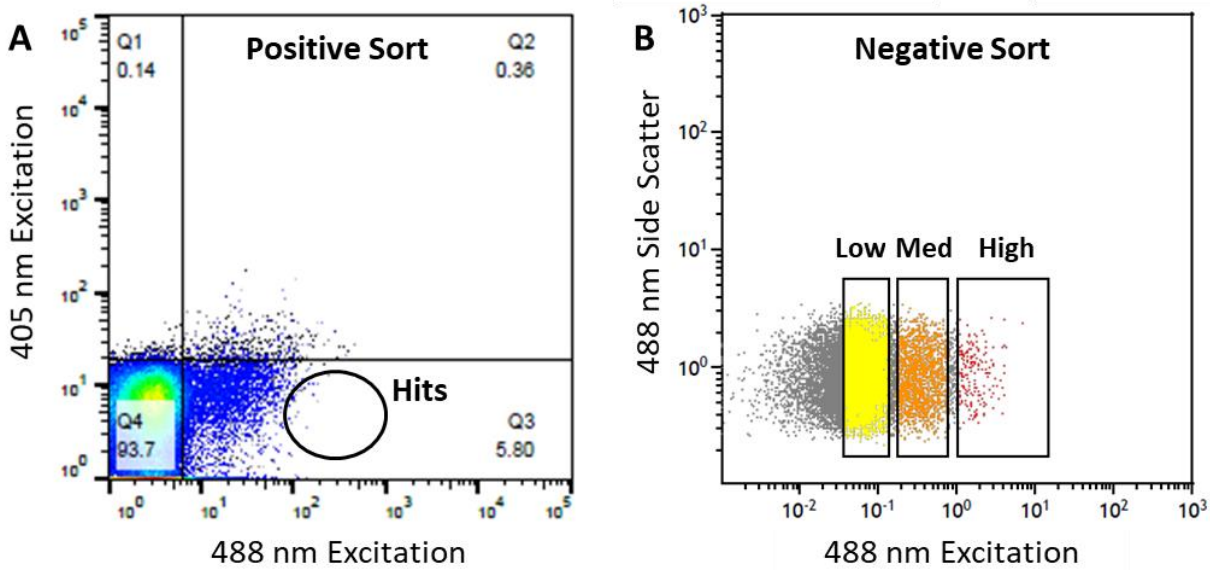


Figure 3.3 Two-step FACS screen of a library of gSTEP0-T1-hBim-L4 mutants. A, The first step of the screen, with cells containing both gSTEP0-T1-hBim-L4 mutants and STEPtag. The cells selected from this screen (black oval, Hits), were those that showed the brightest green fluorescence when excited at 488 nm and also showed relatively low fluorescence when excited at 405 nm, in order to enrich the resulting population in mutants where the anionic chromophore is favoured. **B,** The second step of the screen, with cells containing the hits from the first step in the absence of STEPtag. Three different populations were sorted from this library, based on their brightness relative to the parent protein. The low population contains cells that are dimmer than control cells expressing gSTEP0-T1-hBim-L4 (the off state of the sensor). The medium (Med) population contains cells that are as bright as the off state, and the high population contains cells that are as bright as the on state of the sensor, when STEPtag is present.

For each of the three populations in the negative sort, cells containing the mutant sensor were picked into a 96-well plate, cultured, induced and lysed. For each mutant, an equal volume of cell lysate was then mixed with either phosphate buffer or saturating concentrations of STEPtag, and the fluorescence spectra were measured (Figure 3.4). Unfortunately, none of the mutants from the low population were fluorescent, while the mutants from the medium and high populations showed moderate and high fluorescence, even in the presence of STEPtag. When testing the response to STEPtag, we found that 0% of the low, 27% of the medium and 80% of the high populations showed an increase when STEPtag was added, relative to the signal when phosphate buffer was added. This shows that our two-step screen did not work as intended, as the positive signal in the presence of STEPtag was not maintained from the positive sort to the hits from the negative sort. Despite this, a single mutant from the high library showed promise as a sensor, as it was both brighter and showed a larger $\Delta F/F_0$ in the lysate test than gSTEP0-T1-hBim (Figure 3.5). As gSTEP0-T1-hBim was our current best sensor when exciting the anionic chromophore, we selected this improved mutant for further characterization, and named it gSTEP1.

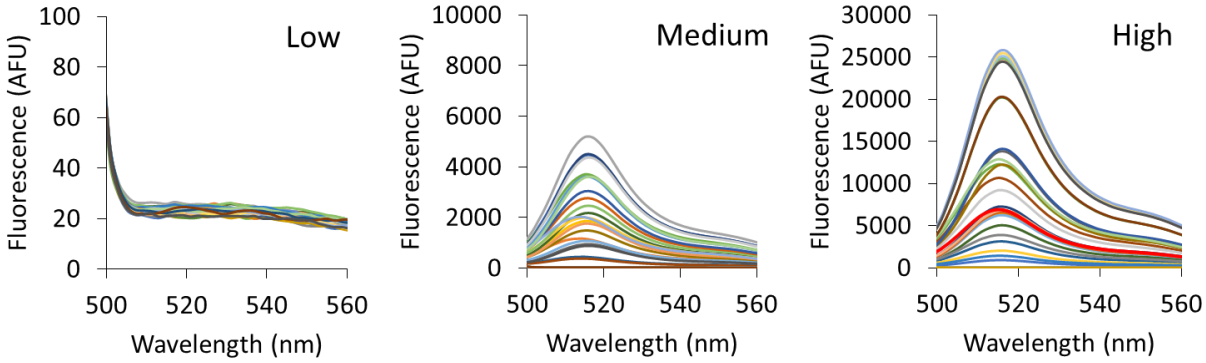


Figure 3.4 Fluorescence emission spectra of mutants selected from the two-step FACS sort. Cells containing individual mutants from the three populations (low, medium and high) selected from the negative sort (Figure 3.3) were cultured in 96-well plates, lysed and equal volumes of lysate were mixed with STEPtag at a final concentration of 7.5 µM. The emission spectra were then measured while exciting at 488 nm. gSTEP1 is indicated in bright red (High panel).

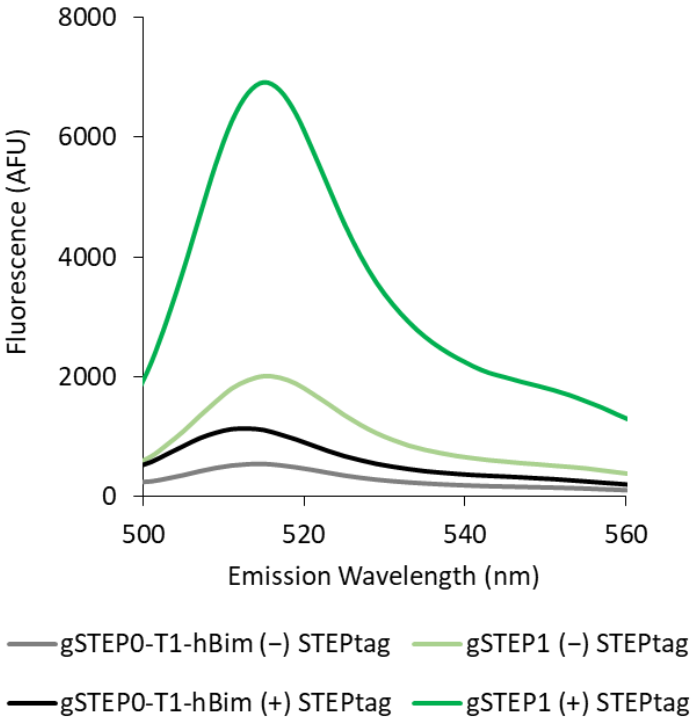


Figure 3.5 Emission spectra comparing gSTEP1 and gSTEP0-T1-hBim in the lysate screen. Cells expressing the gSTEPs were cultured in 96-well plates, lysed and equal volumes of lysate were mixed with either STEPtag at a final concentration of 7.5 µM ((+) STEPtag), or with buffer ((−) STEPtag). The emission spectra was then measured while exciting at 488 nm.

3.3.3 *In vitro* Characterization of gSTEP1

Sequencing of gSTEP1 revealed that the GSGS linker from the parent protein had been mutated to HKIR. This new linker sequence had a significant impact on the excitation spectra of the sensor, restoring the anionic chromophore peak as the major peak, as intended by our positive FACS sort (Figure 3.6). In addition, gSTEP1 appeared to be much brighter than any of the previous mutants, both when comparing the results of the lysate scan (Figure 3.5), as well as qualitatively during purification, as the protein was bright green in colour, even before the addition of STEPtag. We compared the brightness of equal concentrations of gSTEP1, uni-gSTEP0 and GFP S65T, chosen as a positive control with published brightness comparable to that of EGFP (Figure 3.6).[222] We found that in the presence of STEPtag, our sensor now showed similar intensity (roughly 80%) to that of GFP S65T and more than double the brightness of uni-gSTEP0, when exciting at 485 nm. As the spectrum of gSTEP1 is

red-shifted relative to that of GFP S65T (Figure S3.7), this is not actually comparing the peak signal of gSTEP1 to the peak signal of GFP S65T, but once again aiming to compare the signal that would be measured in a standard experiment, where these fluorophores would be excited by a 488 nm laser. Also of note, the increased brightness of gSTEP1 had not come at the expense of a loss in $\Delta F/F_0$, as might be expected if the mutant was simply brighter at all times, leading to a relatively high background signal in the absence of STEPtag. The $\Delta F/F_0$ over six independent purifications was measured to be 3.4 ± 0.5, equivalent to that of gSTEP0-T1-hBim, the previous best mutant. Binding curves also showed a small improvement in K_d relative to that of gSTEP0-T1-hBim, with a measured value of 120 ± 30 nM (Figure 3.7).

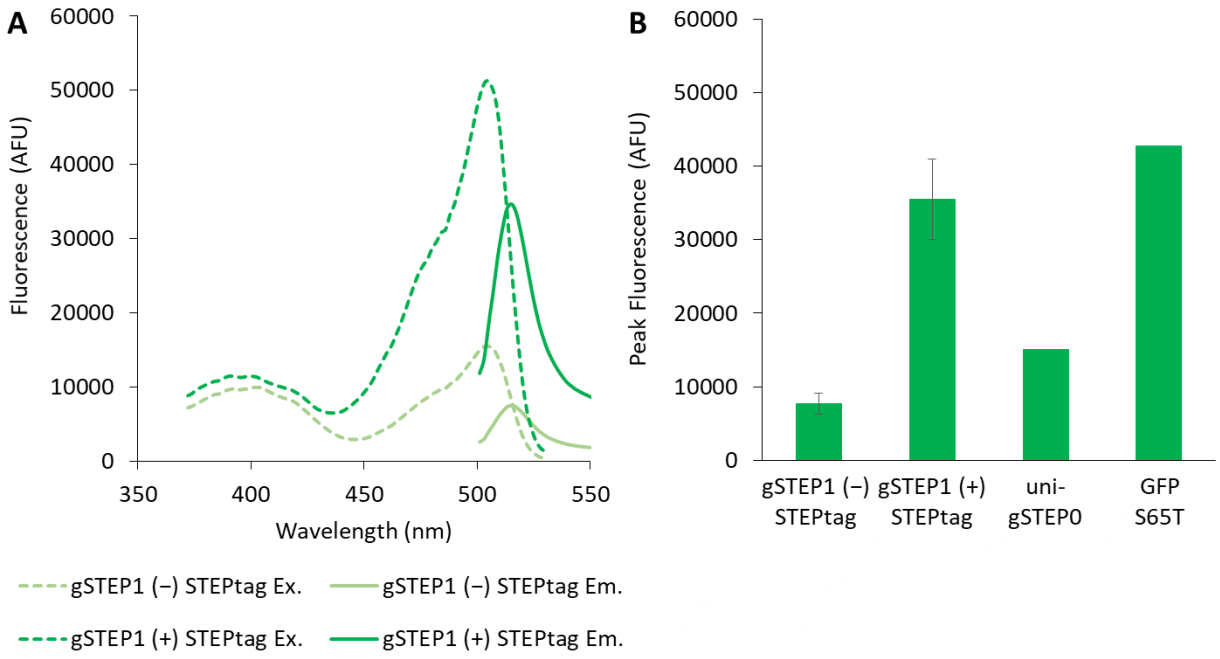


Figure 3.6 Fluorescence intensity of gSTEP1. A, The excitation and emission spectra of gSTEP1 (75 nM), alone and in the presence of 10 µM STEPtag. The excitation spectrum was gathered with an emission wavelength of 550 nm, and the emission spectra was gathered with an excitation wavelength of 485 nm. **B,** The peak fluorescence of gSTEP1 compared to that of uni-gSTEP0 and GFP S65T, when exciting at 485 nm. All fluorescent proteins were measured at 75 nM concentrations. gSTEP1 results are the average of six biological replicates of gSTEP1 and STEPtag. Error bars on the gSTEP1 results represent the standard deviation.

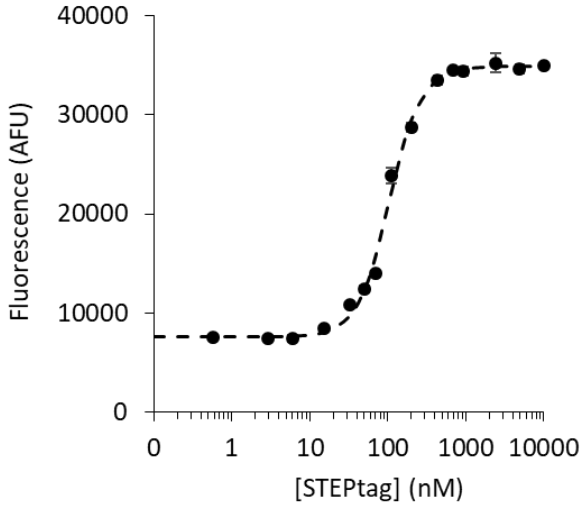


Figure 3.7 Representative binding curve of gSTEP1. Peak fluorescence when exciting at 485 nm of 75 nM of gSTEP1, mixed with a range of concentrations of free STEPtag. The dashed line indicates a fit of the Hill equation to the data (n = 1.8). Error bars are the standard deviation of triplicate measurements.

Having demonstrated that gSTEP1 had improved characteristics relative to the previous generations of the sensor, we measured its binding kinetics, to confirm that it could operate on an appropriate timescale for the detection of transiently expressed proteins. Using a stopped-flow spectrophotometer, gSTEP1 and STEPtag were rapidly mixed while measuring fluorescence (Figure 3.8). We found that the signal from the sensor reached equilibrium within seconds, with the fit of the integrated rate equation (Equation 19 of the introduction) providing a k_{on} of $1.7 \pm 0.1 \times 10^5$ $M^{-1}s^{-1}$, measured across 4 stopped-flow runs with the same preparation of gSTEP1 and STEPtag (Figure 3.8). Given the measured K_d of 120 ± 30 nM from the binding curves, this results in a calculated k_{off} of 0.020 ± 0.005 s^{-1}.

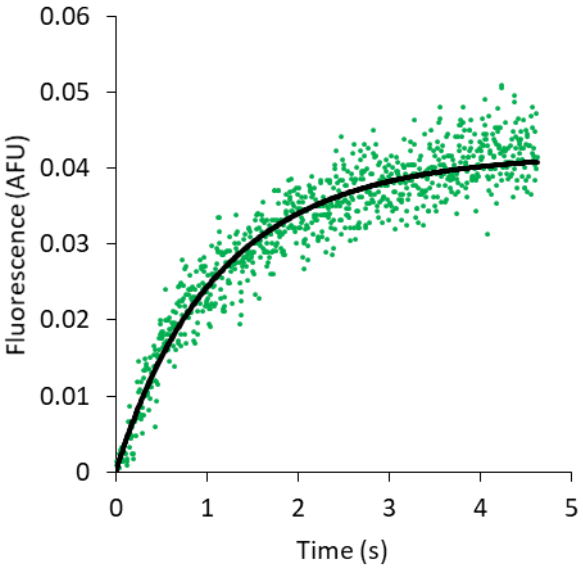


Figure 3.8 Representative stopped-flow experiment with gSTEP1. Measured signal from the stopped-flow photomultiplier tube (PMT) when filtering for excitation at 485 nm, emission at 515 nm. At time = 0s, equal volumes of 1 µM gSTEP1 and 5 µM STEPtag into the mixing chamber were injected, and the fluorescence detector was zeroed, such that the baseline signal from unbound gSTEP1 was removed. The black line represents the fit of Hulme and Trevethick's integrated rate equation (Equation 19 in the introduction)[5] to the data. All measurements were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4), at 37°C

3.3.4 Bacterial Characterization of gSTEP1

While working with the Duet vector in the two-step screening approach, we realized that this cloning strategy was laborious and prone to losses of library DNA during the recloning steps. Additionally, it could not easily be used for *in cellulo* characterization of the STEP, as both gSTEP and STEPtag were under the control of the same promoter. This meant that the only expression states available in a given population of cells were the on state of the STEP, with both halves induced, or no STEP present at all, and thus the off state of the STEP could not be measured. To circumvent this, we identified a system where gSTEP and STEPtag can be expressed independently, by placing the transcription of mRNA for each protein under the control of a separate bacterial promoter. In this system, gSTEP is cloned into the pZA23 vector, under the control of the $P_{A1lacO-1}$ promoter,[223] STEPtag is cloned into the pBAD vector, under the control of the P_{BAD} promoter,[153] and both vectors are transformed into the same *E. coli* cells. Because each of these vectors contains a different origin of

replication (p15A for pZA23 and pBR322 for pBAD) and antibiotic resistance gene (kanamycin for pZA23 and ampicillin for pBAD), they can be stably maintained in the same cell, and the two independent promoters allow induction of either half of the STEP separately, by addition of the appropriate inducer (IPTG for pZA32, and arabinose for pBAD). This creates a system where we can screen gSTEP in the presence and absence of STEPtag in the same cells, without any intermediate cloning steps, simply by modifying the concentration of arabinose present while gSTEP is being induced. In fact, as the pZA23 vector has been designed for use in cell strains with added *laci*^{q} genes, encoding for additional copies of the Lac repressor, we found that when it is present in cell strains without these additional copies of *laci*^{q}, such as the TOP10 cells commonly used for pBAD expression, there is a measurable amount of constitutive expression from the vector in the absence of IPTG. For our goal of *in cellulo* characterization, this is ideal, as the STEP has been designed under the assumption that gSTEP will be expressed long before STEPtag is induced, in order to allow complete maturation of the chromophore, and constitutive expression of gSTEP ensures that this is the case.

Using this pZA23/pBAD expression system, we were able to perform some characterization of the sensor in *E. coli*, using flow cytometry to measure the fluorescence of gSTEP1 in individual cells. With this system, we observed that cells containing gSTEP1 where STEPtag expression was induced by addition of arabinose were considerably brighter than uninduced cells from the same initial culture (Figure 3.9). The $\Delta F/F_0$ observed between the mean fluorescence of these two populations was considerably greater than the *in vitro* values, at 11 ± 4 across two replicates of the experiment. This was sufficient to easily distinguish the two populations by their fluorescence, with little overlap. Also of interest was that under these conditions, cells expressing gSTEP1 and STEPtag were brighter than cells expressing EGFP by about fourfold.

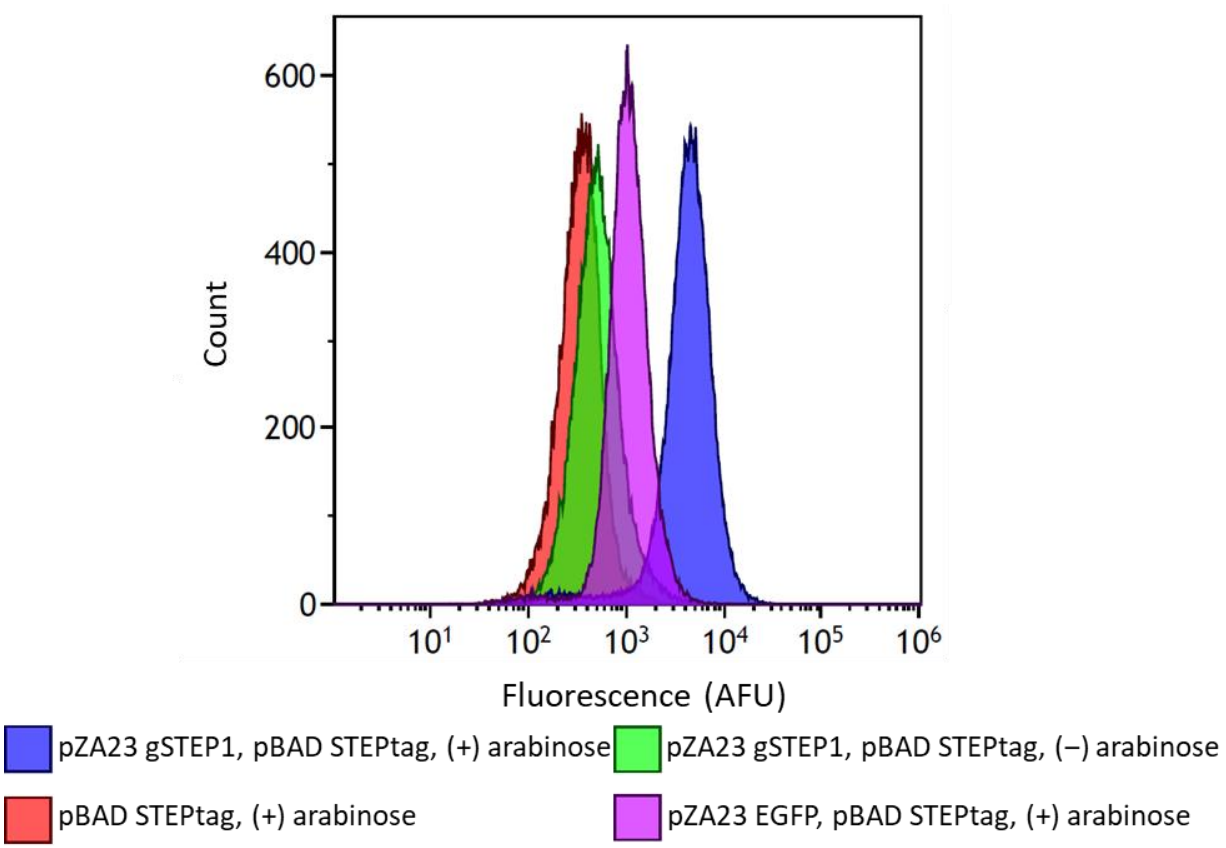


Figure 3.9 Bacterial flow cytometry of cells expressing gSTEP1. Flow cytometry histogram of STEP fluorescence in three cultures of *E. coli* TOP10 cells, containing plasmids expressing either STEPtag alone, gSTEP1 and STEPtag, or EGFP and STEPtag. After growth to mid-log phase, the cells containing both gSTEP1 and STEPtag were split into two cultures, and arabinose was added to the indicated samples to induce STEPtag expression, one hour before the cells were harvested for cytometry. Fluorescence was excited at 488 nm, and emission was measured using a 525BP40 filter.

Having demonstrated that we could detect the increase in bacterial gSTEP1 fluorescence due to STEPtag production after an hour of incubation, our next step was to determine how rapidly this increase could be detected post-induction, as this rate is critical to the end goal of using the STEP to detect transiently expressed proteins. We therefore needed to measure gSTEP1 fluorescence during induction of STEPtag in live cells. To minimize the time between induction and fluorescence measurement, we used a plate reader with an injector system, allowing the first fluorescence readings to be taken seconds after the addition of arabinose. This plate reader was kept at 37°C, with shaking steps every few minutes, to maintain an environment that is reasonable for *E. coli* growth. We once

again compared the gSTEP1 signal to that of EGFP, although in this case EGFP was placed in a pBAD vector, such that it would not be expressed until induction with arabinose. In this way, the assay should mimic the detection of a protein of interest tagged with either STEPtag or EGFP, where the fusion is not expressed until the time of induction. This allowed us to directly compare the time to detection of the STEP versus that of a conventional fluorescent protein fusion tag. Under these conditions, we found that the cells expressing the STEP showed an immediate increase in fluorescence after STEPtag induction (Figure 3.10), while cells expressing EGFP required roughly 8-10 minutes before the signal began to increase. Thus, despite the slightly higher background signal of the STEP, it detected expression more rapidly than EGFP. For instance, the time required for the STEP population to reach 120% of the pre-induction signal was 9 minutes, while the EGFP population required 17 minutes to reach the same relative increase. In fact, the STEP showed a larger relative increase in signal for 25 minutes post-induction, before being overtaken by EGFP. Thus, the STEP provides a signal that cannot be replicated by EGFP for the first 10 minutes post-expression, and will be a more sensitive reporter for the first 25 minutes, satisfying our goal of making a sensor for transiently expressed proteins with lifetimes on the minute timescale.

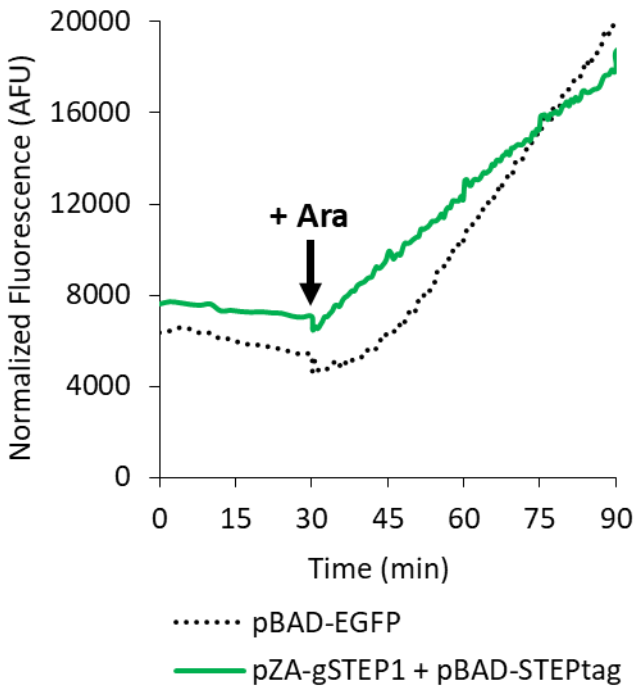


Figure 3.10 Time course of cellular fluorescence during reporter protein induction. Whole-cell fluorescence emitted at 514 nm, measured when exciting at 488 nm, normalized to the OD 600 of the sample to account for variations in culture density. *E. coli* cells were cultured in LB broth to mid-log phase, then transferred to a 96-well plate in a plate reader incubating at 37°C. For the first 30 minutes, fluorescence and absorbance measurements were taken every 2 minutes, shaking prior to measurement. At 30 minutes, 0.23% arabinose was added to induce pBAD expression (+Ara), and measurements were taken every 23 seconds afterwards.

3.4 Discussion

Overall, the two rational design hypotheses did lead to improvements in the STEP, as both gSTEP0-T1-hBim and gSTEP0-T1-L4 outperformed the 250 nM K_d and 2.2 $\Delta F/F_0$ measured for the parent gSTEP0-T1. When testing to see if an increased linker length between the Bim peptide and cpGFP could improve the sensor, gSTEP0-T1-L4 showed the anticipated improvement by lowering the K_d to 130 nM, presumably by allowing the bound STEPtag to adopt a more stable position next to the GFP barrel, without having a large effect on the change in fluorescence. For the mutants with 1 and 2 amino acid linkers, the loss of $\Delta F/F_0$ was interesting, implying that these relatively subtle changes can have a large effect on the functionality of the STEP, potentially by rotating the Bim peptide in a way the moves the bound STEPtag away from the pore. As these mutants were negatively impacting

the functionality of the sensor, we did not continue investigating this effect. Interestingly, when we first tested the unimolecular STEP constructs (Chapter 2), the construct with 5 amino acid linkers and an N-terminal Bim peptide showed no increase in fluorescence relative to cpGFP (Chapter 2), while the split constructs with 4 and 5 amino acid linkers here show an increase that outperforms the 0 amino acid linker sensor. This could be because the 5 amino acid linker between STEPtag and the cpGFP in the unimolecular construct did not allow the proper positioning of the binding pair to improve fluorescence, and so the improvement due to the increased linker length between Bim and cpGFP was missed.

By changing the sequence of Bim peptide used, we also created a gSTEP mutant with a lower K_d (160 nM) when binding STEPtag, where the human Bim peptide is used rather than the mouse sequence. This improvement was not necessarily maintained across the synthetic mutants tested, although they were expected to have similar K_d values to that of the human sequence, based on their published affinities. The increase in $\Delta F/F_0$ to 3.3 when using the human Bim sequence was likewise not anticipated, but it can potentially be explained by changes to the C-terminus of the Bim peptide, closest to the pore. In the original mouse Bim, the C-terminal sequence is NETYTRR, while in human Bim, it is instead NAYYARR. It is possible that these positions are close enough to the chromophore to have an effect on its fluorescence when STEPtag binds, an observation that is supported by the shift in the fluorescence excitation spectra towards the neutral chromophore peak when combining the 4-amino acid linker with the human Bim peptide.

When attempting to revert this change in the spectra through linker saturation, the two-step FACS screen did test considerably more mutants then would have been possible using 96-well plates or purified proteins, screening tens of thousands of mutants. Yet it was not successful at identifying mutants with improved $\Delta F/F_0$ by comparing cells containing STEPtag to cells without it, as the best hit, gSTEP1, came from the brightest cells in the negative sort. In fact, the majority of the mutants in

the dimmer populations were either nonfluorescent or did not show an increase in fluorescence when purified STEPtag was added. This lack of an increase implies that the two-step screen was not effectively retaining mutants with sensor activity, i.e. an increase in fluorescence when STEPtag is added. This loss of sensor activity could result from the positive sort not only detecting mutants that were bright due to the presence of STEPtag, but rather selecting cells that were simply bright overall, either by increased expression or an increased fluorophore brightness that is independent of STEPtag presence. The negative sort also shows that, after the recloning and second expression of the mutants, a large proportion of the cells no longer express fluorescent protein, much less effective sensors. Thus, FACS sorting of STEP mutants would be much more efficient if these steps could be omitted, and both on and off state fluorescence could be measured in the same cell. This could be done with the pZA/pBAD expression system we developed after the original FACS experiment, and we plan on testing this system for future improvements to the STEP.

Nevertheless, by testing multiple 96-well plates of mutants from the two-step screen, a hit was obtained, gSTEP1. This generation of the sensor binds as tightly as gSTEP0-T1-L4, with a K_d of 120 ± 30 nM, and has a change in fluorescence upon binding equivalent to that of gSTEP0-T1-hBim, with a $\Delta F/F_0$ of 3.4 ± 0.5. This improvement is caused by the replacement of the GSGS linker with HKIR, a more positively charged sequence. Interestingly, a number of red and green GECIs contain a residue near the pore that is capable of H-bonding to the chromophore phenolate.[224] A similar mechanism could help modulate fluorescence in gSTEP1, if the added linker residues are able to directly interact with the chromophore. Our efforts to confirm this possibility by obtaining a crystal structure are ongoing, so this remains speculation for now. Regardless, our high-throughput screening method did result in the best gSTEP sensor yet. The fact that the on state of the sensor was of similar brightness to GFP S65T in our *in vitro* test was also heartening, implying that our signal should be detectable using standard fluorescence microscopy methods, as EGFP is still in common use there, and the brightness

of GFP S65T and EGFP is almost identical, at 35.2 for GFP S65T versus 33.5 for EGFP.[1] Finally, the stopped-flow experiment provided confirmation of our hypothesis that the binding of gSTEP to STEPtag leads to an increase in fluorescence that is observable on a sub-minute timescale, reaching equilibrium in a few seconds at the concentrations tested. This rate is considerably faster than the maturation rate of even the fastest-maturing fluorescent protein, Venus, a yellow fluorescent protein that still has a maturation half-time of 2 minutes.[44]

Having demonstrated that we could detect the STEP *in vitro*, we then moved on to bacterial characterization. For the gSTEP1 experiments, the pZA23/pBAD expression system that we developed allowed us to compare cells from the same population before and after STEPtag induction, which was not previously effective in the gSTEP0-T1 tests (Chapter 2), performed with the pOXB/pACYC-Duet vectors, and impossible with the pET-Duet vector used for the FACS sort. When comparing cells using flow cytometry, the cells containing gSTEP1 with STEPtag induced by arabinose addition were clearly distinct from the same cells that had not had STEPtag induced, and the cells without STEPtag induced did show the expected low background fluorescence, as they were only slightly brighter than the control cells expressing only STEPtag. Interestingly, in this experiment the cells containing both gSTEP1 and STEPtag were brighter than the cells expressing EGFP, unlike in the *in vitro* data, where GFP S65T was found to be brighter than gSTEP1. Initially, we considered that the red-shift in the gSTEP1 spectra could contribute to this increase. Our flow cytometer uses a 525BP40 emission filter, meaning that it allows light between 505 and 545 nm to pass through to the detector. As the EGFP emission peak can be found at 511 nm, and the emitted fluorescence is still at 90% of the peak value at 505 nm,[1] a considerable amount of the emission spectra is being cut off by the filter. The gSTEP1 emission peak, on the other hand, occurs at 515 nm, and the emitted fluorescence drops to 50% of maximum at 505 nm, allowing a larger fraction of total gSTEP1 fluorescence to be detected (Figure S3.13). But the EGFP peak is also wider than that of gSTEP1, and so the total integrated signal between

505 and 545 nm ended up being quite similar, with the gSTEP1 signal being about 6% greater. As the increased signal was reproducible over two separate cultures (Figure 3.9 and Figure S3.16), it seems that some part of our preparation for flow cytometry is particularly conducive to visualizing the gSTEP1 signal. Most importantly, we demonstrated that the on and off states of the STEP could be directly compared between cells containing sequences for both gSTEP1 and STEPtag, where in the previous experiments we could only distinguish populations where the vector containing the STEPtag gene had been added or removed by cloning.

This ability to compare the on and off states of the STEP in the same cell population finally allowed us to perform a kinetic *in cellulo* experiment, in order to determine the time required for gSTEP1 to detect expression of STEPtag. As shown in Figure 3.10, the STEP signal begins to increase linearly almost immediately after induction of STEPtag, while the EGFP signal shows a lag phase, remaining relatively stable for the first 8-10 minutes before curving upwards. The experimental setup allowed us to measure the fluorescent signal every 23 seconds post-induction, and these individual data points show that the STEP culture does require about three points, or just over one minute, to stabilize and begin to increase post-induction (Figure S3.11), although some of this instability could be due to the injection and mixing of the arabinose used to induce STEPtag production, as both the STEP and EGFP cultures show a small drop in fluorescence during the induction step. In the original pBAD paper,[153] the authors were able to detect expression from the pBAD promoter within roughly a minute, so it seems that STEPtag is being detected about as early as possible. It should be clarified that the time to reach the 120% fluorescence threshold reported in section 3.3.4 considers the baseline fluorescence to be the pre-induction level, before the drop due to induction. If we instead consider the baseline to be the values just after induction, the STEP reaches the 120% threshold in 6 minutes, versus 11 for EGFP. In either case, the STEP attains this early threshold in slightly over half the time required for

EGFP. This experiment provides the proof-of-principle that our sensor can outperform a fluorescent protein at the rapid detection of protein expression in a living organism.

Having demonstrated that gSTEP1 can rapidly detect STEPtag both *in cellulo* and *in vitro*, the binding kinetics and signal-to-noise ratio of the sensor are therefore sufficient for our tests, although they could still be improved. Thus, the gSTEP1 and STEPtag pair open the door to the use of our sensor for transiently expressed proteins for its intended purpose: detecting transiently expressed proteins. With a theoretical response on the timescale of seconds, and a detectable signal that appears minutes earlier than EGFP at the population level, the STEP is ready to be tested in research model organisms.

3.5 Materials and Methods

3.5.1 Construct Preparation

gSTEP0-T1 mutant sequences were obtained using splicing by overlap extension (SOE) mutagenesis,[225] in order to insert the appropriate linkers or perform site-directed mutagenesis on the sequence. Briefly, synthetic oligonucleotides (Eurofins) were ordered containing the desired mutations, as well as sequences complementary to the parent gene flanking the mutations, for both the sense and anti-sense strands. These oligonucleotides were then used in conjunction with oligonucleotides flanking the gSTEP0-T1 gene to perform PCR amplification of gSTEP0-T1, generating two sequences, one of which contains the desired sequence from the 5’ end to the mutation site, and the other that contains the desired sequence from the mutation site to the 3’ end. The full mutant genes were then generated by mixing these two sequences, which overlap at the desired mutation site, and performing PCR amplification using the previous flanking oligonucleotides. These sequences were then subcloned into pET11a vectors (Novagen), using the NdeI and BamHI restriction sites. The mutant library of gSTEP0-T1-L4-hBim was generated in the same manner as above, using NNS degenerate codons in place of all four of the linker amino acids. The library was then subcloned

into MCS2 of a pETDuet-1 vector (Novagen), using the NdeI and XhoI restriction sites, along with STEPtag containing a HindII restriction site upstream of the coding sequence, which was cloned into MCS1 using the NcoI and BamHI restriction sites. All reagents used for subcloning were obtained from New England Biolabs. A pET11a vector containing the gene encoding GFP S65T was a generous gift from Stephen L. Mayo, and a pET11a vector containing the gene encoding EGFP was a generous gift from James A. Davey. Constructs for flow cytometry and bacterial characterization were subcloned into either pBAD/His A (Invitrogen), or pZA23MCS (EXPRESSYS), in the same manner as above, using either NcoI/EcoRI or HindIII/BamHI restriction sites, respectively, except for EGFP, where XhoI/EcoRI were used to subclone the gene into pBAD/His A. This cloning strategy removed the His tag, Xpress epitope and EK site from the pBAD vector for the STEP constructs, as our constructs already contained a His tag for purification. These vector components were retained for EGFP, as it was cloned in-frame with them.

3.5.2 Protein Expression and Purification for *in vitro* screening and characterization

All vectors for *in vitro* use were transformed into chemically competent *E. coli* BL21 (DE3) Gold (Agilent). Expression was performed by culturing cells in Luria-Bertani (LB) broth, supplemented with 100 µg/mL ampicillin. Cells were grown at 37°C with shaking to an OD600 of 0.6-0.8, at which point protein expression was induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were then incubated with shaking overnight at 16°C, harvested by centrifugation and lysed using an EmulsiFlex-B15 cell disruptor (Avestin). Proteins were purified by immobilized metal affinity chromatography using Profinity Nickel-charged IMAC resin (Bio-Rad), in a gravity flow column, according to the manufacturer's protocol. Buffer exchange into 20 mM sodium phosphate containing 50 mM NaCl (pH 7.4) and sample concentration was performed using Amicon Ultra-15 centrifugal filters with a 3K molecular weight cut-off (Millipore) for STEPtag, and Microsep Advance centrifugal filters with a 10K molecular weight cut-off (Pall) for all other proteins.

3.5.3 Fluorescence Assays

Proteins were quantified by measuring absorbance at 280 nm in a cuvette using a SpectraMax Plus^{384} microplate spectrophotometer (Molecular Devices). Extinction coefficients for the proteins were estimated using the ProtParam online tool.[221] All measurements were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4), at room temperature. Fluorescence spectra of all proteins were recorded on an Infinite M1000 microplate reader (Tecan). All fluorescence measurements were performed using 75 nM of fluorescent protein, and each sample well was prepared in triplicate. To calculate the K_d, the fluorescence signal when exciting at 485 nm and emitting at 515 nm was used to fit the Hill equation, accounting for ligand depletion,[5] across the range of STEPtag concentrations used. $\Delta F/F_0$ values for all the binding curves were calculated using the Hill equation fit to determine F_{max} and F_{min}, where F_{max} is the maximum signal of the fit, and F_{min} is the minimum signal of the fit.

3.5.4 FACS Sorting and Library Screening

For the positive sort, the mutant linker library with STEPtag present in pET-Duet-1 was transformed into *E. coli* E. cloni ELITE electrocompetent cells (Lucigen), and plated onto agar plates supplemented with 100 µg/mL ampicillin. Roughly 100 000 colonies were collected, precultured overnight in LB broth supplemented with ampicillin, and the preculture was miniprepped using an Omega miniprep kit. The miniprepped DNA was transformed into electrocompetent *E. coli* BL21 (DE3) Gold for expression, and plated onto agar plates supplemented with ampicillin. Roughly 100 000 colonies were again collected, cultured overnight in LB broth supplemented with ampicillin, and the overnight cultures were used to seed fresh cultures in LB broth supplemented with ampicillin. These cultures were grown to an OD600 of 0.5-0.9, then were diluted to a concentration of roughly 5 x 10^7 CFU/mL, assuming that 1 OD600 = 8 x 10^8 CFU/mL for *E. coli*.[226] The cells were washed twice with filtered PBS, then passed through a 40 µm Falcon Cell Strainer (Fisher) twice. FACS sorting was

done at the University of Ottawa Faculty of Medicine Core Facility, using a MoFlo AstriosEQ Cell Sorter (Beckman Coulter). Fluorescence was detected by exciting with a 488 nm laser, and detecting emission with a 513/26 nm filter. Data analysis was performed using the FlowJo software package (BD). The top 20 000 cells were sorted, diluted into 50 mL of LB broth supplemented with ampicillin, and cultured overnight.

For the negative sort, the overnight culture of the positively sorted cells was miniprepped, and the STEPtag sequence was removed from the pET-Duet-1 vector by digestion with HindIII. The vector was purified by agarose gel electrophoresis, followed by gel extraction. The purified vector was recircularized by ligation with the Quick Ligation Kit (NEB), and retransformed into electrocompetent BL21 (DE3) Gold cells. The remaining steps were performed as for the positive FACS sort, with the exception that the FACS sorting was performed at the University of Ottawa Cellular Imaging and Cytometry Facility, using the same make of instrument, and data analysis was performed using the Kaluza software package.

To screen the $\Delta F/F_0$ of the sorted cells, the overnight cultures from the negative sort were used to streak LB agar plates supplemented with ampicillin, and colonies were picked into 96-well plates containing LB broth supplemented with ampicillin. Plates were grown overnight at 37°C with shaking, then were used to seed fresh plates containing TB Overnight Express (Novagen) for expression. The expression plate was also grown overnight at 37°C with shaking, then the cells were lysed with Bugbuster (Millipore). The clarified lysate was then mixed with either phosphate buffer or STEPtag, and fluorescence scans were recorded on an Infinite M1000 microplate reader (Tecan).

3.5.5 Stopped-Flow Kinetic Assay

Stopped-flow measurements were performed using an RSM1000 Stopped-flow spectrophotometer (Olis). Plane gratings used were 400 lines/mm, 500 nm blaze wavelength, all fixed

slits were set to 3.16 mm to maximize signal, and the fixed-disk was used for single wavelength measurements, with a 1.24 mm slit. gSTEP1 and STEPtag were purified and quantified as described in section 3.5.2, and were loaded into the stopped-flow unit, which was kept at 37°C using a temperature control unit (Julabo). 300 µL of each sample was fired into the mixing chamber per shot, and the fluorescence was measured with the monochromators set to an excitation wavelength of 485 nm, and an emission wavelength of 515 nm. For each combination of samples, one shot was fired prior to data collection, to clear the dead volume of the system. All data were exported to Excel for analysis.

3.5.6 Flow Cytometry

Constructs used were transformed into chemically competent TOP10 *E. coli* cells (Invitrogen). Cells were cultured in LB broth supplemented with 100 µg/mL ampicillin (for cells containing pBAD STEPtag) and 50 µg/mL Kanamycin (for cells containing pZA23 gSTEP1 or EGFP). Cells were grown with shaking at 37°C to an OD600 between 0.4-0.8, then the pZA23 gSTEP1 pBAD STEPtag culture was split equally into two flasks, the (+) arabinose flasks were induced with 0.2% arabinose, then all cultures were incubated for a further 60 minutes. Cells were then harvested by centrifugation, washed twice in filtered 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4), and diluted to an approximate concentration of 10^7 cells/mL with filtered 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). All cells were passed through a 40 µm Falcon Cell Strainer (Fisher) immediately before beginning flow cytometry. Flow cytometry measurements were performed using a Gallios flow cytometer (Beckman Coulter), set to detect either 10 000 or 100 000 events per run. Fluorescence was detected by exciting with a 488 nm laser, and detecting emission with a 525BP40 filter. Data analysis was performed using the Kaluza software package (Beckman Coulter).

3.5.7 Plate-Based Bacterial Kinetic Assay

Constructs used were transformed into chemically competent TOP10 *E. coli* cells (Invitrogen). Cells were cultured in LB broth supplemented with 100 µg/mL ampicillin (for cells containing pBAD

STEPtag) and 50 µg/mL Kanamycin (for cells containing pZA23 gSTEP1 or EGFP). Cells were grown with shaking at 37°C to an OD600 between 0.5-0.6, then 200 µL of cells were loaded into a 96-well plate. Fluorescence and absorbance measurements were recorded on an Infinite M1000 microplate reader (Tecan), preheated to 37°C. Measurements were taken every 2 minutes for 30 minutes, shaking the plate before measuring, then expression was induced by injecting 6 µL of 8% arabinose into the wells, for a final concentration of 0.23%. Fluorescence and absorbance measurements were then taken as rapidly as possible for 1 hour, shaking the plate every 15 minutes. Finally, measurements were taken every 2 minutes for the next 5 hours, with shaking prior to every measurement, to track cell growth over time.

3.6 Supplementary Information

3.6.1 Amino Acid Sequences

gSTEP0-T1-L1

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRGLENVYIKADKQKNGIKANFKIRHNIEDGGVQLA
YHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMV
SKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGV
QCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGN
ILGHKLEY*

gSTEP0-T1-L2

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRGSLENVYIKADKQKNGIKANFKIRHNIEDGGVQL
AYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSM
VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYG
VQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG
NILGHKLEY*

gSTEP0-T1-L3

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRGSGLENVYIKADKQKNGIKANFKIRHNIEDGGVQ
LAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGS
MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTY
GVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKED
GNILGHKLEY*

gSTEP0-T1-L4

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRGSGSLENVYIKADKQKNGIKANFKIRHNIEDGGV
QLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGG
SMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLT
YGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEY*

gSTEP0-T1-L5

MHHHHHHDLRPEIRIAQELRRIGDEFNETYTRRGSGSGLENVYIKADKQKNGIKANFKIRHNIEDGG
VQLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTG
GSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFK
EDGNILGHKLEY*

gSTEP0-T1-hBim

MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAY
HYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI
LGHKLEY*

gSTEP0-T1-XXA1

MHHHHHHDLRPEIWYAQGLKRFGDEFNAYYARRLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAY
HYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI
LGHKLEY*

gSTEP0-T1-XXA4

MHHHHHHDLRPEIWYAQWLKRFGDQFNAYYARRLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAY
HYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI
LGHKLEY*

gSTEP0-T1-G2gE

MHHHHHHDLRPEIWYAQELKRFGDEFNAYYARRLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAY
HYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI
LGHKLEY*

gSTEP0-T1-Y4eK

MHHHHHHDLRPEIWYAQGLKRFGDEFNAYKARRLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAY
HYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI
LGHKLEY*

gSTEP0-T1-hBim-L4

MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRGSGSLENVYIKADKQKNGIKANFKIRHNIEDGGV
QLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGG
SMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLT
YGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEY*

gSTEP1

MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRLENVYIKADKQKNGIKANFKIRHNIEDGGV
QLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGG
SMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLT
YGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEY*

GFP S65T

MASMGSHHHHHHSGSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLP
VPWPTLVTTFTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVN
RIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPI
GDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK*

EGFP (in pZA23)

MASMGSHHHHHHSGSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLP
VPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVN
RIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPI
GDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK*

EGFP (in pBAD)

MGGSHHHHHHGMASMTGGQQMGRDLYDDDDKDRWGSELEGSKGEELFTGVVPILVELDGDVNGHKFS
VSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQE
RTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIK
VNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGI
TLGMDELYK*

3.6.2 Sample SDS-PAGE Gels

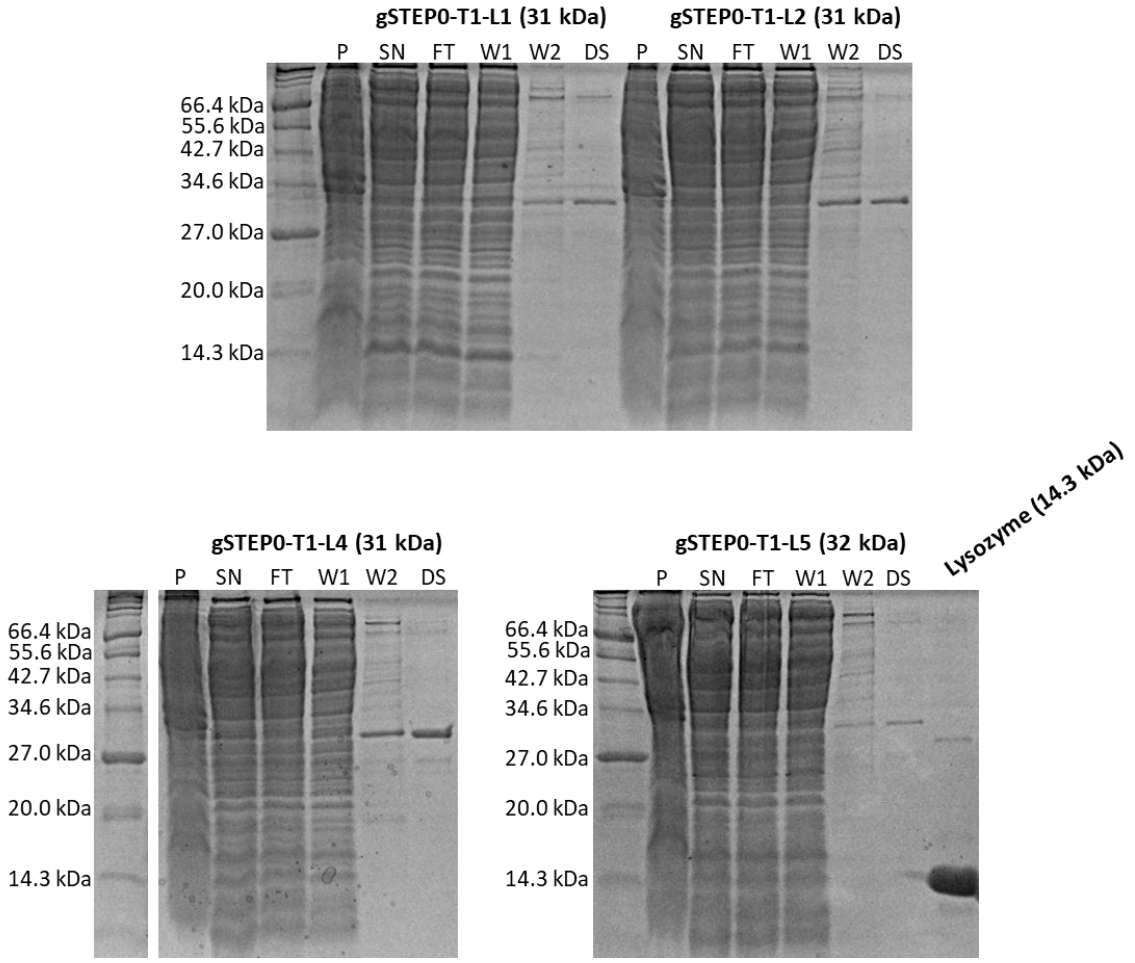


Figure S3.1 Sample SDS-PAGE gels of purification steps. Samples taken during protein purification were mixed 1:1 by volume with SDS-PAGE loading dye, heated to 95°C for 5 minutes, then loaded into 15% acrylamide SDS-PAGE gels. The ladder used is the P7702 Unstained Protein Marker, Broad Range (NEB). Electrophoresis was performed at 120 volts until the dye front had migrated to the end of the gel. Abbreviations are as follows: P – pellet, SN – supernatant, FT – flowthrough, W1 – wash 1, W2 – wash 2, DS – desalted (purified). Lysozyme is present on the gSTEP0-T1-L5 gel for an unrelated experiment, it was included here as it seems to have spread into the desalted lane.

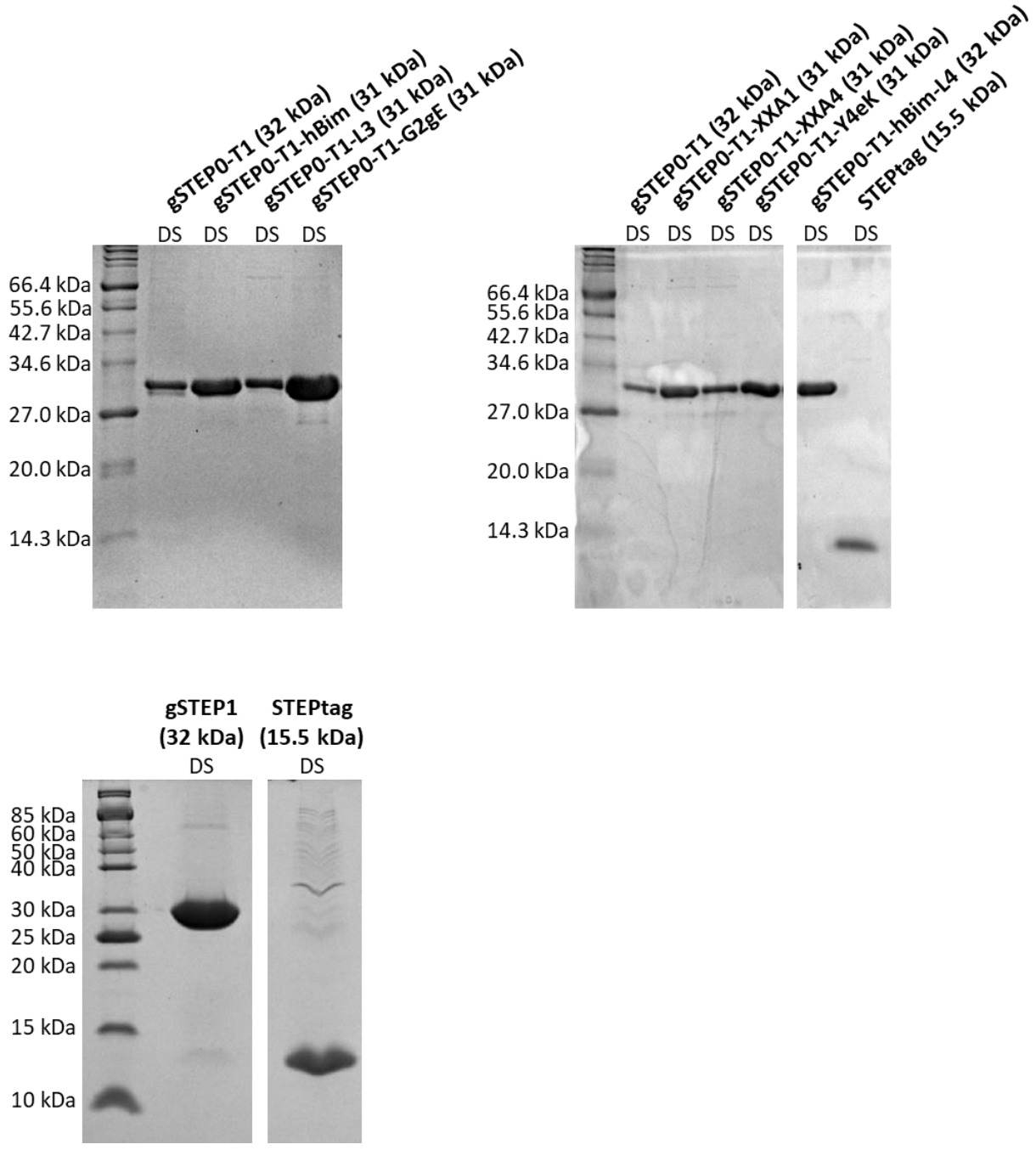
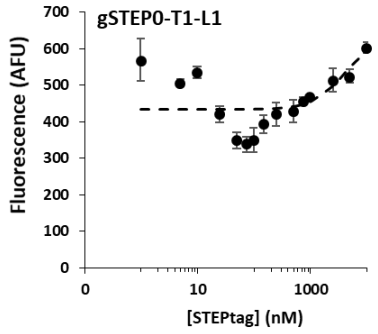
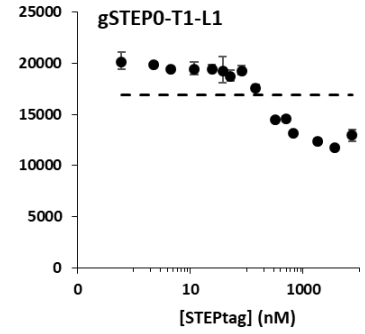
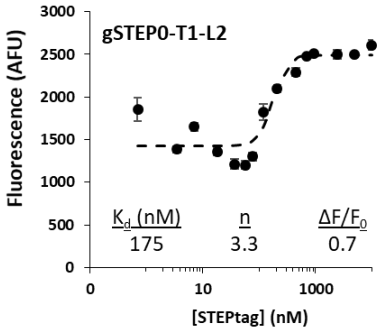


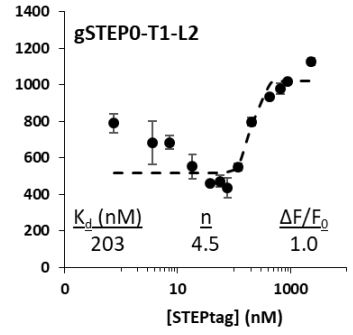
Figure S3.2 Sample SDS-PAGE gels of purified protein. Samples taken of desalted (DS) protein were mixed 1:1 by volume with SDS-PAGE loading dye, heated to 95°C for 5 minutes, then loaded into 15% acrylamide SDS-PAGE gels. The ladder used is the P7702 Unstained Protein Marker, Broad Range (NEB) (top gels), or the P7704 Unstained Protein Marker, Broad Range (NEB) (bottom gel). Electrophoresis was performed at 120 volts until the dye front had migrated to the end of the gel.

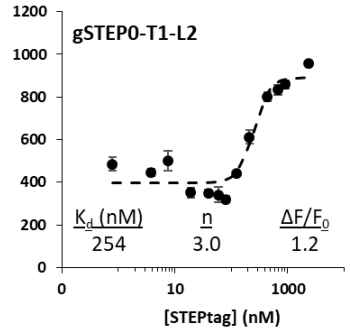
3.6.3 Raw Fluorescence Data

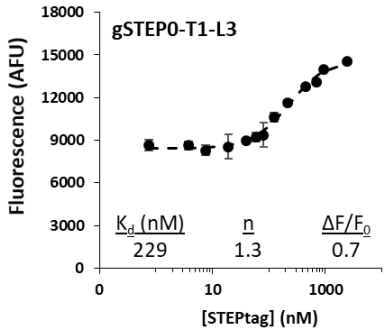


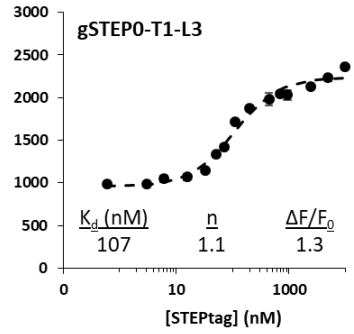


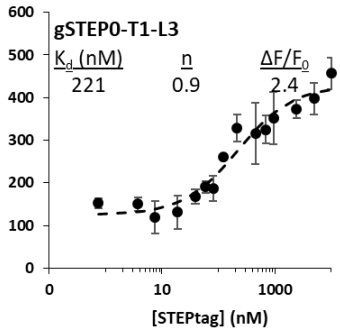












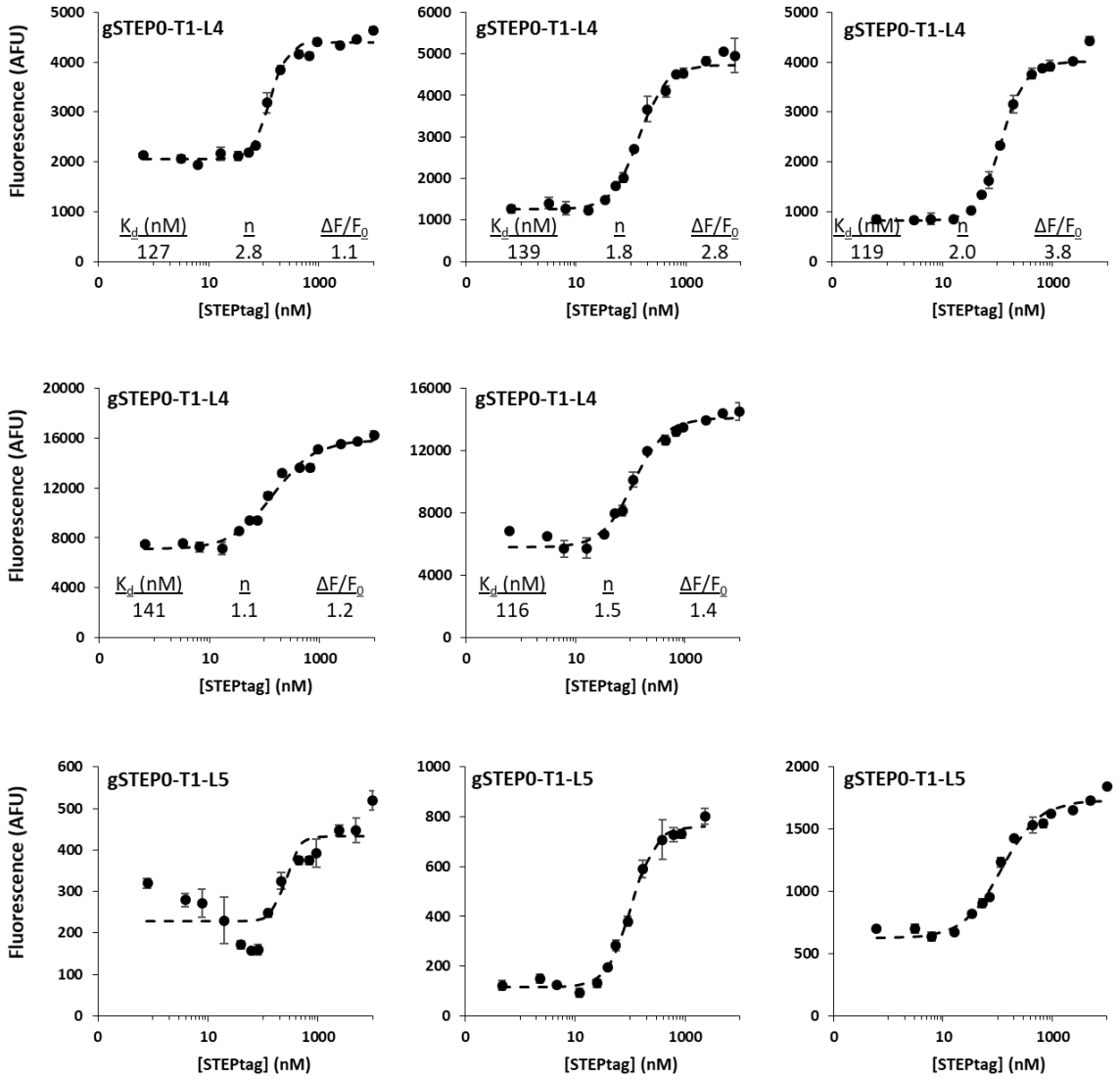
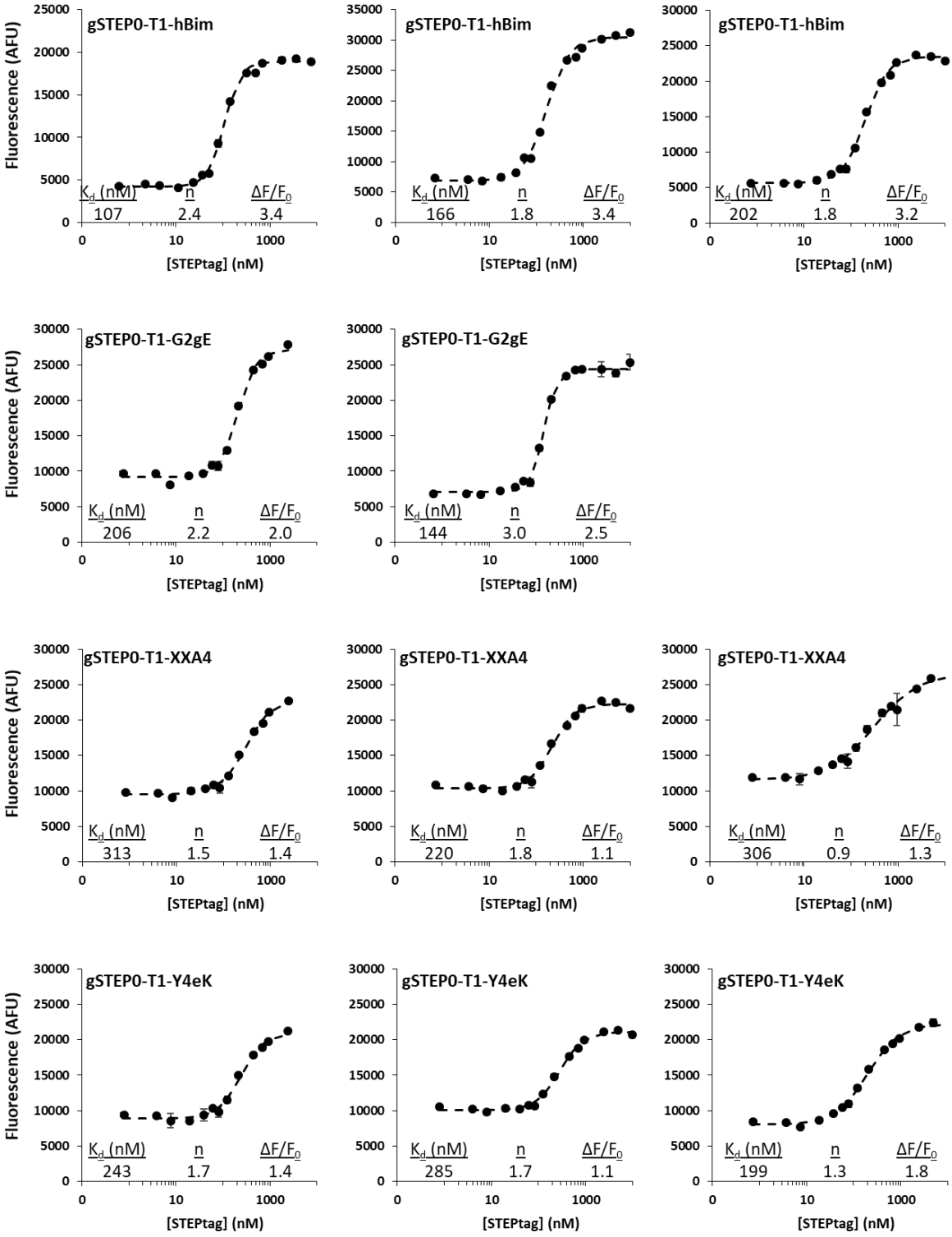


Figure S3.3 Binding curves of gSTEP0-T1 constructs with increased linker length. Emission scans of 75 nM of biological replicates of purified gSTEP0-T1 mutants, mixed with varying concentrations of STEPtag, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). For each mutant, each experiment was performed with a fresh biological replicate of STEPtag. The fluorescence signal is the emission at 515 nm, when exciting at 485 nm. The dashed lines are the results of a fit of the Hill equation to the data, and K_d, $\Delta F/F_0$ and the value of n calculated from the Hill equation fit are reported for each curve. Error bars are the standard deviation of triplicate measurements.


gSTEP0-T1-hBim
Fluorescence (AFU)
[STEPtag] (nM)
Kd (nM) n ΔF/F0
107 2.4 3.4
gSTEP0-T1-hBim
[STEPtag] (nM)
Kd (nM) n ΔF/F0
166 1.8 3.4
gSTEP0-T1-hBim
[STEPtag] (nM)
Kd (nM) n ΔF/F0
202 1.8 3.2
gSTEP0-T1-G2gE
Fluorescence (AFU)
[STEPtag] (nM)
Kd (nM) n ΔF/F0
206 2.2 2.0
gSTEP0-T1-G2gE
[STEPtag] (nM)
Kd (nM) n ΔF/F0
144 3.0 2.5
gSTEP0-T1-XXA4
Fluorescence (AFU)
[STEPtag] (nM)
Kd (nM) n ΔF/F0
313 1.5 1.4
gSTEP0-T1-XXA4
[STEPtag] (nM)
Kd (nM) n ΔF/F0
220 1.8 1.1
gSTEP0-T1-XXA4
[STEPtag] (nM)
Kd (nM) n ΔF/F0
306 0.9 1.3
gSTEP0-T1-Y4eK
Fluorescence (AFU)
[STEPtag] (nM)
Kd (nM) n ΔF/F0
243 1.7 1.4
gSTEP0-T1-Y4eK
[STEPtag] (nM)
Kd (nM) n ΔF/F0
285 1.7 1.1
gSTEP0-T1-Y4eK
[STEPtag] (nM)
Kd (nM) n ΔF/F0
199 1.3 1.8

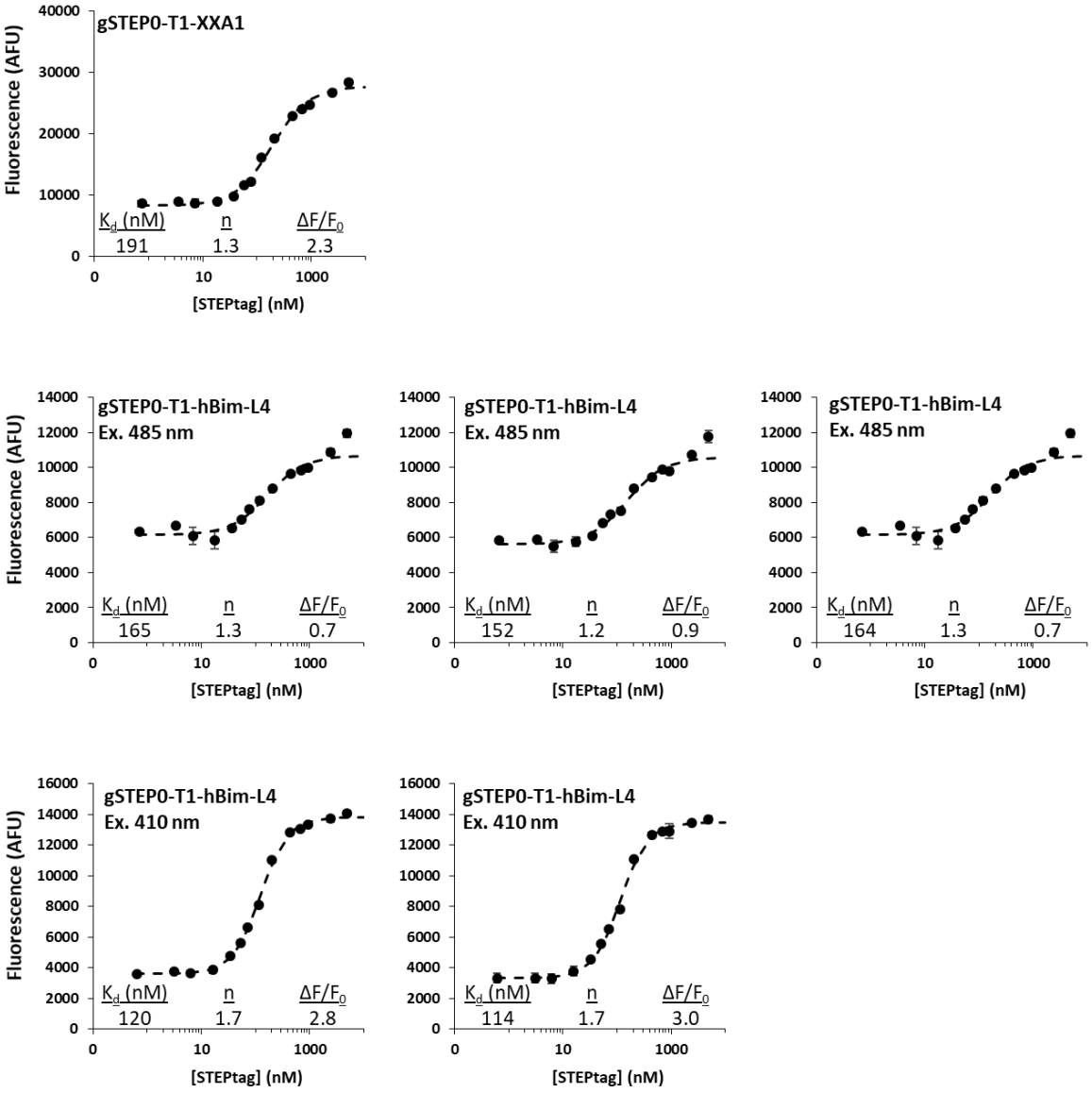


Figure S3.4 Binding curves of gSTEP0-T1 constructs with different binding peptides. Scans of 75 nM of biological replicates of purified gSTEP0-T1 mutants, mixed with varying concentrations of STEPtag, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). For each mutant, each experiment was performed with a fresh biological replicate of STEPtag. The fluorescence signal is the emission at 515 nm, when exciting at 485 nm, or 410 nm where indicated for gSTEP0-T1-hBim-L4. The dashed lines are the results of a fit of the Hill equation to the data, and K_d, $\Delta F/F_0$ and the value of n calculated from the Hill equation fit are reported for each curve. Error bars are the standard deviation of triplicate measurements.

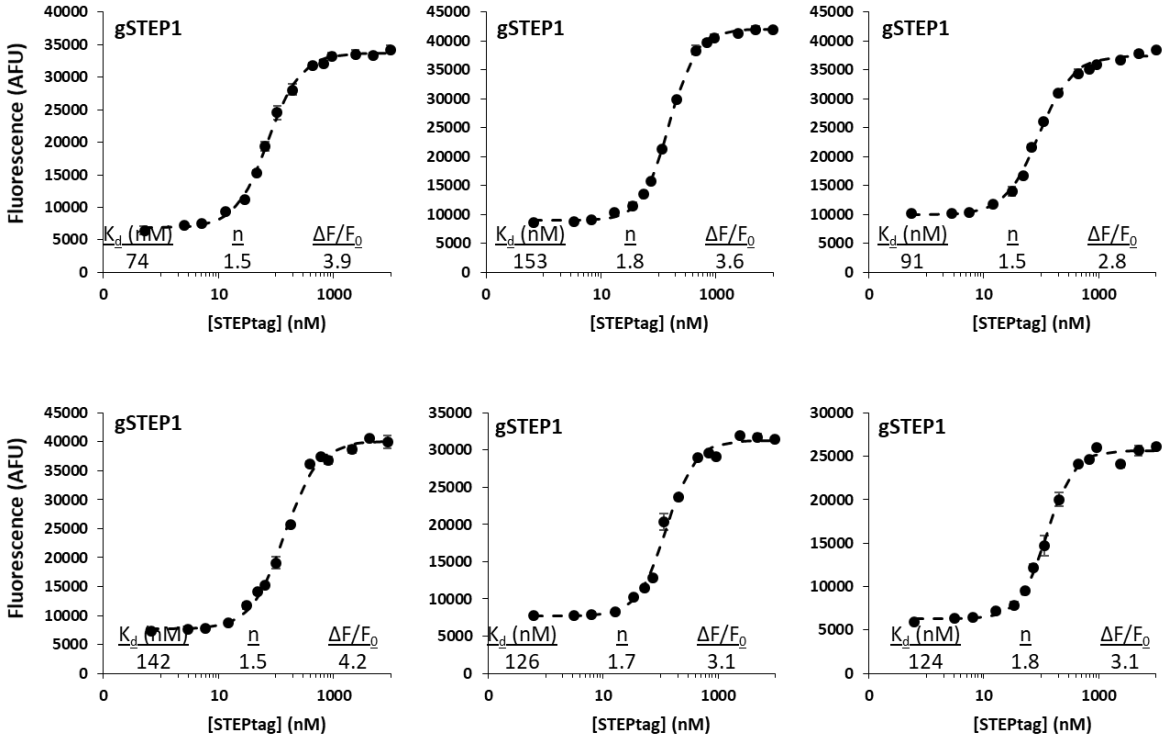


Figure S3.5 Binding curves of gSTEP1. Scans of 75 nM of biological replicates of purified gSTEP1, mixed with varying concentrations of STEPtag, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). For each mutant, each experiment was performed with a fresh biological replicate of STEPtag. The fluorescence signal is the emission at 515 nm, when exciting at 485 nm. The dashed lines are the results of a fit of the Hill equation to the data, and K_d, $\Delta F/F_0$ and the value of n calculated from the Hill equation fit are reported for each curve. Error bars are the standard deviation of triplicate measurements.

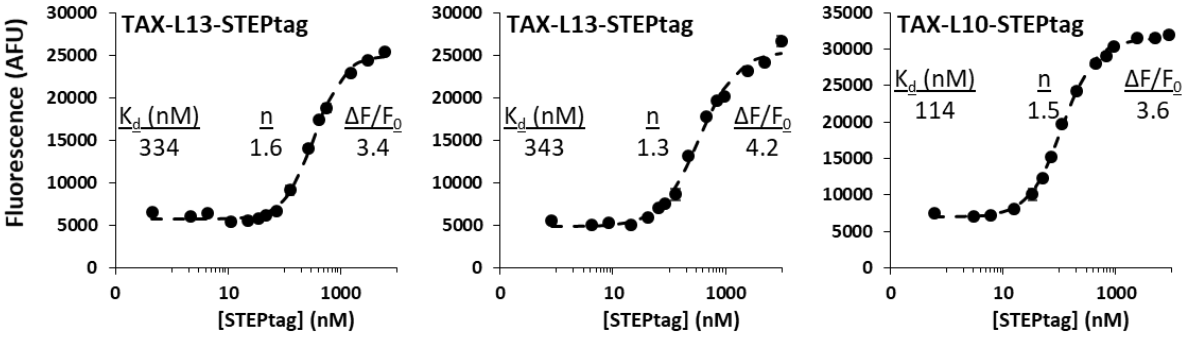


Figure S3.6 Binding curves of gSTEP1 with TAX-STEPtag fusions. Scans of 75 nM of biological replicates of purified gSTEP1, mixed with varying concentrations of TAX-STEPtag fusions, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). The fluorescence signal is the emission at 515 nm, when exciting at 485 nm. The dashed lines are the results of a fit of the Hill equation to the data, and K_d, $\Delta F/F_0$ and the value of n calculated from the Hill equation fit are reported for each curve. Error bars are the standard deviation of triplicate measurements. Note that the behaviour of the fusions is similar to that seen with gSTEP0-T1, confirming that gSTEP1 binds similarly, and should likewise perform well with a 10 amino-acid linker between STEPtag and a POI.

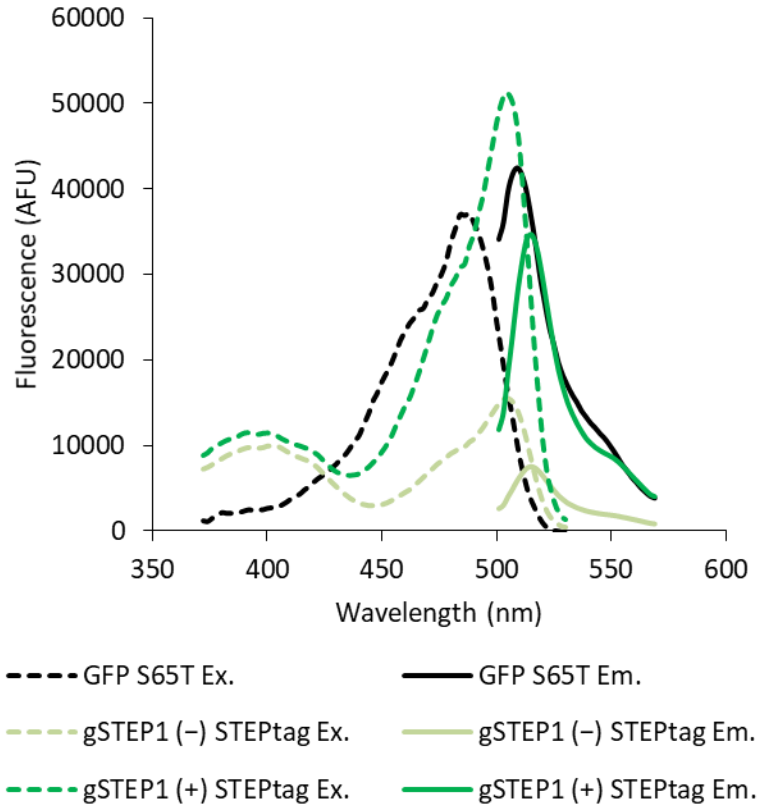


Figure S3.7 Fluorescence spectra of gSTEP1 compared to GFP S65T. Excitation and Emission spectra of purified GFP S65T (75 nM) or gSTEP1 (75 nM), alone and in the presence of 10 µM STEPtag. The excitation spectrum was gathered with an emission wavelength of 550 nm, and the emission spectra was gathered with an excitation wavelength of 485 nm. The GFP S65T spectra are from a single purification, while the gSTEP1 spectra are the average of six purifications. All measurements were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4), at room temperature.

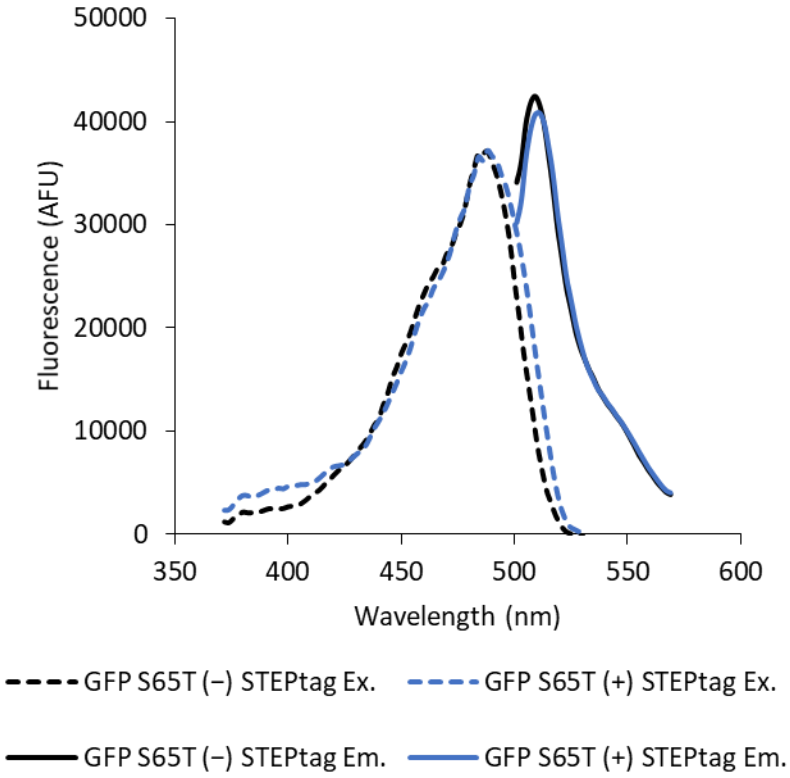


Figure S3.8 STEPtag does not affect the fluorescence spectra of GFP S65T. Excitation and Emission spectra of purified GFP S65T (75 nM) alone and in the presence of 10 µM STEPtag. The excitation spectrum was gathered with an emission wavelength of 550 nm, and the emission spectra was gathered with an excitation wavelength of 485 nm. These spectra are from a single purification. All measurements were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4), at room temperature.

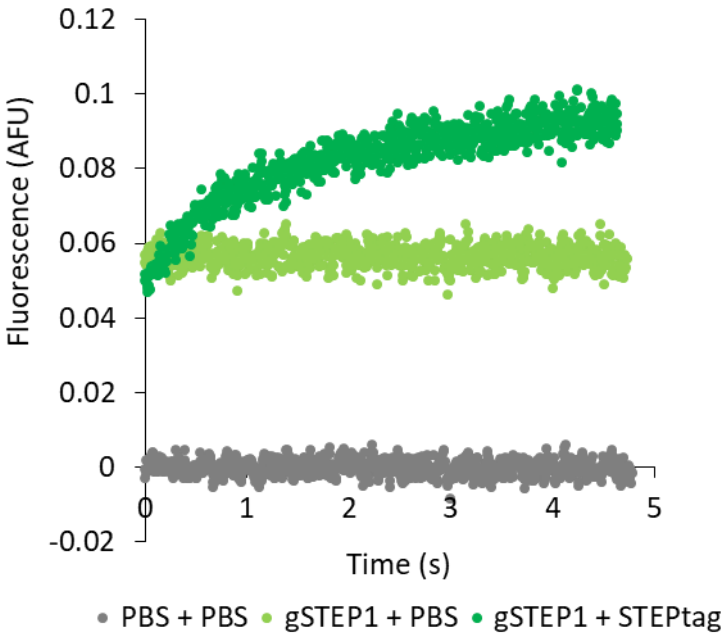
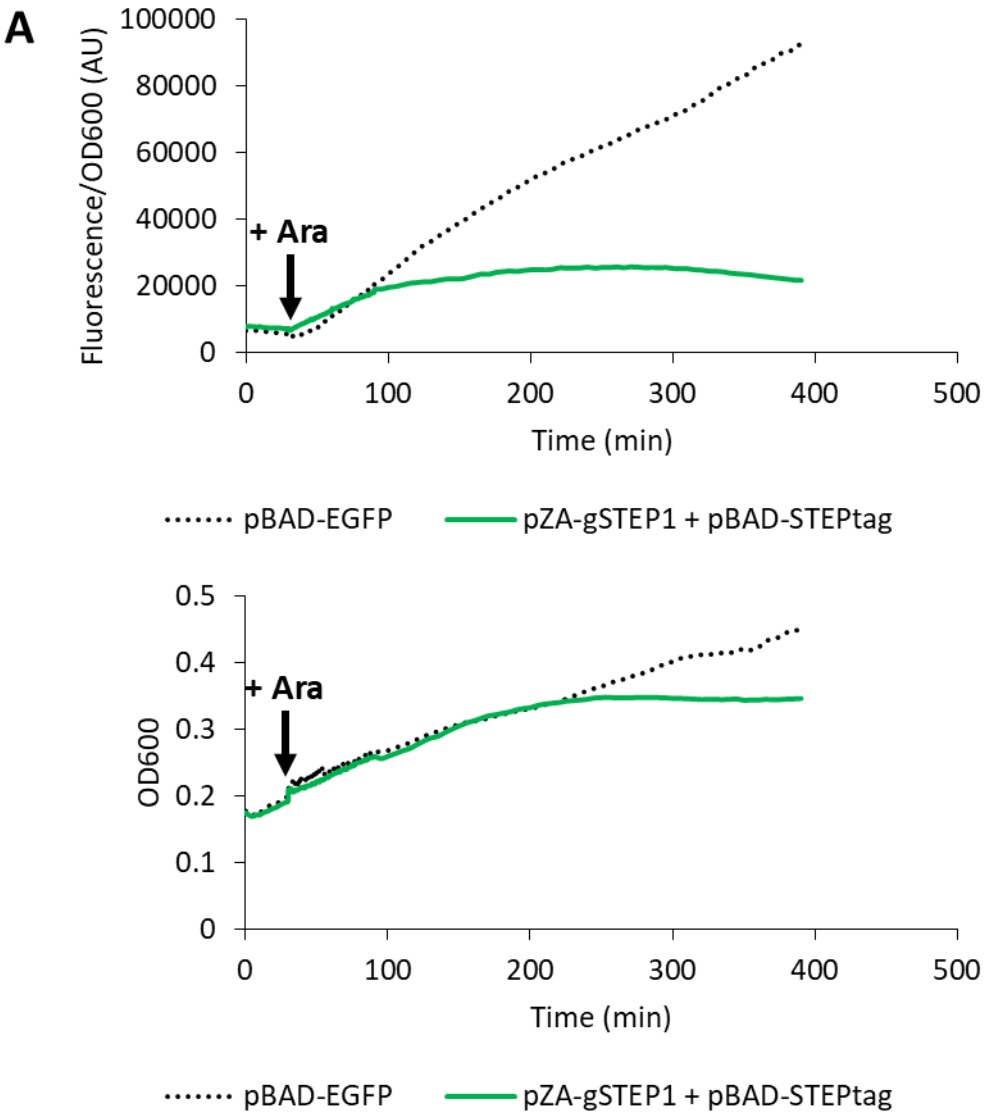


Figure S3.9 Raw signal from the stopped-flow experiment. Measured signal from the stopped-flow PMT when filtering for excitation at 485 nm, emission at 515 nm. Injection of equal volumes of samples into the mixing chamber occurs at time = 0 s. Samples are named as the contents of syringe 1 + syringe 2, with PBS being 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). gSTEP1 was mixed at an initial concentration of 1 µM, and STEPtag at an initial concentration of 5µM. All measurements were performed at 37°C.


A
Fluorescence/OD600 (AU)
100000
80000
60000
40000
20000
0
+ Ara
0
100
200
300
400
500
Time (min)
pBAD-EGFP
pZA-gSTEP1 + pBAD-STEPtag
OD600
0.5
0.4
0.3
0.2
0.1
0
+ Ara
0
100
200
300
400
500
Time (min)
pBAD-EGFP
pZA-gSTEP1 + pBAD-STEPtag

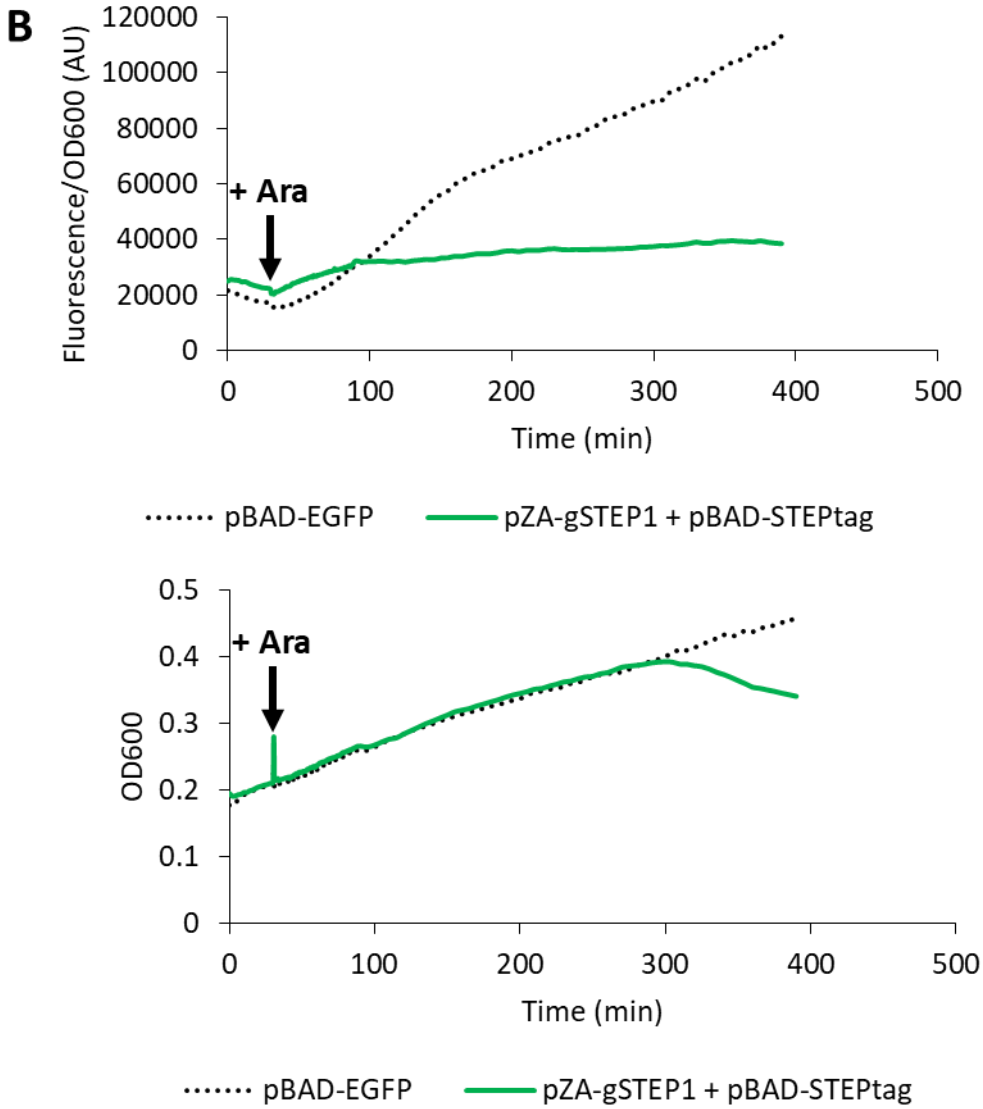


Figure S3.10 Full time courses of cellular fluorescence during reporter protein induction. A, The time course used for Figure 3.10 Peak fluorescence measurements when exciting at 488 nm, normalized to the OD 600 of the sample to account for variations in cell density, as well as OD 600 measurements, over the full 390 minute time course. Cells were cultured in LB broth to mid-log phase, then transferred to a 96-well plate in a plate reader, incubating at 37°C. For the first 30 minutes, measurements were taken every 2 minutes, shaking prior to measurement. At 30 minutes, 0.23% arabinose was added to induce pBAD expression (+Ara), and measurements were taken every 23 seconds afterwards for 1 hour, shaking every 15 minutes. For the final 5 hours, measurements were taken every 2 minutes, shaking prior to measurement. **B,** A replicate assay, run under the same conditions as in Panel A.

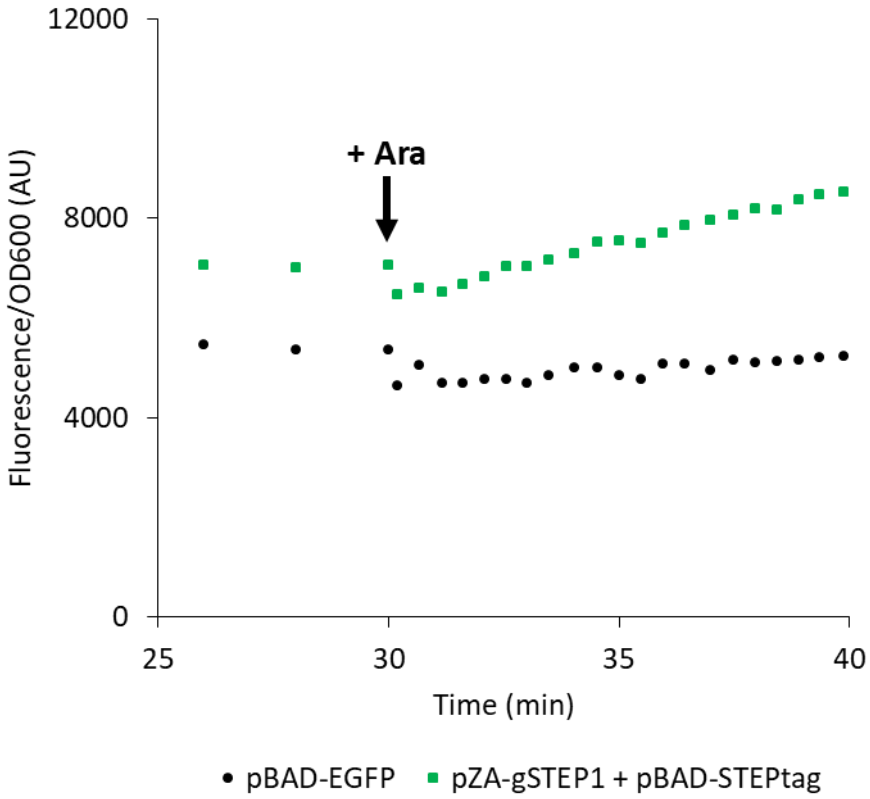


Figure S3.11 Cellular fluorescence change at the moment of reporter protein induction. Peak fluorescence measured when exciting at 488 nm, normalized to the OD 600 of the sample to account for variations in cell density, as well as OD 600 measurements, over the full 390 minute time course. The selected data is the same as was used for Figure 3.10, in this figure the individual data points are shown to highlight the effect of induction on the signals. Cells were cultured in LB broth to mid-log phase, then transferred to a 96-well plate in a plate reader, incubating at 37°C. For the first 30 minutes, measurements were taken every 2 minutes, shaking prior to measurement. At 30 minutes, 0.23% arabinose was added to induce pBAD expression (+Ara), and measurements were taken every 23 seconds afterwards for 1 hour, shaking every 15 minutes.

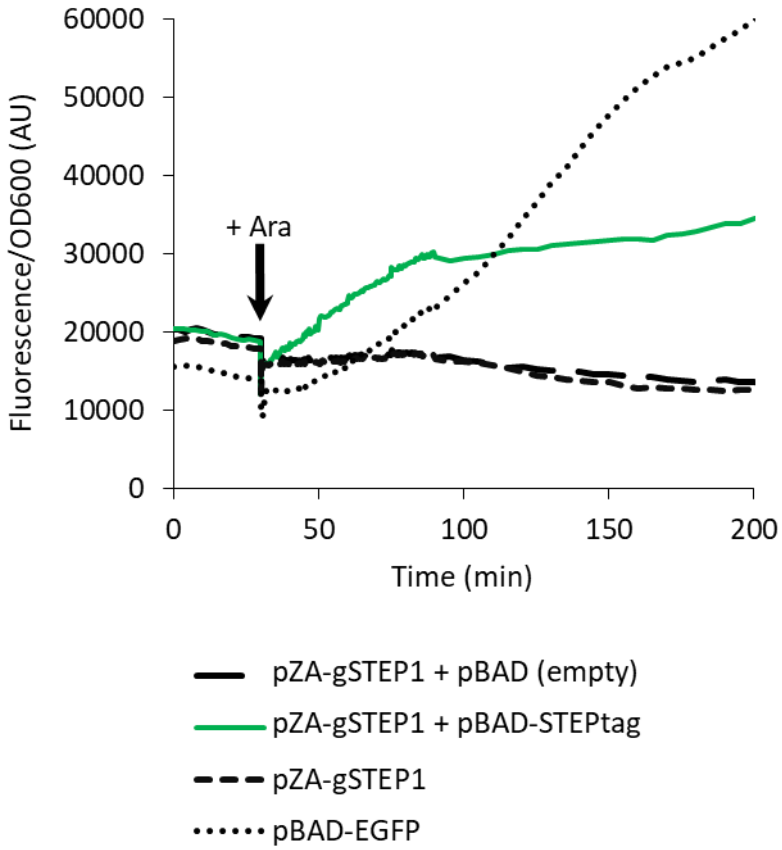


Figure S3.12 pZA-gSTEP1 expression is not induced by arabinose. Peak fluorescence measured when exciting at 488 nm, normalized to the OD 600 of the sample to account for variations in cell density, as well as OD 600 measurements. The data for pZA-gSTEP1 + pBAD-STEPtag and pBAD-EGFP is the same as was used for Figure 3.10, also included are negative controls expressing gSTEP1 without STEPtag. Cells were cultured in LB broth to mid-log phase, then transferred to a 96-well plate in a plate reader, incubating at 37°C. For the first 30 minutes, measurements were taken every 2 minutes, shaking prior to measurement. At 30 minutes, 0.23% arabinose was added to induce pBAD expression (+Ara), and measurements were taken every 23 seconds afterwards for 1 hour, shaking every 15 minutes.

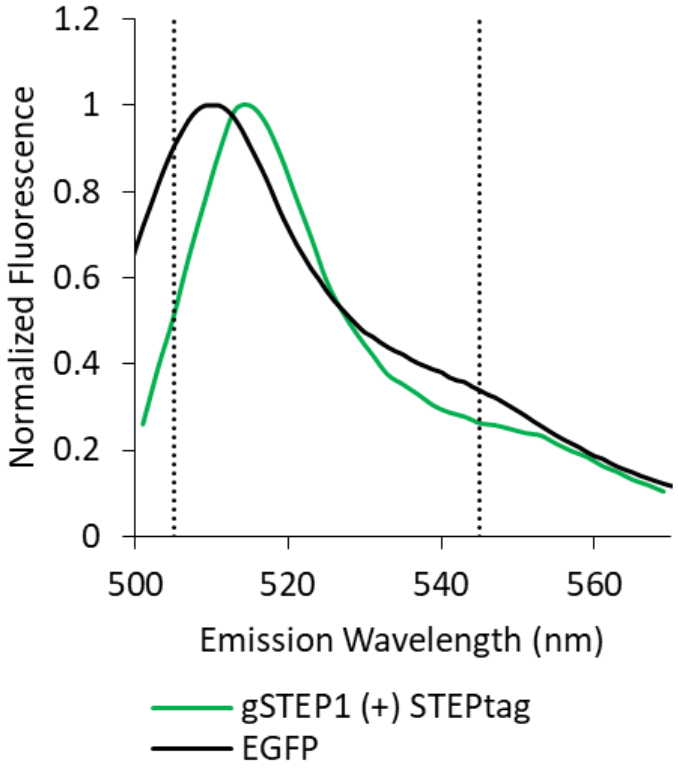
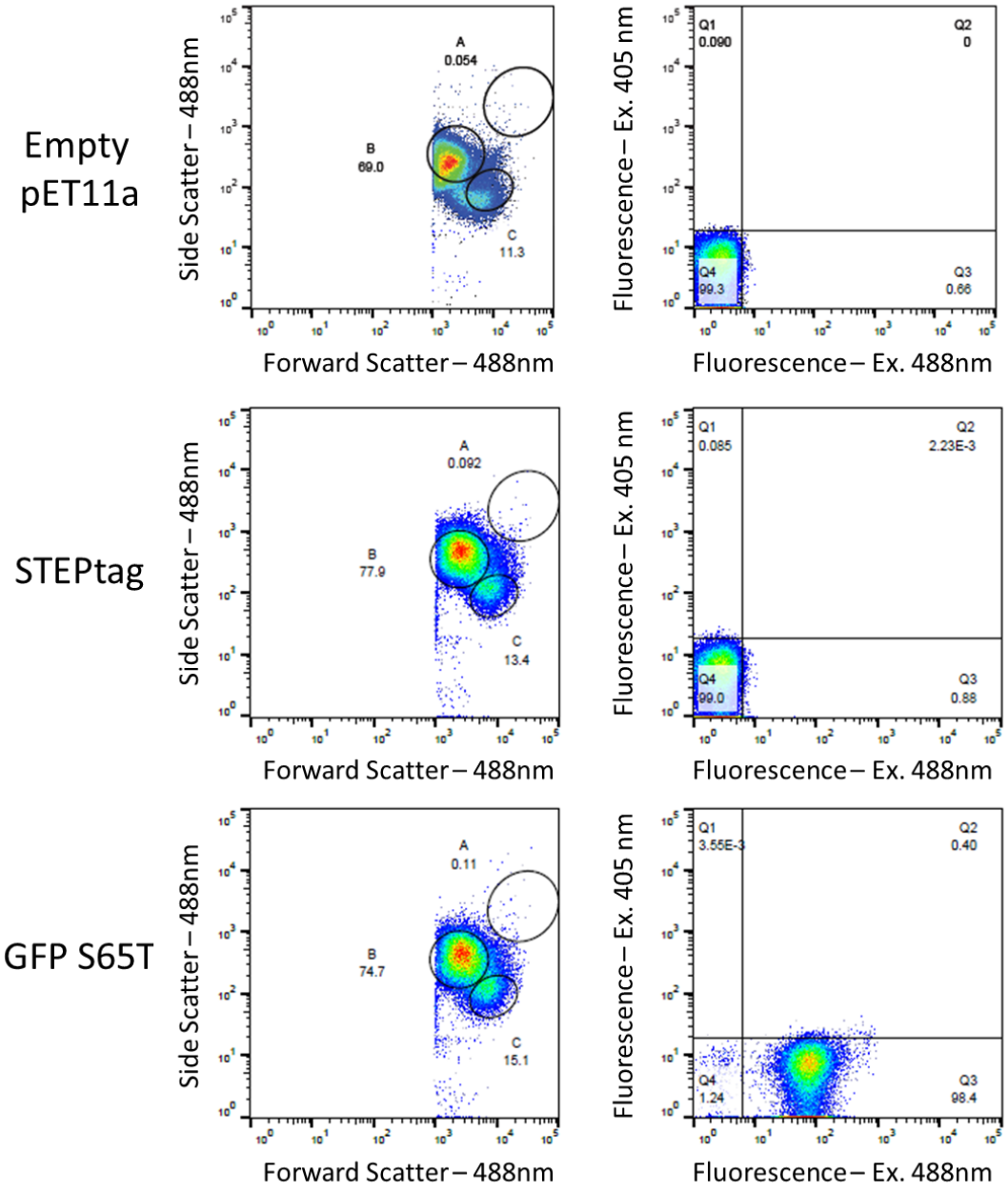


Figure S3.13 gSTEP1 and EGFP emission when passed through the flow cytometry filter. Normalized emission spectra for gSTEP1 saturated with STEPtag and for EGFP. The EGFP spectra was obtained from FPbase.[1] The vertical dashed lines indicate the 505 and 545 nm wavelengths, between which the 525BP40 filter used for flow cytometry would allow fluorescence to be detected.

3.6.4 Additional Flow Cytometry Data



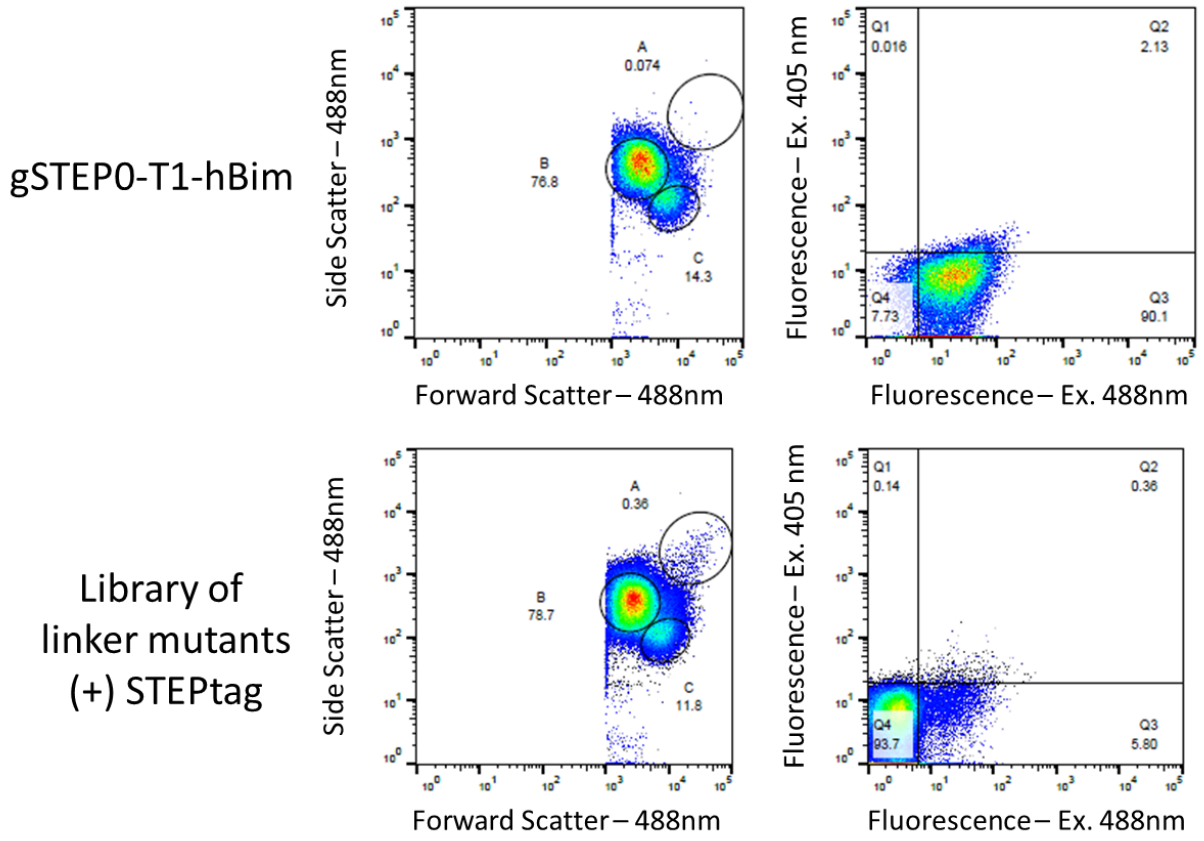


Figure S3.14 Flow cytometry controls before the positive FACS sort of linker mutants. All cells sorted are *E. coli* BL21 (DE3) Gold. Dot plots of forward scatter height vs side scatter height (Left) were used to select for well-formed cells (Gate B). The gated cells were then tested for green fluorescence when exciting at both 405 nm and 488 nm (right).

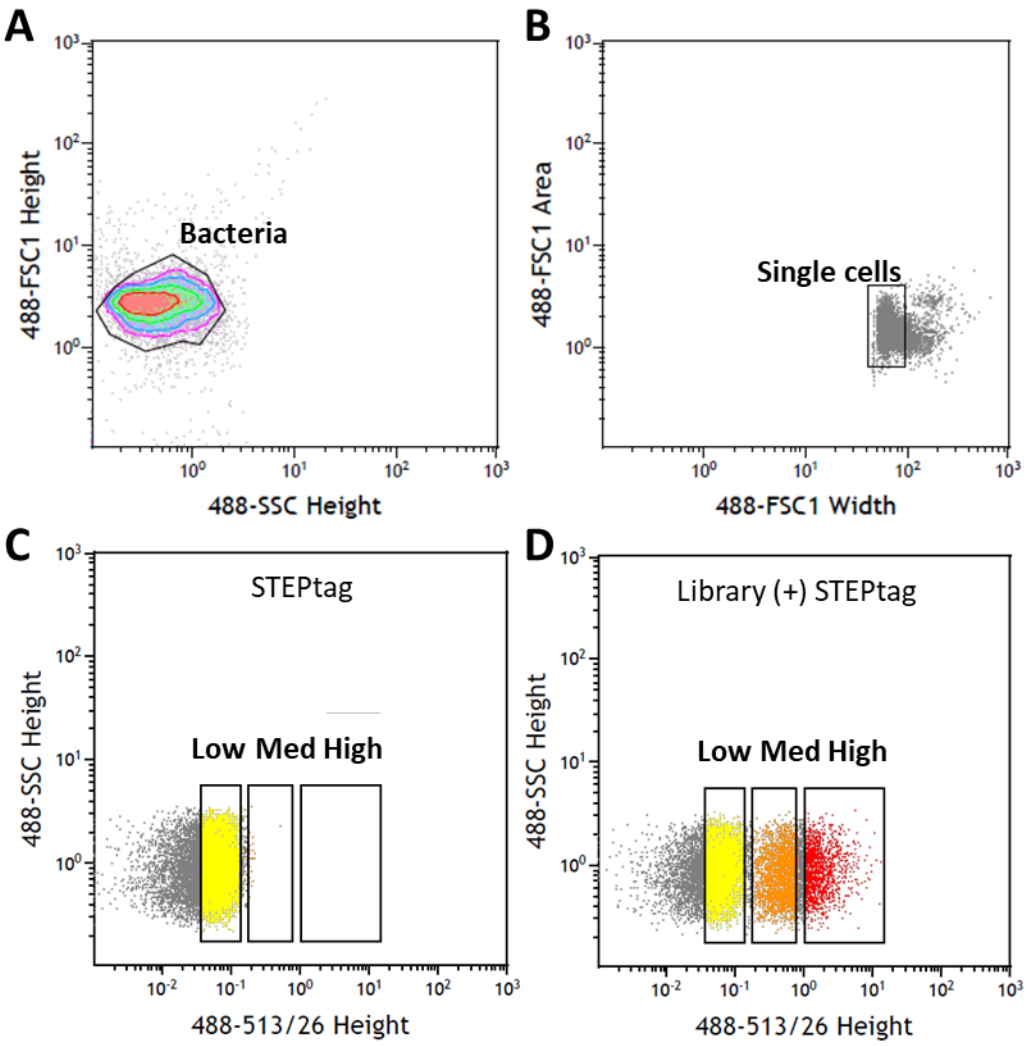


Figure S3.15 Flow cytometry controls for the negative FACS sort of linker mutants. All cells sorted are *E. coli* BL21 (DE3) Gold. **A,** The dot plot of side scatter (SSC) versus forward scatter (FSC1) of cells containing empty pET11a vector, used to select for well-formed cells. The Bacteria gate was applied to all populations during the fluorescence measurements. **B,** The dot plot of forward scatter width versus area of cells containing empty pET11a vector, used for doublet discrimination. The Single cells gate was also applied to all sorted populations during the fluorescence measurements. **C,** The dot plot of green fluorescence when exciting at 488 nm, using a 513/26 bandpass filter, versus side scatter of cells expressing STEPtag from a pET11a vector. This population was used to establish the Low gate for the final sort. **D,** The same dot plot as panel C, but for cells expressing the positively sorted library and STEPtag from a pET Duet vector. This population was used to establish the Med and High gates for the final sort.

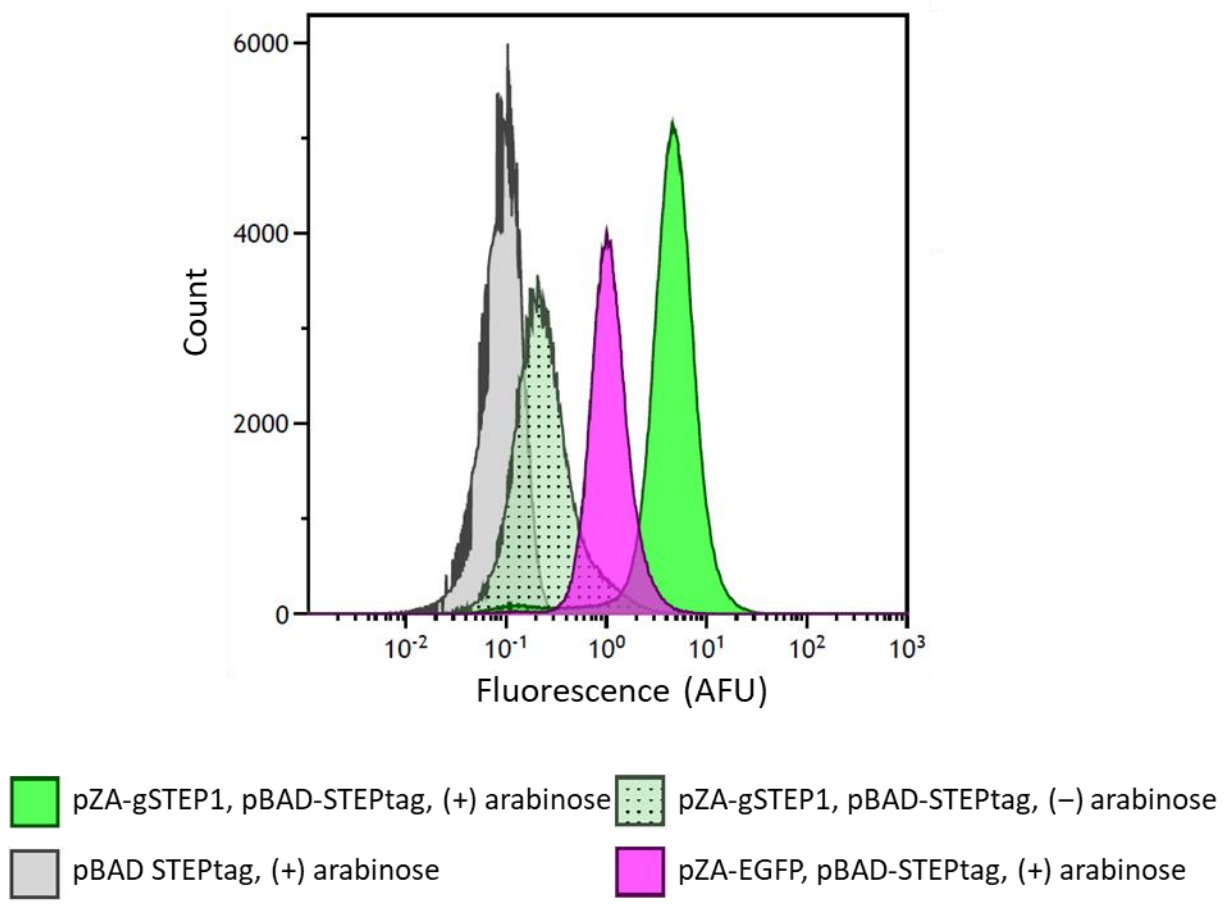
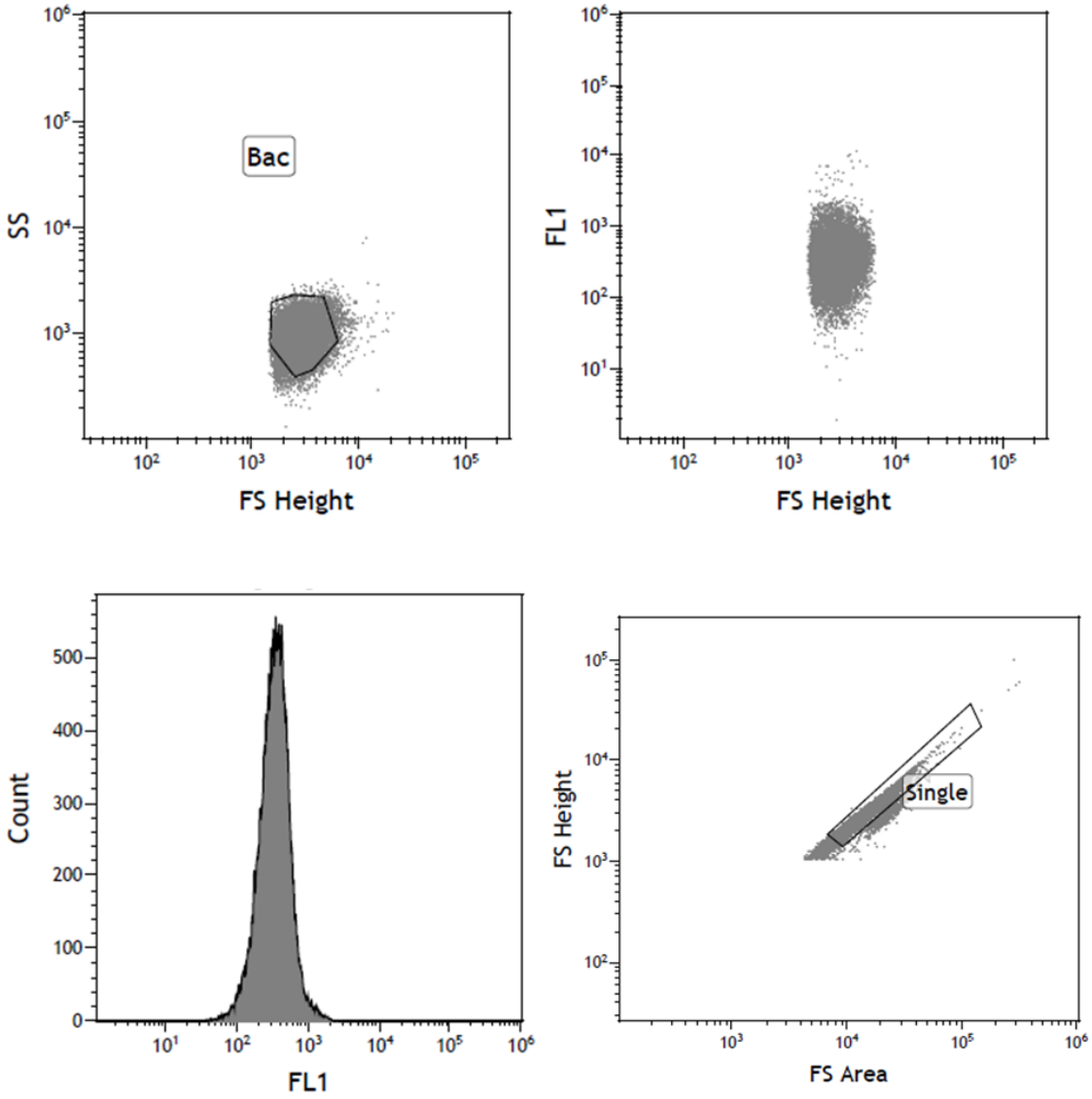
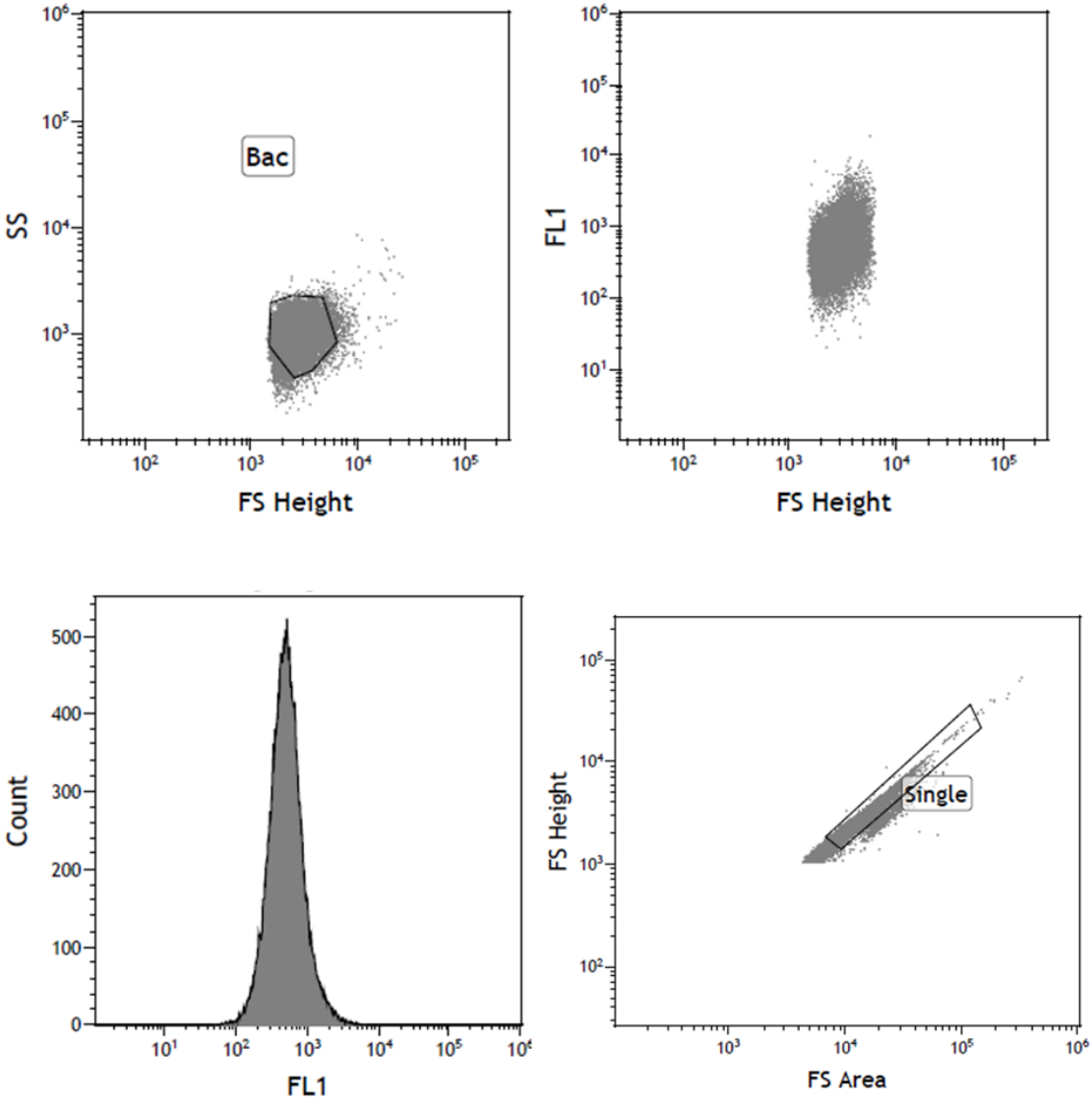


Figure S3.16 Bacterial flow cytometry of cells expressing gSTEP1. Replicate of the experiment performed in Figure 3.9. Flow cytometry histogram of STEP fluorescence in three cultures of *E. coli* TOP10 cells, containing plasmids expressing either STEPtag alone, gSTEP1 and STEPtag, or EGFP and STEPtag. After growth to mid-log phase, the cells containing both gSTEP1 and STEPtag were split into two cultures, and arabinose was added to the indicated samples to induce STEPtag expression, one hour before the cells were harvested for cytometry. Fluorescence was excited at 488 nm, and emission was measured using a 525BP40 filter.

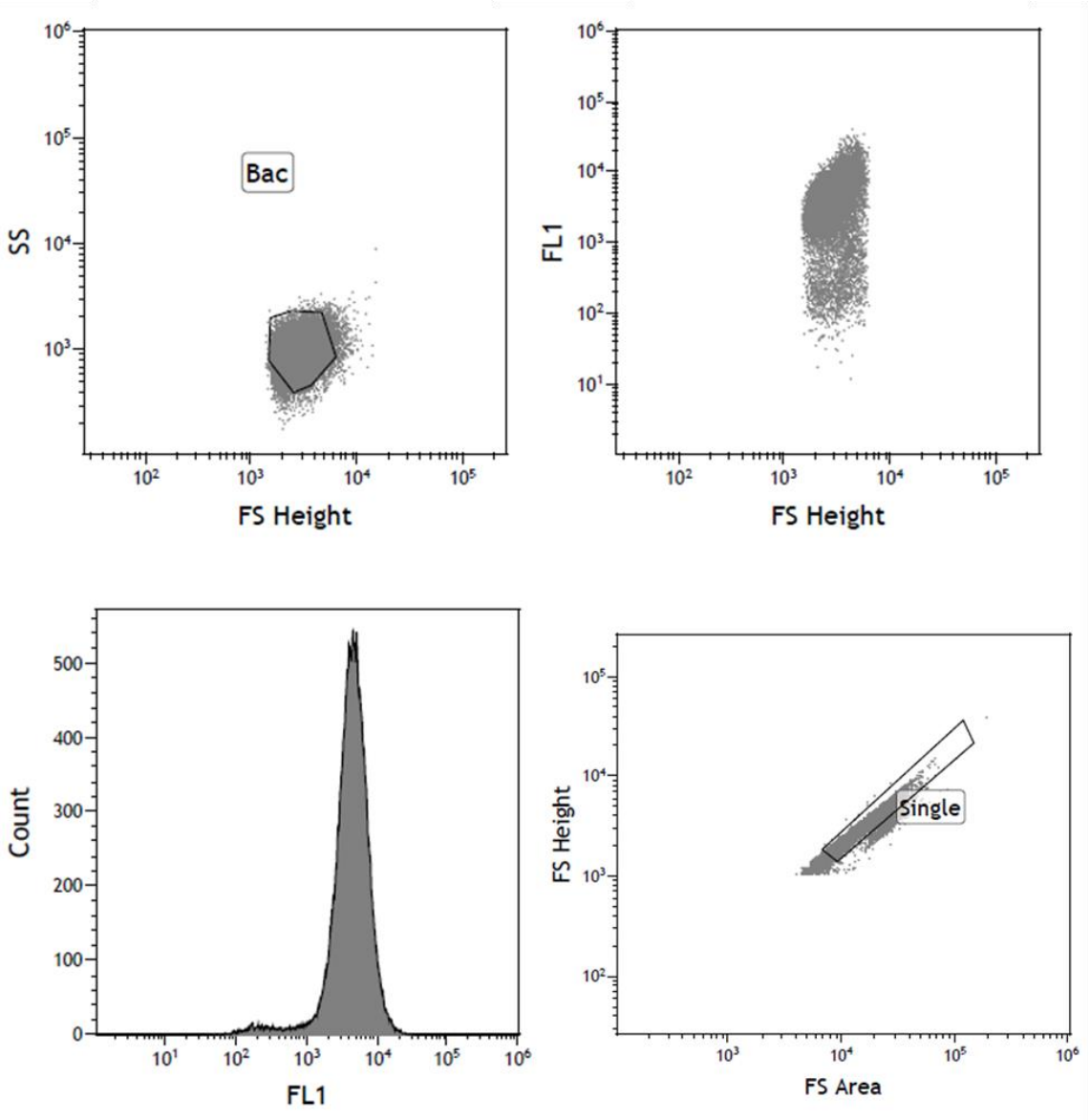
Top10, pBAD STEPtag, (+) Arabinose



Top10, pZA gSTEP1, pBAD STEPtag, (−) Arabinose



Top10, pZA gSTEP1, pBAD STEPtag, (+) Arabinose



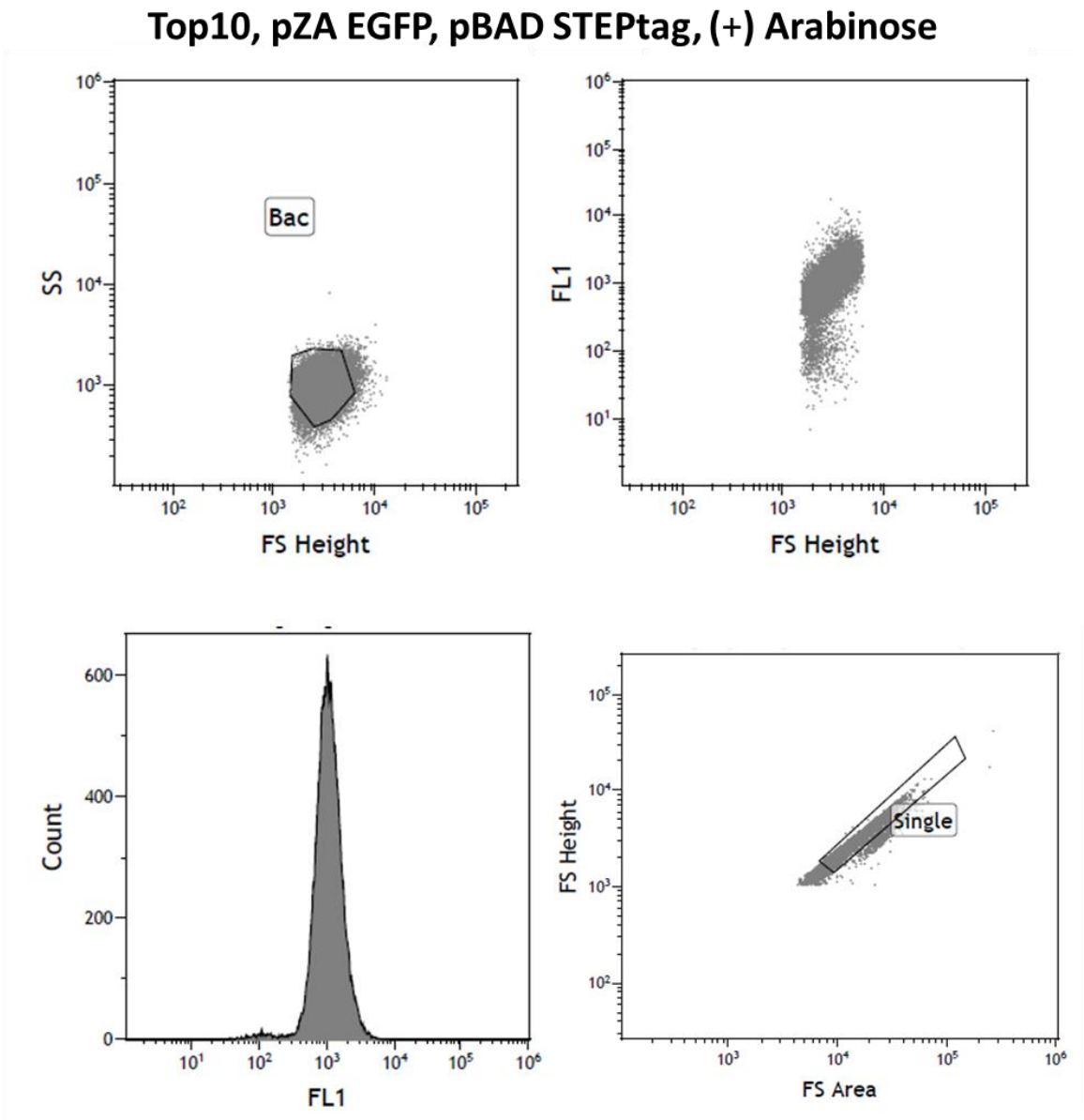
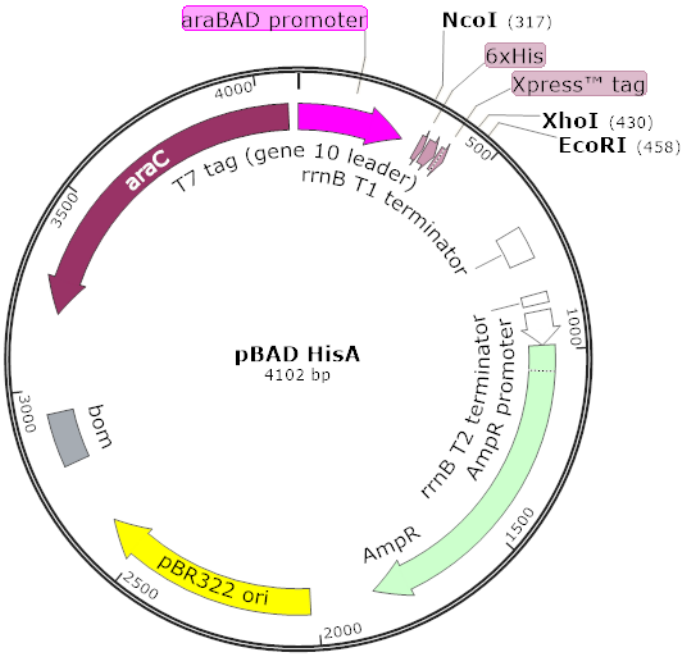
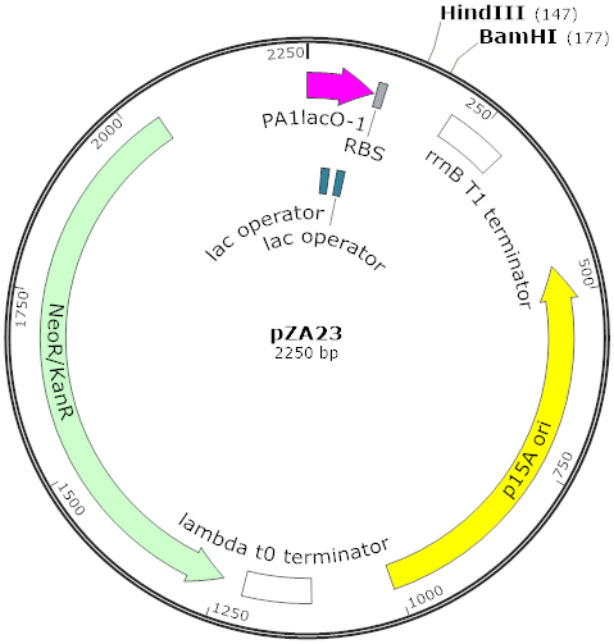


Figure S3.17 Flow cytometry data. Cell strain, plasmids and presence of arabinose are all indicated in the figure titles. Dot plots of forward scatter (FS) area versus height for doublet discrimination, forward scatter height versus side scatter height (SS), for selection of well-formed cells, and forward scatter versus green fluorescence (FL1) for the flow cytometry runs shown in Figure 3.9. Events with a large forward scatter area to forward scatter height ratio are more likely to be doublets, i.e. multiple cells passing through the detector simultaneously, so they were removed with the "Single" gate, which also removes the population of small, nonfluorescent particles, which we hypothesize to be dust. The "Bac" gate from the forward scatter versus side scatter plot was used for the fluorescence histogram, to screen out outliers with anomalous shapes or sizes.

3.6.5 Vector Maps





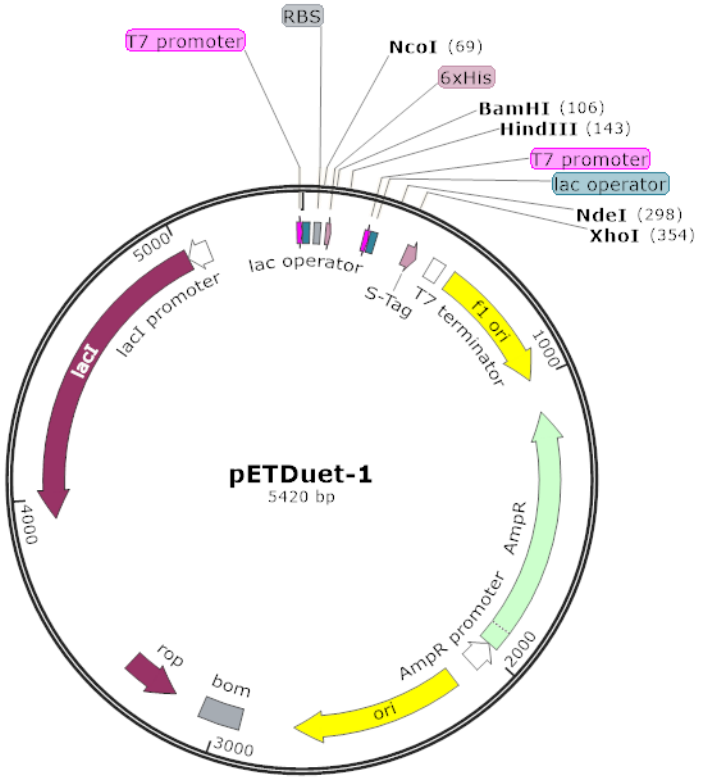


Figure S3.18 Vector maps of new plasmids used in this chapter. Vector maps were generated using SnapGene Viewer. Promoters used are highlighted in magenta, and restriction sites used for cloning are indicated in bold. Note that pBAD HisA and pZA23 have different antibiotic resistances (Ampicillin and Kanamycin), as well as compatible origins of replication (pBR322 and p15A), such that they can both be stably maintained within the same cells.

Chapter 4: Side STEPs: Expansion of the STEP family

Matthew G. Eason, Antonia T. Pandelieva, Eduardo Ramirez & Roberto A. Chica

Manuscript in preparation

4.1 Statement of Contribution

Matthew G. Eason designed all protein sequences. Matthew G. Eason and Eduardo Ramirez performed cloning experiments. Matthew G. Eason and Antonia T. Pandelieva performed characterization experiments and analyzed data. Matthew G. Eason and Roberto A. Chica selected binding partners to test. Matthew G. Eason wrote the chapter.

4.2 Introduction

While the green version of the STEP was being created and optimized, we also considered other permutations of the sensor. As the STEP is made up of defined motifs: a binding protein (STEPtag), binding peptide (Bim) and a circularly permuted fluorescent protein (cpGFP), we considered different designs that varied one or more of these motifs to change the function or fluorescence characteristics of the STEP. Ideally, this could lead to a toolbox of STEP designs, where different binding peptide-protein pairs could be used to alter the binding kinetics or different fluorescent proteins could be used to change the colour of the signal produced. Combining these designs would then lead to multiplexing of the STEP, by using different coloured fluorescent proteins and orthogonal binding peptide-protein pairs, such that each STEP detects a different protein during the same experiment.[227]

When considering alternative binding peptide-protein pairs, we turned to the original proteins on which we based the design of the STEP: the family of genetically encoded calcium ion indicators, in particular GCaMP3. These proteins already contain such a binding pair, although they are fused together next to the cpGFP pore. In the case of GCaMP3, the peptide is a 13-residue fragment of myosin light chain kinase (M13),[228] the protein is calmodulin,[229] and together they form a binding pair

that has been shown to bind very tightly, with a K_d on the sub-nanomolar scale when calcium ions are present.[230] Thus, we hypothesized that by splitting GCaMP3 at the cpGFP-calmodulin linker, the resulting M13-cpGFP construct would be a sensor for calmodulin, much like the STEP is a sensor for STEPtag, but this new construct would only function in the presence of calcium ions, ideally showing a similar change in fluorescence to that of GCaMP3, i.e. a $\Delta F/F_0$ of 12.[86] A proof-of-concept of a similar sensor using a yellow fluorescent protein-based sensor, split-pericam, has been shown to have partial function as a calcium ion sensor when both halves are co-expressed, although it was split at the linker between the two circularly permuted halves of the fluorescent protein, and thus likely did not mature until after association.[231] A split-GCaMP construct could then open the door to creating conditional sensors, which will only detect tagged proteins when other physiological conditions are met. Alternatively, any protein that binds specifically to a peptide could potentially be used to create a STEP, for instance PDZ and SH3 domains, which have previously been used to create molecular switches with a range of K_d values.[110] For the purposes of multiplexing, two sets of peptide-protein pairs are required that are orthogonal, i.e. each pair will not interact with the members of the other pair. Conveniently, the Bcl-2 family used to create gSTEP1 and STEPtag is comprised of multiple homologs with different specificities,[119] such that orthogonal pairs can be identified, and we have also tested some of these pairs.

In addition to orthogonal binding pairs, multiplexing the STEP to detect multiple proteins in the same experiment would also require spectrally distinct fluorophores, such that the signal from each POI is different. Having access to multiple colours of STEP is also a relevant stand-alone goal, as it would allow the sensor to be used with a wider range of optical hardware, or simultaneously with other fluorophores in imaging experiments. To achieve this goal, we began to create a toolbox of different colours of STEP, by introducing mutations to the chromophore and surrounding residues that have previously been shown to affect the fluorescence spectra. For instance, the T203Y mutation in GFP

has been shown to red-shift its fluorescence by π-stacking with the chromophore, creating a yellow fluorescent protein.[22] An opposite effect can be seen in the cyan fluorescent proteins, where the Y66H or Y66W mutation directly alters the aromatic side chain present in the chromophore, blue-shifting the spectra.[25] Finally, although red-shifting GFP to higher wavelengths than YFP proved difficult, red fluorescent proteins have been derived from homologous proteins in species of coral and anemones,[33, 34, 50] with extended chromophore conjugation along the protein backbone. By introducing the equivalent YFP and CFP mutations into gSTEP1, or by replacing the cpGFP with a circularly permuted RFP sequence, we created versions of the STEP with a range of different colours.

4.3 Results

4.3.1 Alternative Binding Partners

The first set of alternative binding partners that we tested was the M13 peptide and calmodulin pair from GCaMP3. To create the biomolecular system, the GCaMP3 sequence was split at the linker between the cpGFP and calmodulin, resulting in an M13-cpGFP fluorescent protein, and a calmodulin binding partner. When these two parts were mixed together in the presence of calcium ions, we found that there was indeed an increase in fluorescence (Figure 4.1). Additionally, this interaction was shown to only occur in the presence of calcium ions, as when EGTA was added to chelate calcium ions, the signal remained at baseline levels. The measured binding affinity was much weaker than the reported sub-nanomolar value for free M13 and calmodulin, as we found that the signal was not saturated at 600 nM of calmodulin, and did not noticeably increase above baseline until more than 20 nM of calmodulin was added. As the signal could not be saturated with the amount of purified calmodulin available, we were unable to determine a K_d or a $\Delta F/F_0$, but some lower bounds can be placed on these values. Given the results obtained, the $\Delta F/F_0$ is at least 2.0 ± 0.7, with a K_d of at least 120 ± 50 nM, based on two biological replicates. We also verified that the signal increase was specific to calmodulin binding, by performing binding curves with BSA in parallel (Figure S4.6), and found that the addition

of BSA did not lead to an increase in fluorescence, regardless of the calcium ion concentration. Thus, the split-GCaMP system does demonstrate specific, measurable binding, with the possibility of multi-parameter detection, i.e. detecting a POI only when there are also calcium ions present. As we expected that the added complexity of requiring calcium ions to be added would not generally be an asset to the sensor, we did not continue with further characterization or optimization of this binding pair.

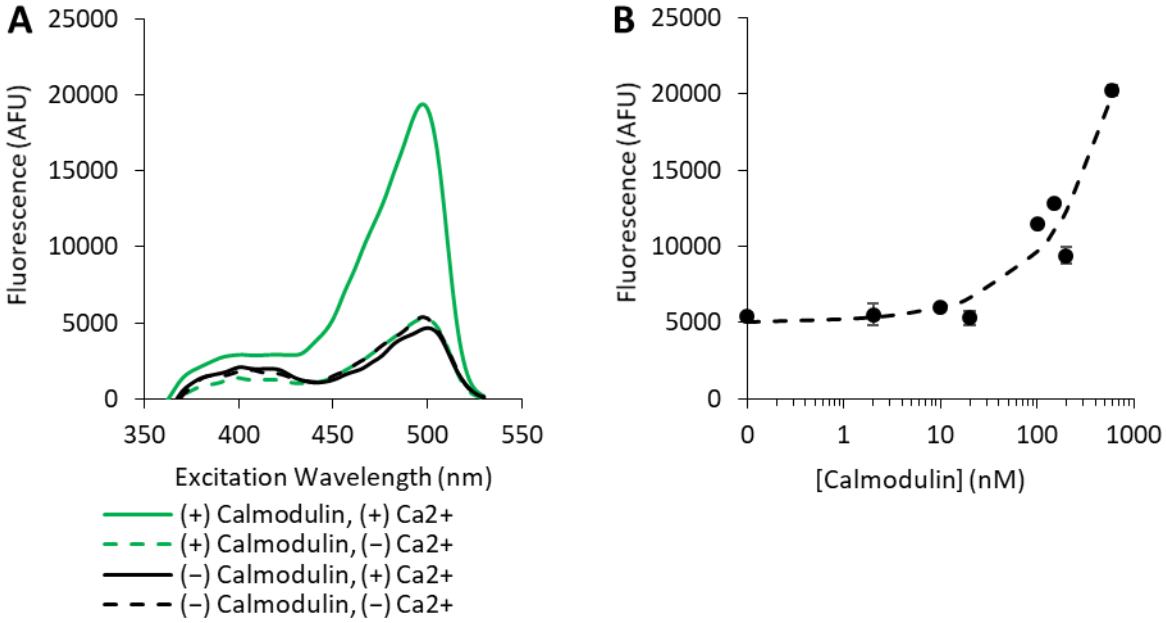


Figure 4.1 Fluorescence spectra and binding curve of M13-cpGFP. A, The measured green excitation spectrum (emitting at 550 nm) of purified M13-cpGFP (75 nM). The (+) Calmodulin curves have 600 nM of calmodulin added, the (+) Ca2+ curves have 1 mM of $CaCl_2$ added, and the (**−**) Ca2+ curves have 1 mM of EGTA added. **B,** A representative binding curve, measuring maximum emission at 550 nm when exciting at 500 nm. The dashed line indicates a fit of the Hill equation to the data (n = 0.6). Error bars are the standard deviation of duplicate measurements.

For our next set of alternative binding pairs, we turned to a PDZ domain from mouse α-1-syntrophin, that had been previously used in the development of a modular protein switch.[110] This PDZ domain, which we called PDZtag, binds to a short peptide sequence, with a measured K_d of 8 μM. By fusing this short peptide to either the N- or C-terminus of the cpGFP sequence present in gSTEP0-T1, we created PDZ-cpGFP and cpGFP-PDZ, respectively. Both of these constructs were shown to increase in fluorescence in the presence of PDZtag, with a $\Delta F/F_0$ of 1.7 for PDZ-cpGFP and 1.0 for cpGFP-PDZ (Figure 4.2). Interestingly, the measured K_d of binding was lower than that reported by

Dueber *et al.*, at 630 nM for PDZ-cpGFP and 380 nM for cpGFP-PDZ (Figure 4.3). Thus, the PDZ binding pair seems to be a possible pair for future use in a STEP that has different specificity to gSTEP1. Although neither the binding affinity nor the signal increase outperforms gSTEP1, much less optimization has been performed with this binding pair, and therefore the current values could likely be improved.

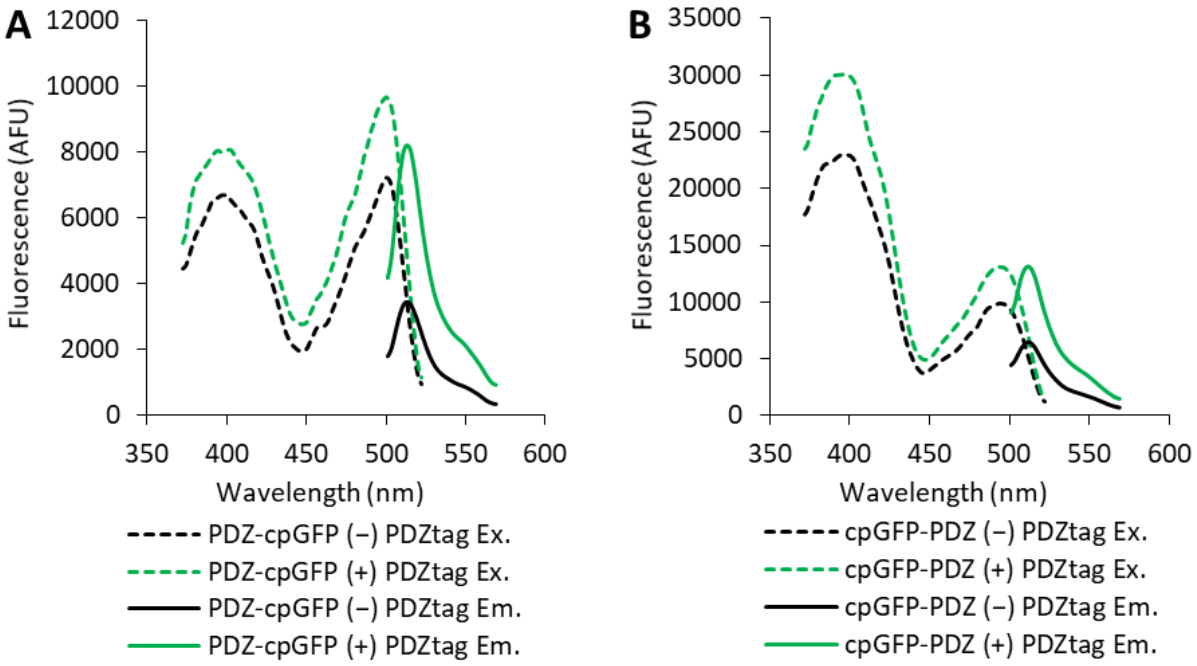


Figure 4.2 Fluorescence spectra of PDZ-based binding pairs. A, The measured green excitation spectrum (emitting at 550 nm) and green emission spectrum (exciting at 485 nm) of purified PDZ-cpGFP (75 nM). The (+) PDZtag curves have 5 µM of PDZtag added. **B,** The measured green excitation spectrum (emitting at 550 nm) and green emission spectrum (exciting at 485 nm) of purified cpGFP-PDZ (75 nM). The (+) PDZtag curves have 5 µM of PDZtag added.

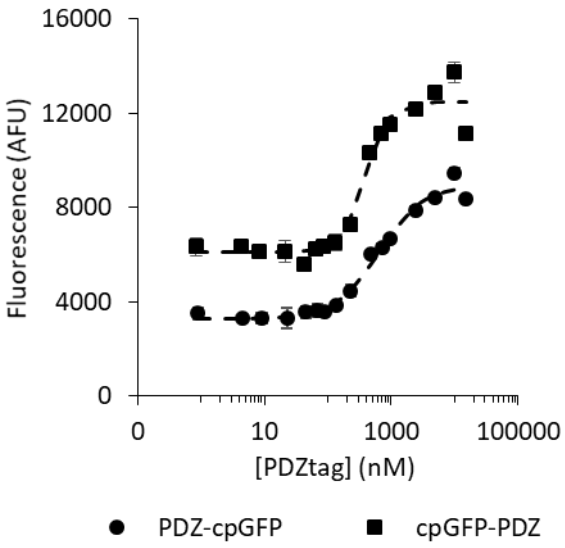


Figure 4.3 Binding curves of PDZ-based binding pairs. Measured binding curves for PDZ-cpGFP and cpGFP-PDZ, measuring emission at 515 nm when exciting at 485 nm. The dashed line indicates a fit of the Hill equation to the data (n = 1.3 for PDZ-cpGFP and n = 2.4 for cpGFP-PDZ). Error bars are the standard deviation of triplicate measurements.

Our final set of alternative binding pairs were the other members of the Bcl-2 family. When comparing the affinities of the different peptides and proteins,[119] it has been noted that the Noxa peptide and Mcl-1 protein bind specifically to one another, as Noxa binds very poorly to other proteins in the family, including Bcl-x_L, which was used as the template for STEPtag. The Bim peptide used in gSTEP1 (Chapter 3) binds to Mcl-1 with high affinity, and thus could not be used as part of the other binding pair. Instead, we chose the Bad peptide, as it does not show measurable binding to Mcl-1, yet maintains similar affinity to Bcl-x_L. Thus, the two predicted orthogonal binding pairs we tested were: pair a, the Bad peptide with Bcl-x_L, and pair b, the Noxa peptide with Mcl-1. To differentiate the gSTEPs and STEPtags by their specificity, the letter associated with the binding pair is appended to the name, thus, gSTEPs with the Bad peptide are known as gSTEP.a, and STEPtags based on Bcl-x_L will be STEPtag.a. Because no modifications were made to the Bcl-x_L based tag, STEPtag.a is identical to STEPtag (Chapters 2 and 3), but will be referred to as STEPtag.a for the rest of the chapter. STEPtag.b is simply Mcl-1, with the C-terminal membrane binding domain removed.

To test binding pair a, with the Bad peptide and Bcl-x_L binding partner, we generated gSTEP0.a, based on gSTEP0, our optimal sensor at the time. To create gSTEP0.a, we replaced the Bim peptide sequence with a Bad peptide sequence, and also included the mutations to the GFP barrel reported in GCaMP8,[87] as they were reported to improve the $\Delta F/F_0$ of the calcium ion sensor, and we hoped to see a similar increase in the STEP. STEPtag.a was simply STEPtag, with no further changes added. When we tested this binding pair, we found that gSTEP0.a did show an increase in signal when STEPtag.a was added, with a K_d of 200 ± 100 nM, and a $\Delta F/F_0$ of 0.8 ± 0.3 (Figure 4.4). This was a reasonable result, with a K_d similar to that of gSTEP0 (240 ± 50 nM), and a small loss in $\Delta F/F_0$ relative to gSTEP0 (1.4 ± 0.1), so the gSTEP0.a sensor does seem to function as desired with STEPtag.a. But when it was tested with STEPtag.b, an Mcl-1 sequence that should not bind to Bad, we also measured an increase in signal, with a K_d of 480 nM, and a $\Delta F/F_0$ of 1.4. Thus, this iteration of the sensor does not seem have the expected specificity for STEPtag.a.

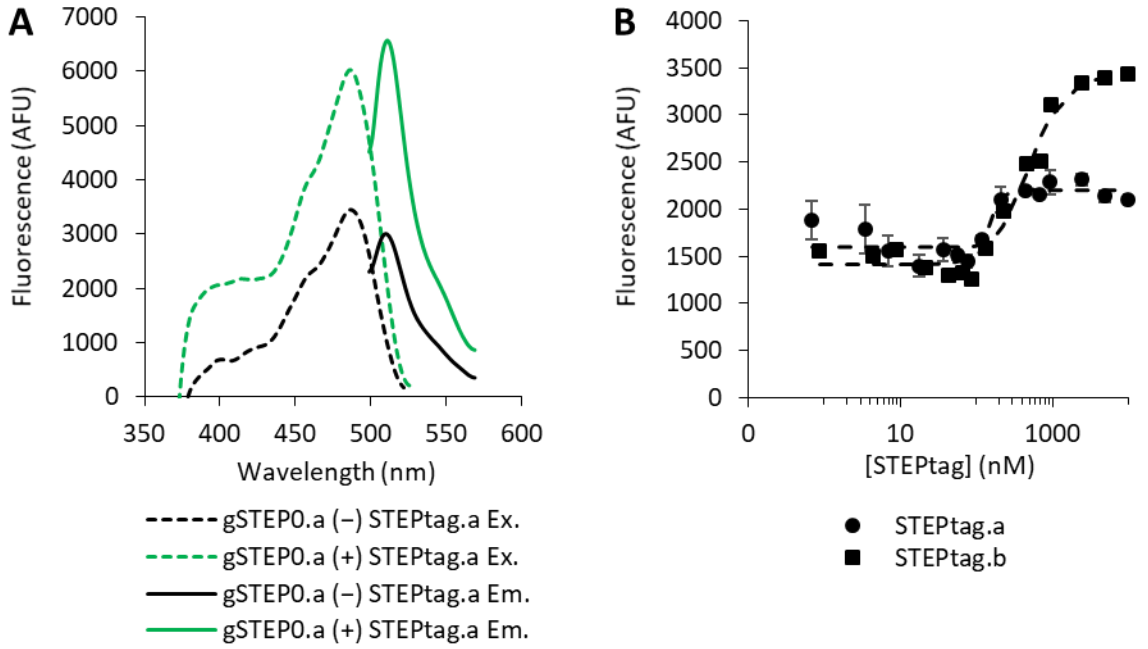


Figure 4.4 Fluorescence spectra and binding curve of gSTEP0.a. **A,** The measured green excitation spectrum (emitting at 550 nm) and green emission spectrum (exciting at 485 nm) of purified gSTEP0.a (75 nM). The (+) STEPtag.a curves have 10 µM of STEPtag.a added. **B,** Representative binding curves, measuring emission at 515 nm when exciting at 485 nm. The dashed line indicates a fit of the Hill equation to the data (n = 6.9 for STEPtag.a, 1.8 for STEPtag.b). Error bars are the standard deviation of triplicate measurements.

To test binding pair b, with the Noxa peptide and Mcl-1 binding partner, we generated gSTEP0.b, this time replacing the Bim peptide with Noxa. We also introduced the C-terminal truncation present in gSTEP0-T1, as we had been simultaneously optimizing the gSTEP0 sensor, and found this truncation to be beneficial (Chapter 2). Although we did not extensively test the effects of the GCaMP8 mutations in gSTEP0.a, we decided not to include them in gSTEP0.b, as they had not initially improved the $\Delta F/F_0$ as we had hoped. As we had had some difficulties expressing STEPtag.b alone, we obtained a maltose-binding protein fusion with Mcl-1, termed MBP-STEPtag.b, with the MBP tag present to increase the yield during purification.[191] Unlike STEPtag.a, where truncations were made to the Bcl-x_L sequence, no further engineering was performed to create STEPtag.b. We found that gSTEP0.b was green fluorescent, although it showed a shift in the peak excitation wavelength to that of the neutral chromophore at 396 nm (Figure 4.4), much like cpGFP-PDZ and gSTEP0-T1-hBim-L4 (Chapter 3). When the anionic chromophore was excited at 485 nm, it showed an increase in fluorescence when MBP-STEPtag.b was added, with a K_d of 38 ± 9 nM and a $\Delta F/F_0$ of 5 ± 2, outperforming gSTEP1 in both affinity and signal increase (Figure 4.5). When STEPtag.a was added to test the specificity of binding, it was found that there is still a measurable response, with a K_d of 500 ± 200 nM, and a $\Delta F/F_0$ of 1.3 ± 0.4. Thus, the binding affinity of gSTEP0.b is not completely orthogonal to STEPtag.a, but it is a full order of magnitude stronger with MBP-STEPtag.b than with STEPtag.a, and shows a considerably larger signal increase.

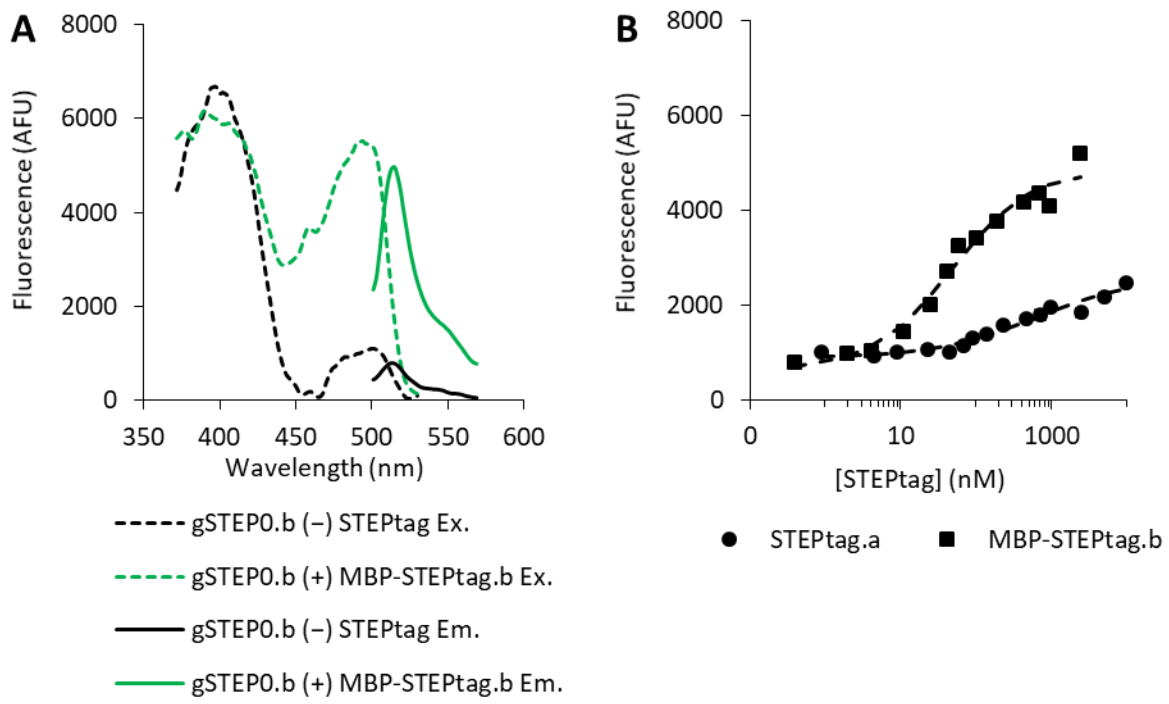


Figure 4.5 Fluorescence spectra and binding curves of gSTEP0.b. **A,** The measured green excitation spectrum (emitting at 550 nm) and green emission spectrum (exciting at 485 nm) of purified gSTEP0.b (75 nM). The (+) MBP-STEPtag.b curves have 2.5 μM of MBP-STEPtag.b added. **B,** A set of representative binding curves, measuring emission at 515 nm when exciting at 485 nm. The dashed line indicates a fit of the Hill equation to the data (n = 0.6 for STEPtag.a, 0.8 for MBP-STEPtag.b). Error bars are the standard deviation of triplicate measurements.

While gSTEP.b was being tested, we verified that the Bim peptide would indeed bind to the MBP-STEPtag.b construct, by performing binding curves with gSTEP0-T1-hBim, our best Bim-based sensor at the time (Chapter 3). We found that there was indeed detectable binding, with a K_d of 90 ± 40 nM and a $\Delta F/F_0$ of 1.25 ± 0.05 (Figure 4.6). This is comparable to the binding with STEPtag.a, with a K_d of 160 ± 40 and a $\Delta F/F_0$ of 3.3 ± 0.1, meaning that the Bim peptide in gSTEP0-T1-hBim could not effectively distinguish between STEPtag.a and MBP-STEPtag.b, and so another binding peptide would be needed if we wished to use both of those binding proteins in the same experiment.

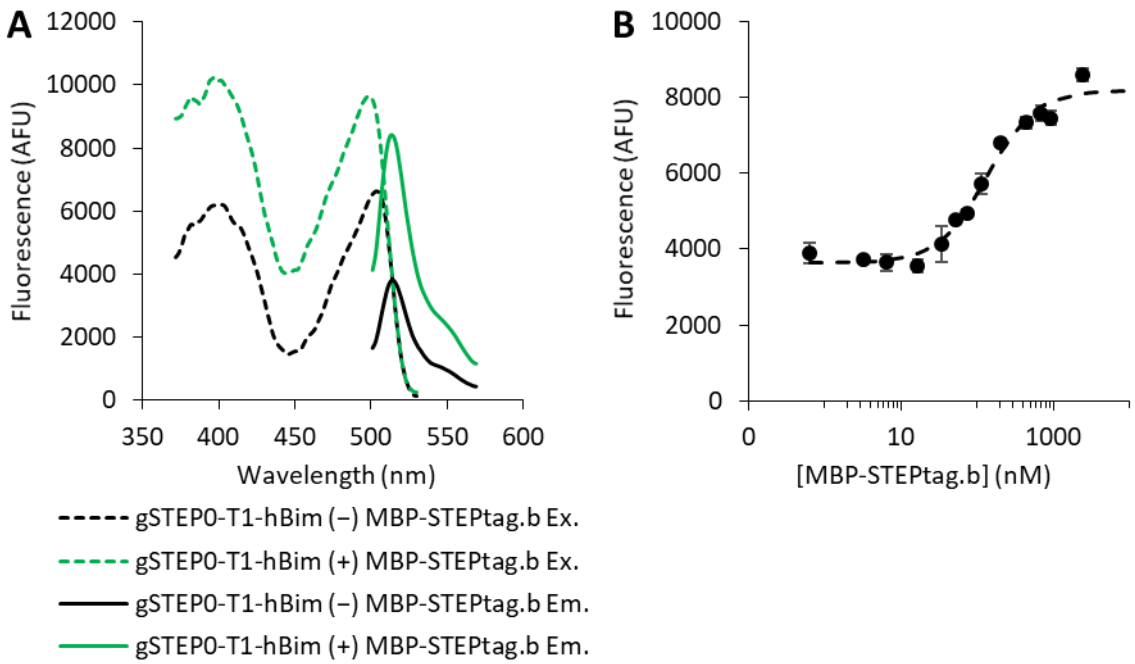


Figure 4.6 Fluorescence spectra and binding curve of gSTEP0-T1-hBim with MBP-STEPtag.b. A, The measured green excitation spectrum (emitting at 550 nm) and green emission spectrum (exciting at 485 nm) of purified gSTEP0-T1-hBim (75 nM). The (+) MBP-STEPtag.b curves have 2.5 µM of MBP-STEPtag.b added. **B,** A representative binding curve, measuring emission at 515 nm when exciting at 485 nm. The dashed line indicates a fit of the Hill equation to the data (n = 1.4). Error bars are the standard deviation of triplicate measurements.

4.3.2 Increasing the Colour Palette of the STEP

With the overarching goal of multiplexing multiple STEPs, we chose to begin expanding the colour palette with a red fluorescent protein, as the red fluorescent proteins have fluorescence spectra that are distinct from those of GFP. As with the creation of gSTEP0, we began with the circularly permuted fluorescent protein from a genetically encoded calcium ion sensor, in this case R-GECO1.[90] R-GECO1 is based on a circularly permuted mApple,[45] which was then optimized through directed evolution, introducing 17 mutations relative to the initial mApple version of GCaMP. This optimization process produced an effective sensor for the detection of calcium ion-mediated binding between calmodulin and the M13 peptide. To make a red STEP, the binding interaction would instead be between the Bim peptide and STEPtag, which would have a different binding surface against the pore of the fluorescent protein. We hypothesized that some of the surface mutations in R-GECO1

would be specific to the M13-calmodulin binding interaction, as the use of directed evolution tends to lead to proteins that are very specifically optimized for the screening conditions. Thus, to begin with a more general template for the red STEP, we reverted most of the R-GECO1 mutations back to their mApple counterparts, keeping only the N-terminal residue modifications, S63V/E64S, as these comprise the linkers next to the pore, and are therefore likely involved in the modulation of signal upon binding, as well as the internal V228A mutation, which seems likely to be involved in core packing, and thus could be generally beneficial to any sensor based on this scaffold (Figure S4.15).

To this cpRFP, we appended the human Bim peptide and the 4 amino acid linker found to be effective in gSTEP1, assuming that the improvements made during the engineering of gSTEP would be beneficial to the sensors of other colours. As with the original development of gSTEP, we began with a unimolecular version of the sensor, with STEPtag.a directly fused to the C-terminus of the circularly permuted RFP, creating uni-rSTEP0. This construct was used to verify that the sensor could express, fold and mature properly, as the maturation of the red chromophore is less efficient process than that of the green, with the possibility of incomplete maturation leading to the formation of a green chromophore by-product.[49] We found that uni-rSTEP0 matured mostly to red chromophore, as shown by an absorption spectrum with a visible light peak at 574 nm (Figure 4.7). The protein was also fluorescent in the red portion of the spectrum, with an excitation peak at 574 nm, and an emission peak at 592 nm. We also found that there was some green fluorescence visible when exciting at 485 nm, indicating that some incomplete maturation was occurring, leading to the presence of green chromophores in the mixture as well.

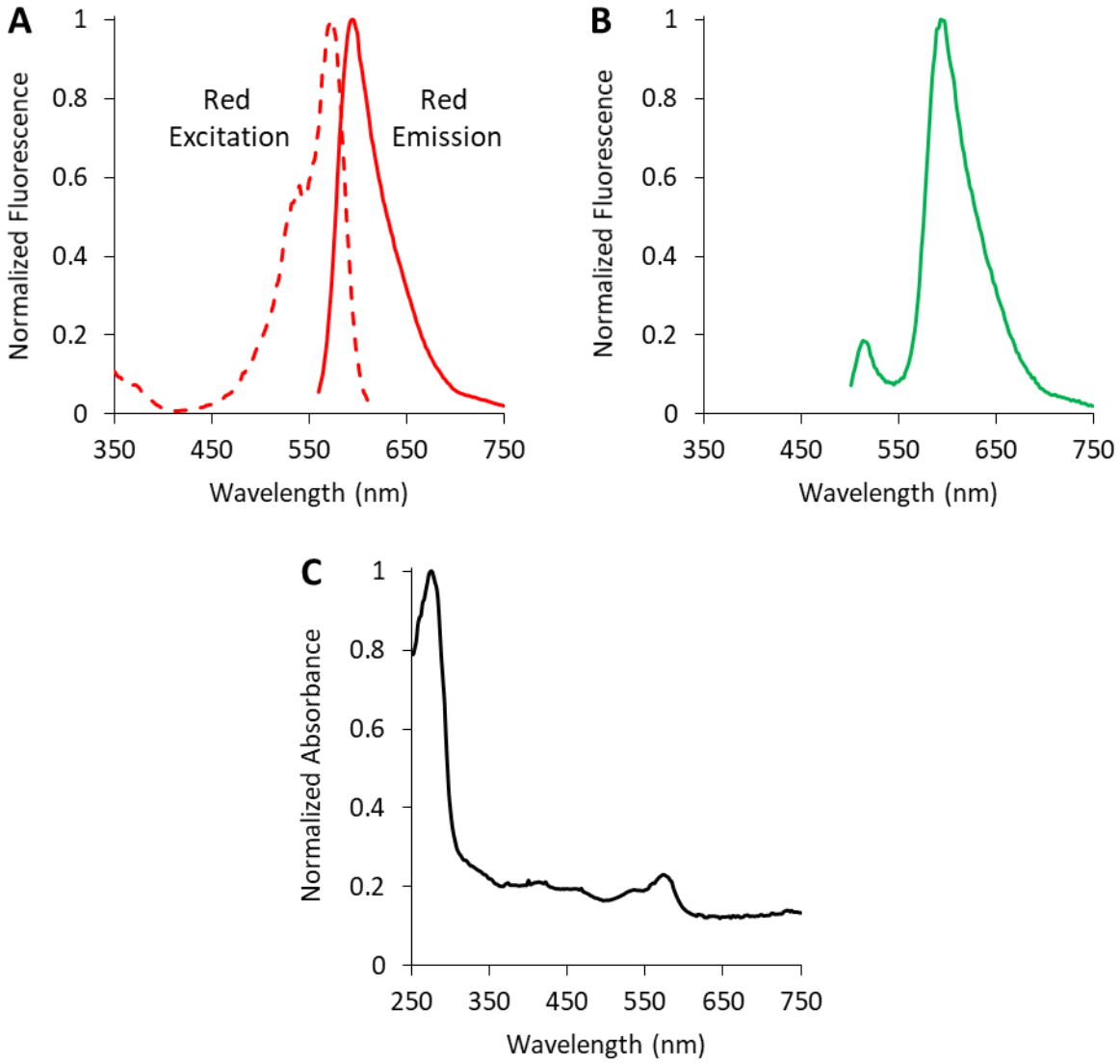


Figure 4.7 Measured spectra of purified uni-rSTEP0. A, The measured red excitation spectrum (emitting at 620 nm) and red emission spectrum (exciting at 540 nm) of uni-rSTEP0 (3.7 µM). **B,** The measured emission spectrum when exciting at 485 nm, to check for green fluorescence. Note that the green and red fluorescence measurements were performed at different gains, and so are not directly comparable in intensity, hence they have been normalized. **C,** The measured absorbance spectra, with a small peak at 574 nm.

Having demonstrated that the unimolecular rSTEP0 construct could mature and fluoresce, we then tested the split construct, rSTEP0, with STEPtag.a removed. We found that rSTEP0 did still mature to a red fluorescent chromophore, albeit with qualitatively lower brightness compared to the gSTEPs, requiring more than 75 nM of FP to obtain a measurable signal (Figure 4.8). Additionally, it

did not show any significant change in fluorescence upon addition of STEPtag, up to a maximum tested concentration of 10 µM STEPtag.a.

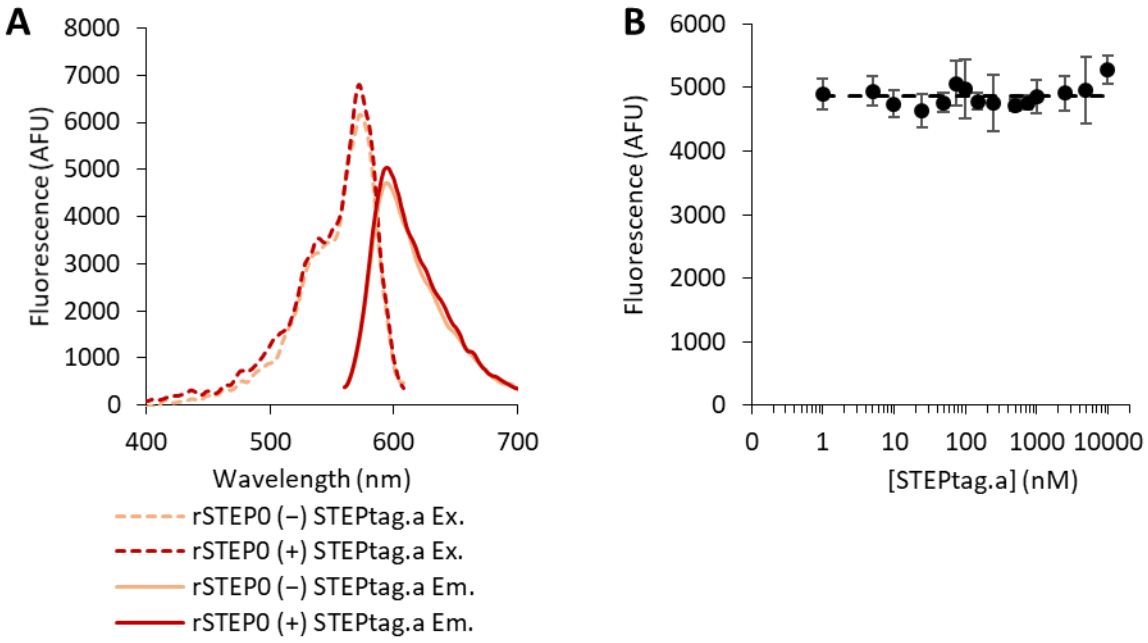


Figure 4.8 Fluorescence spectra and binding curve of rSTEP0. **A**, The measured red excitation spectrum (emitting at 620 nm) and red emission spectrum (exciting at 540 nm) of purified rSTEP0 (1.5 µM). **B,** A representative binding curve, measuring emission at 592 nm when exciting at 540 nm. The dashed line indicates a fit of the Hill equation to the data (n = 1). Error bars are the standard deviation of triplicate measurements.

In an attempt to create a functional red sensor, we created a series of C-terminal truncations of rSTEP0, equivalent to the truncations of gSTEP0 performed during its optimization. We hypothesized that by increasing the size of the pore, fluorescence in the absence of STEPtag could be further quenched, leading to a more detectable increase in fluorescence upon binding. Truncations from one to five amino acids were tested, named rSTEP0-T1 to rSTEP0-T5, respectively. When testing these truncation mutants, we found that they generally expressed and purified poorly, leading to impure samples that were difficult to accurately quantify (Figure S4.5). In addition, none of the truncation mutants showed an increase in fluorescence when STEPtag was added (Figure 4.9). We did discover that the rSTEP0-T4 and rSTEP0-T5 mutants were effectively nonfluorescent, with less than 5% of the rSTEP0 signal for a similar amount of protein. We also found that the truncations seem to change the

maturation kinetics of the protein, as the ratio of green and red fluorescence varies between different mutants, with rSTEP0 showing the least green fluorescence relative to the amount of red, rSTEP0-T2 showing the most signal overall, and rSTEP0-T3 showing a greater relative proportion of green signal. As none of the mutants showed useful sensor activity, and they generally showed less complete maturation, we concluded that none of the rSTEP0 truncations were good candidates for further development of a red STEP, and future efforts will likely be devoted to the improvement of rSTEP0 itself.

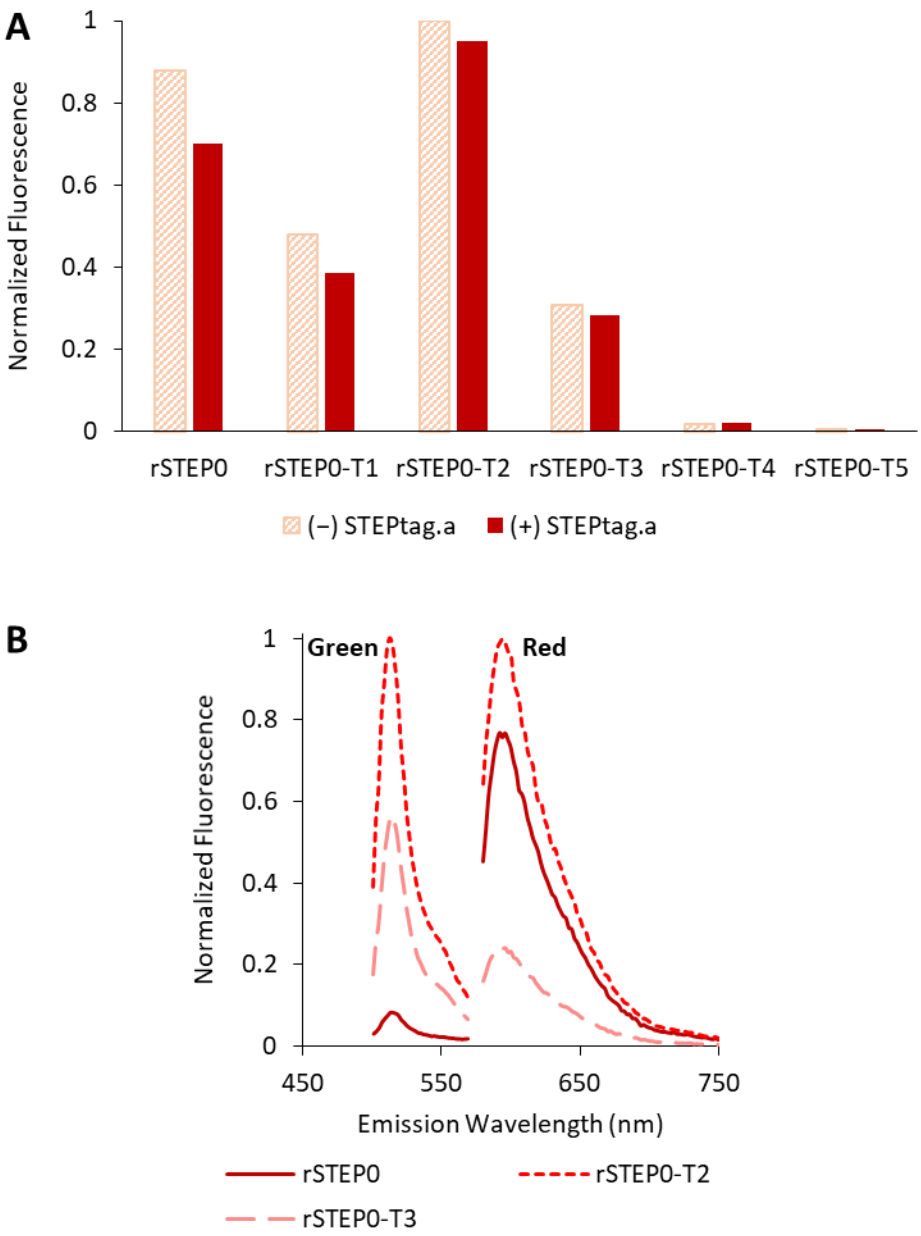


Figure 4.9 Fluorescence of rSTEP0 truncation mutants. **A,** The relative peak red fluorescence emission (exciting at 540 nm) for purified rSTEP0 truncation mutants. (+) STEPtag results are in the presence of 26 µM STEPtag. The fluorescence was normalized as different concentrations of fluorescent protein were used for the various mutants, in order to maximize the measured signal. These measurements were not replicated, thus no error bars are reported. **B,** The relative green and red emission spectra (exciting at 485 nm and 540 nm, respectively) for selected truncation mutants. The fluorescence was normalized as different concentrations of fluorescent protein were used for the various mutants, in order to maximize the measured signal. Each spectrum is normalized separately, as the green and red signals were measured with different gains, and thus are not directly comparable.

As the red STEPs were not particularly effective, we turned to other potential colours of fluorescent protein, namely cyan and yellow. Both of these colours of fluorescent protein have been well studied, and once again, genetically encoded calcium ion sensors containing circular permutations of the fluorescent proteins have been successfully created.[84] To create a cyan STEP, we began with the circularly permuted cyan fluorescent protein barrel used in CyCaMP1a,[84] which is a single mutant of the GCaMP3 barrel with the chromophore tyrosine mutated to a tryptophan residue, a mutation which is known to blue-shift the green fluorescent protein chromophore.[76] As gSTEP1 is originally based on GCaMP3, we therefore created our cyan sensor, cSTEP0, by creating the analogous mutation in gSTEP1, Y202W. When expressed, we found that cSTEP0 did in fact have a blue-shifted fluorescence profile, relative to gSTEP1, with an excitation peak at 442 nm and an emission peak at 490 nm (Figure 4.10). But it also does not seem to change in fluorescence in the presence of STEPtag, with a $\Delta F/F_0$ of 0.1. By fitting a binding curve to the data, a K_d of 1.3 μM was obtained, yet the error on these values is quite high, due to the small change in fluorescence, and so they are unlikely to represent a significant change. Thus, much like rSTEP0, cSTEP0 will require further engineering to increase the $\Delta F/F_0$ and K_d before it can be used as a sensor, to ensure that there is truly a measurable increase when STEPtag is bound.

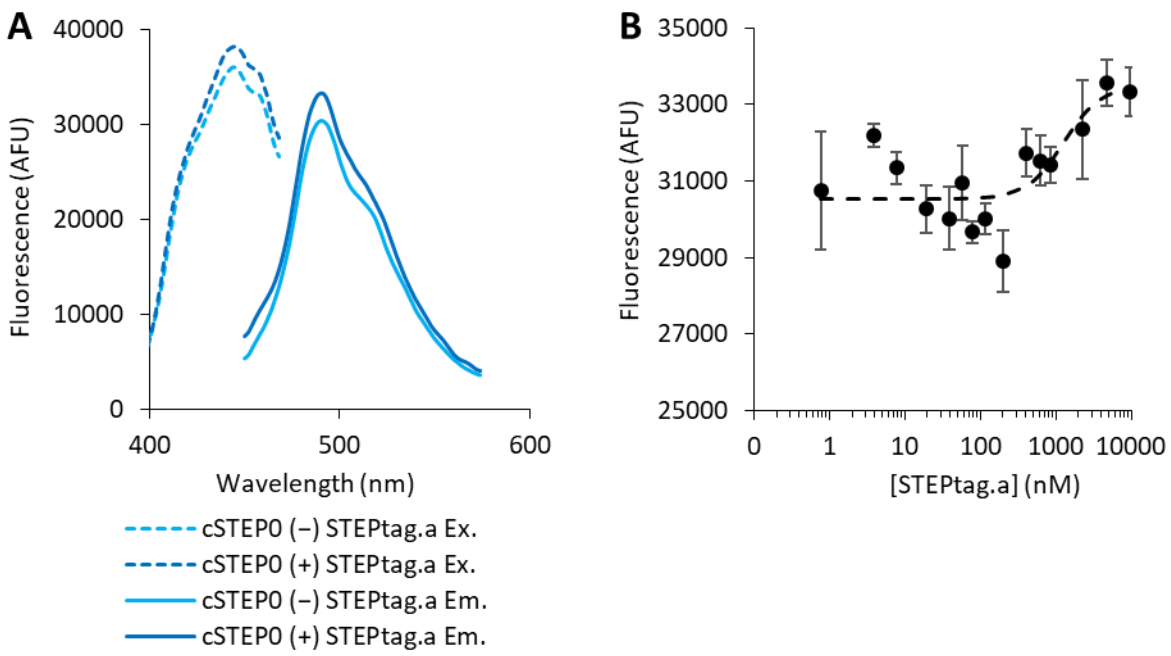


Figure 4.10 Fluorescence spectra and binding curve of cSTEP0. **A,** The measured cyan excitation spectrum (emitting at 490 nm) and cyan emission spectrum (exciting at 430 nm) of purified cSTEP0 (375 nM). **B,** The measured binding curve of cSTEP0 (375 nM), measuring peak emission (490 nm) when exciting at 430 nm. The dashed line indicates a fit of the Hill equation to the data (n = 1.8). Error bars are the standard deviation of triplicate measurements.

For the yellow version of the STEP, we chose to use circularly permuted Venus as the fluorescent protein. Although the yellow versions of the genetically encoded calcium ion sensors similar to GCaMP, such as the pericams,[231] and YCaMP1a,[84] are based on EYFP, Venus is an improved variant of EYFP, with increased brightness and faster maturation.[44] As Venus is one of the most rapidly-maturing fluorescent proteins known, it is the fusion tag that would be used for rapid detection of a protein *in vivo*, and so we felt that it would be a relevant scaffold for our yellow sensor. We therefore introduced the required mutations to convert the green fluorescent protein in gSTEP1 into Venus, creating ySTEP0. Conveniently, gSTEP1 already contains a lysine at the position corresponding to 206 in Venus, a mutation which is expected to reduce aggregation, and has been found to improve *in vivo* expression, in the variant known as mVenus.[184] Thus, we created ySTEP0, containing a circularly permuted mVenus, and found that it was a bright, functional yellow sensor,

with an excitation peak at 512 nm and an emission peak at 524 nm, slightly red-shifted relative to gSTEP1 (Figure 4.11). Unlike rSTEP0 and cSTEP0, ySTEP0 showed sufficient fluorescence to measure binding curves at 75 nM of fluorescent protein, and there was an obvious increase in fluorescence in the presence of STEPtag, with a measured K_d of 280 ± 80 nM, and a $\Delta F/F_0$ of 1.5 ± 0.1. While this is both a higher K_d and a lower $\Delta F/F_0$ than gSTEP1, this is the first iteration of the yellow sensor, and so we expect to be able to optimize it with additional rounds of mutagenesis and design.

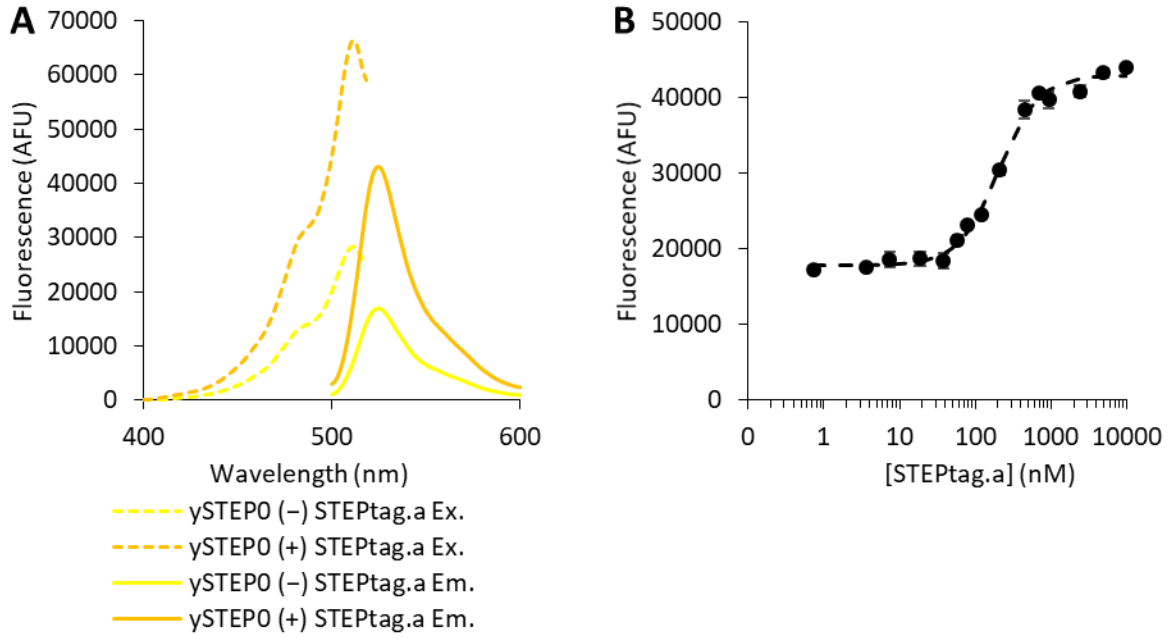


Figure 4.11 Fluorescence spectra and binding curve of ySTEP0. **A,** The measured yellow excitation spectrum (emitting at 530 nm) and yellow emission spectrum (exciting at 490 nm) of purified ySTEP0 (75 nM). **B,** A representative binding curve of ySTEP0 (75 nM), measuring peak emission (526 nm) when exciting at 490 nm. The dashed line indicates a fit of the Hill equation to the data (n = 1.6). Error bars are the standard deviation of triplicate measurements.

Table 4.1 Summary of STEP characteristics.

Sensor	Binding Partner	Excitation Maximum (nm)	Emission Maximum (nm)	K_d (nM)	Hill Coefficient	$\Delta F/F_0$
gSTEP1	STEPtag.a[a]	504	515	120 ± 30	1.6 ± 0.1	3.4 ± 0.5
M13-cpGFP	Calmodulin[b]	498	512	>120 ± 50[f]	1.3 ± 0.1	>2.0 ± 0.7[f]
PDZ-cpGFP	PDZtag[c]	500	513	630	1.3	1.7
cpGFP-PDZ	PDZtag[c]	398/494[g]	511	380	2.4	1.0
gSTEP0-T1-hBim	STEPtag.a[d]	502	515	160 ± 40	2.0 ± 0.3	3.3 ± 0.1
	MBP-STEPtag.b[b]			90 ± 40	1.4 ± 0.1	1.25 ± 0.05
gSTEP0.a	STEPtag.a[d]	486	509	200 ± 100	5 ± 2	0.8 ± 0.3
	STEPtag.b[c]			480	1.8	1.4
gSTEP0.b	STEPtag.a[e]	396/506[g]	515	500 ± 200	1.0 ± 0.2	1.3 ± 0.4
	MBP-STEPtag.b[b]			38 ± 9	1.0 ± 0.2	5 ± 2
rSTEP0	STEPtag.a[b]	574	592	N/A[h]	N/A[h]	N/A[h]
cSTEP0	STEPtag.a[c]	442	490	1300	1.8	0.1
ySTEP0	STEPtag.a[b]	512	524	280 ± 80	1.3 ± 0.2	1.5 ± 0.1

[a] Data obtained from six biological replicates, error shown is the standard deviation
[b] Data obtained from two biological replicates, error shown is the standard deviation
[c] Data obtained from a single biological replicate
[d] Data obtained from three biological replicates, error shown is the standard deviation
[e] Data obtained from four biological replicates, error shown is the standard deviation
[f] Sensor was not confirmed to be saturated, so minimum values are reported, based on a fit of the Hill equation that assumes that saturation occurs at the maximum concentration of binding partner tested.
[g] Green fluorescent proteins with an excitation maximum around 400 nm also have the secondary peak wavelength reported, as the peak around 500 nm was used for the binding curves
[h] No change in fluorescence was detected

4.4 Discussion

Our efforts to expand the variety of STEPs available by varying the binding partners and fluorescent proteins used have had some success, but have also shown that the parts are not fully modular, i.e. each time they are changed, the sensor will likely require optimization.

While most of the different binding pairs tested could function as sensors, they generally did not show the expected affinities or specificities. For instance, the split-GCaMP3 pair has an expected K_d in the sub-nanomolar range, based on the affinities of the M13 peptide and calmodulin in the presence of calcium ions, yet we were unable to saturate the M13-cpGFP sensor with hundreds of nM of calmodulin. Likewise, gSTEP0.b binds MBP-STEPtag.b with a K_d of 38 ± 9 nM, making it the tightest binding sensor that we have tested, yet this K_d is still an order of magnitude greater than the expected value for the Noxa peptide binding to Mcl-1. These results are similar to those seen in gSTEP0 and gSTEP1, where the measured K_d of the sensor is considerably greater than that of the binding pair alone. It seems that the presence of the cpGFP surface has an effect on the binding of the tags to the peptides, which could be mitigated by increasing the linker length between the peptide and the cpGFP, as was shown with gSTEP0-T1 (Chapter 3).

Interestingly, the PDZ-based sensors show increased affinity relative to the literature values, with K_d values in the hundreds of nM, rather than the 8 µM measured previously. Dueber *et al.*[110] state that they used a peptide with the N-terminal glycine replaced by a tyrosine when measuring affinity, to be able to more accurately measure the peptide concentration, which could have an effect on their measured K_d. It could also be that in this case the presence of the fluorescent protein barrel is positively affecting binding. This has been seen in the GCaMP family of sensors, which has calmodulin interacting directly with the surface of the fluorescent protein.[11] Thus, the outer surface of cpGFP near the pore represents another potential region for improvement for all the versions of the STEP, where

affinity and specificity could be increased by adding residues that can directly interact with the binding partner.

In particular, further improvements in specificity are desired, as the expected pair of orthogonal binding partners, gSTEP0.a/STEPtag.a and gSTEP0.b/STEPtag.b, do still show some cross-interactions. Despite the fact that Chen *et al.*[119] measured no detectable binding (IC_{50} > 100 μM) between Mcl-1 and Bad peptide, or Bcl-x_L and Noxa peptide, the STEPs show off-target binding with K_d values around 500 nM, and $\Delta F/F_0$ values around 1.4. While gSTEP0.b shows a distinguishable signal in the presence of the appropriate binding partner, with a K_d an order of magnitude lower and a $\Delta F/F_0$ more than double when adding MBP-STEPtag.b rather than STEPtag.a, any increase in signal due to STEPtag.a will hinder analysis efforts in a multiplexing experiment. It is also possible that some of the affinity is due to the presence of the maltose-binding protein fusion, as we have not yet had the opportunity to test gSTEP0.b with just the STEPtag.b sequence. Should this be the case, it would highlight the possibility of interactions between a fused POI and the STEP, meaning that STEPtag-POI fusions may need to be calibrated on an individual basis. Also, given that the cross-interaction with gSTEP0.b could be problematic, that of gSTEP0.a is even more of an issue, as the K_d with gSTEP0.a is only just below half that with gSTEP0.b, and the $\Delta F/F_0$ was measured to be lower with the "appropriate" binding partner. Thus, gSTEP0.a will have difficulty distinguishing the two tags at all if they are both present in similar concentrations (Figure 4.4), and cannot be used for multiplexing in its current state.

Thus, there are not currently two proven orthogonal binding partners to use with the STEP. While the PDZ-based sensors do seem to function, they have not yet been tested for interactions with STEPtag.a or STEPtag.b, and have higher K_d values than the current gSTEPs. The gSTEP0.a and gSTEP0.b constructs currently show some cross-interactions, but as they are the first generation of STEPs to use their respective peptides, it is likely that they could be improved through either rational

design or directed evolution techniques, as was gSTEP0. Given that the Bad peptide-based gSTEP0.a seems to have almost no selectivity for Bcl-x_L-based STEPtag.a, it could be that another peptide would perform better in its place, for instance the synthetic peptides designed by Dutta *et al.*[120] Some of these peptides were tested in Chapter 3, and were not found to outperform the Bim peptide present in gSTEP0-T1-hBim when comparing the K_d and $\Delta F/F_0$, but they still functioned as sensors, and are reported to bind to Bcl-x_L with at least three orders of magnitude more affinity than Mcl-1. If this specificity is maintained in the STEP scaffold, a trade-off of sensor efficacy for specificity could potentially be made by using these peptides when multiplexing is required.

The other part of the sensor that was varied here is the circularly permuted fluorescent protein, in order to change the colour of the STEP. For the colours tested, only ySTEP0, based on the mVenus scaffold, maintained its ability to detect STEPtag.a. rSTEP0 showed no increase at all, nor did any of its truncation mutants, and neither did cSTEP0, despite it only being a single point mutation away from the parent gSTEP1. A certain loss in $\Delta F/F_0$ was expected, based on similar efforts to create multiple colours of calcium ion sensors based on GCaMP3,[84, 90] which saw $\Delta F/F_0$ values drop to roughly 25% of the green parent value for yellow and cyan sensors. For the red versions in that experiment, many of the initial sensors failed to mature at all, generally requiring optimization of the linker regions and/or directed evolution to obtain variants that showed a response to calcium ions. Thus, the next steps for cSTEP0 and rSTEP0 will likely be randomization of the linker regions, using a high-throughput screen similar to that used in Chapter 3 to identify variants with improvements in $\Delta F/F_0$.

Another possibility for creating an improved cyan STEP would be to change the scaffold being used. cSTEP0 is based on CyCaMP1a (itself based on the original CFP), and required only a single point mutation to generate from gSTEP1, hence why it was chosen as the initial cyan STEP. There are improved variants of cyan fluorescent proteins, such as mTurquoise2,[42] with increased quantum yields, which could potentially become sensors with increased overall brightness. Much of the improvements

in these proteins arises from stabilization of the seventh β-strand of the barrel,[232] leading to improved packing and a more planar chromophore, which contributes to an increase in quantum yield. This seventh β-strand happens to be the position of the circular permutation in cSTEP0, so it is possible that the mutations will not show equivalent improvements to the STEP. On the other hand, if the appropriate configuration of the strand is preferentially stabilized when STEPtag.a is bound, and destabilized when it is unbound, this could provide a mechanism for a large increase in the $\Delta F/F_0$, leading to a greatly improved sensor.

In a similar fashion, a functional red STEP could potentially be generated using a different red fluorescent protein as a scaffold, and we have already generated a version of the sensor based on rCaMP1h, rather than R-GECO1, to see if it showed better results. Unfortunately, it did not express well enough to measure a red fluorescence spectrum, and so we were not able to test its response to STEPtag.a (Data not shown). Another consideration for rSTEP0 is identifying the cause of the lack of change when STEPtag.a is added. Because red fluorescent proteins have sequences that are considerably different from those of the cyan, green and yellow fluorescent proteins, the equivalent STEP sequences are also quite different (Figure S4.16). Thus, if the surface of the barrel is involved in modulating binding between the STEP and the STEPtag, as we hypothesize based on some of the results seen with the different binding partners, the red STEP would likely have different interactions being made, and the barrel may in fact be destabilizing the binding between STEPtag.a and the target Bim peptide. Thus, it may be valuable to perform some fluorescence-independent test of binding, such as surface-plasmon resonance, or isothermal titration calorimetry, to determine if rSTEP0 binds to STEPtag.a at all. If STEPtag.a does indeed bind, then the problem is one of fluorescence modulation, and thus the planned linker modifications are likely to help. If STEPtag.a does not bind, then the interaction surface would be a more effective target for improvement.

Overall, while the STEP cannot yet be used in a multiplexed experiment to detect two proteins at a time, and the replacement of binding partners and fluorescent proteins has shown that the resulting STEP will not necessarily have the characteristics expected from the parts used, a number of new scaffolds have been generated that can be used to develop new sensors. The gSTEP0.b/MBP-STEPtag.b binding pair show the lowest K_d and highest $\Delta F/F0$ of any of the STEPs, even if gSTEP.a does not show the orthogonal binding that was intended. Cyan and red versions of the STEP have been created that at least express and form mature chromophore, providing a baseline that can be improved through mutagenesis. For situations where a yellow fluorescent sensor would be preferred, ySTEP0 functions with similar effectiveness to the original gSTEP0, and the M13-cpGFP construct has opened the door to creating conditional sensors, by demonstrating a change in fluorescence that only occurs when both the binding protein (calmodulin) and an effector molecule (Ca^{2+}) are present.

4.5 Materials and Methods

4.5.1 Construct Preparation

GCaMP3 and mouse α-1-syntrophin PDZ domain sequences were a generous gift from Hernan G. Garcia. M13-cpGFP and calmodulin sequences were generated by PCR amplification of the appropriate regions of GCaMP3 using synthetic oligonucleotides (Eurofins) with additional nucleotides added to insert His-tags and NdeI/BamHI restriction sites for cloning. cpGFP-PDZ, PDZ-cpGFP and gSTEP0.b were generated in the same manner, using gSTEP0-T1 as a template. The Mcl-1 (STEPtag.b) sequence was a generous gift from Samuel H. Gellman, and MBP-Mcl-1 (MBP-STEPtag.b) was a generous gift from Amy E. Keating. gSTEP0.a, uni-rSTEP0 and ySTEP0 sequences were all codon optimized for *E. coli*, synthesized and cloned into pET11a vectors by Genscript. rSTEP0 was generated by PCR amplification of the appropriate regions of uni-rSTEP0, and all truncations were generated in the same way using rSTEP0 as a template. cSTEP0 was generated using splicing by overlap extension (SOE) mutagenesis,[225] in order to perform site-directed mutagenesis of the gSTEP1

protein sequence. All sequences were subcloned into pET11a vectors, using NdeI and BamHI restriction sites. All reagents used for subcloning were obtained from New England Biolabs.

4.5.2 Protein Expression and Purification for *in vitro* screening and characterization

All vectors were transformed into chemically competent *E. coli* BL21 (DE3) Gold (Agilent). Expression was performed by culturing cells in Luria-Bertani (LB) broth, supplemented with 100 µg/mL ampicillin. Cells were grown at 37°C with shaking to an OD600 of 0.6-0.8, at which point protein expression was induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were then incubated with shaking overnight at 16°C, harvested by centrifugation and lysed using an EmulsiFlex-B15 cell disruptor (Avestin). Proteins were purified by immobilized metal affinity chromatography using Profinity Nickel-charged IMAC resin (Bio-Rad), in a gravity flow column, according to the manufacturer's protocol. Buffer exchange into 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4) and sample concentration was performed using Amicon Ultra-15 centrifugal filters with a 3K molecular weight cut-off (Millipore) for STEPtag.a and PDZtag, and Microsep Advance centrifugal filters with a 10K molecular weight cut-off (Pall) for all other proteins.

4.5.3 Fluorescence Assays

PDZtag and Calmodulin were quantified using the Quick-Start Protein Assay Kit (Bio-Rad) to perform the linearized Bradford Assay,[220] while all other proteins were quantified by measuring absorbance at 280 nm in a cuvette using a SpectraMax Plus[384] microplate spectrophotometer (Molecular Devices). Extinction coefficients for the proteins were estimated using the ProtParam online tool.[221] All measurements were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4), at room temperature. Fluorescence spectra of all proteins were recorded on an Infinite M1000 microplate reader (Tecan). Unless otherwise specified, all fluorescence measurements were performed using 75 nM of fluorescent protein, and each sample well was prepared in triplicate. To calculate the K_d, the fluorescence signal was used to fit the Hill equation, accounting for ligand

depletion,[5] across the range of binding partner concentrations used. $\Delta F/F_0$ values for all the binding curves were calculated using the Hill equation fit to determine F_{max} and F_{min}, where F_{max} is the maximum signal of the fit, and F_{min} is the minimum signal of the fit.

4.6 Supplementary Information

4.6.1 Amino Acid Sequences

M13-cpGFP
MGSHHHHHHGMASMTGGQQMGRDLYDDDDKDLATMVDSSRRKWNKTGHAVRAIGRLSSLENVYIKAD
KQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLE
FVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTL
KFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAE
VKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNTR*

Calmodulin
MTRDQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGDGTIDF
PEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREADIDG
DGQVNYEEFVQMMTAKHHHHHH*

PDZ-cpGFP
MHHHHHHGSGSGGVKESLVGSGSLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGD
GPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVV
PILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMK
QHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEY*

cpGFP-PDZ
MLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPN
EKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGE
GDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFK
DDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYGSGSGGVKESLVGSGSHHHHHH*

PDZtag
MHHHHHHLQRRRVTVRKADAGGLGISIKGGRENKMPILISKIFKGLAADQTEALFVGDAILSVNGED
LSSATHDEAVQALKKTGKEVVLEVKYMKEVSPYFKSGSG*

gSTEP0.a
MHHHHHHNLWAAQRYGRELRRMSDEFVDSFKKGLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAY
HYQQNTPIGDGPVLLPDNHYLSTQNKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFFCTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGYYKTRAEVKFEGDTLVNRIVLKGIDFKEDGNI
LGHKLEYNTR*

gSTEP0.b
MHHHHHHPAELEVECATQLRRFGDKLNFRQKLLLENVYIKADKQKNGIKANFKIRHNIEDGGVQLAY
HYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVS
KGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQ
CFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNI
LGHKLEY*

MBP-STEPtag.b

MGSSHHHHHHGSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAAT
GDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKD
LLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAG
AKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQP
SKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATM
ENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGIEENLY
FQGSDELYRQSLEIISRYLREQATGAKDTKPMGRSGATSRKALETLRRVGDGVQRNHETAFQGMLRK
LDIKNEDDVKSLSRVMIHVFSDGVTNWGRIVTLISFGAFVAKHLKTINQESCIEPLAESITDVLVRT
KRDWLVKQRGWDGFVEFFHVEDLEGG*

STEPtag.b

MGSHHHHHHHHGSDYDIPTTENLYFQGSEDELYRQSLEIISRYLREQATGAKDTKPMGRSGATSRKA
LETLRRVGDGVQRNHETAFQGMLRKLDIKNEDDVKSLSRVMIHVFSDGVTNWGRIVTLISFGAFVAK
HLKTINQESCIEPLAESITDVLVRTKRDWLVKQRGWDGFVEFFHVEDLEGG*

uni-rSTEP0

MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRPVVSERMYPEDGALKSEIKKRLKLKDGGHY
AAEVKTTYKAKKPVQLPGAYIVDIKLDIVSHNEDYTIVEQYERAEGRHSTGGMDELYKGGTGGSMVS
KGEENNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEAFQTAKLKVTKGGPLPFAWDILSPQF
MYGSKAYIKHPADIPDYFKLSFPEGFRWERVMNFEDGGIIHVNQDSSLQDGVFIYKVKLRGTNFPSD
GPVMQKKTMGWEATRREVIPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNE
LFRDGVNWGRIVAFFSFGGALCVESVDKEMQVLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYG
NNAAAESRK*

rSTEP0

MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRPVVSERMYPEDGALKSEIKKRLKLKDGGHY
AAEVKTTYKAKKPVQLPGAYIVDIKLDIVSHNEDYTIVEQYERAEGRHSTGGMDELYKGGTGGSMVS
KGEENNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEAFQTAKLKVTKGGPLPFAWDILSPQF
MYGSKAYIKHPADIPDYFKLSFPEGFRWERVMNFEDGGIIHVNQDSSLQDGVFIYKVKLRGTNFPSD
GPVMQKKTMGWEATR*

rSTEP0-T1

MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRPVVSERMYPEDGALKSEIKKRLKLKDGGHY
AAEVKTTYKAKKPVQLPGAYIVDIKLDIVSHNEDYTIVEQYERAEGRHSTGGMDELYKGGTGGSMVS
KGEENNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEAFQTAKLKVTKGGPLPFAWDILSPQF
MYGSKAYIKHPADIPDYFKLSFPEGFRWERVMNFEDGGIIHVNQDSSLQDGVFIYKVKLRGTNFPSD
GPVMQKKTMGWEAT*

rSTEP0-T2

MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRPVVSERMYPEDGALKSEIKKRLKLKDGGHY
AAEVKTTYKAKKPVQLPGAYIVDIKLDIVSHNEDYTIVEQYERAEGRHSTGGMDELYKGGTGGSMVS
KGEENNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEAFQTAKLKVTKGGPLPFAWDILSPQF
MYGSKAYIKHPADIPDYFKLSFPEGFRWERVMNFEDGGIIHVNQDSSLQDGVFIYKVKLRGTNFPSD
GPVMQKKTMGWEA*

rSTEP0-T3

MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRPVVSERMYPEDGALKSEIKKRLKLKDGGHY
AAEVKTTYKAKKPVQLPGAYIVDIKLDIVSHNEDYTIVEQYERAEGRHSTGGMDELYKGGTGGSMVS
KGEENNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEAFQTAKLKVTKGGPLPFAWDILSPQF
MYGSKAYIKHPADIPDYFKLSFPEGFRWERVMNFEDGGIIHVNQDSSLQDGVFIYKVKLRGTNFPSD
GPVMQKKTMGWE*

rSTEP0-T4

MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRPVVSERMYPEDGALKSEIKKRLKLKDGGHY
AAEVKTTYKAKKPVQLPGAYIVDIKLDIVSHNEDYTIVEQYERAEGRHSTGGMDELYKGGTGGSMVS
KGEENNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEAFQTAKLKVTKGGPLPFAWDILSPQF
MYGSKAYIKHPADIPDYFKLSFPEGFRWERVMNFEDGGIIHVNQDSSLQDGVFIYKVKLRGTNFPSD
GPVMQKKTMGW*

rSTEP0-T5

MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRPVVSERMYPEDGALKSEIKKRLKLKDGGHY
AAEVKTTYKAKKPVQLPGAYIVDIKLDIVSHNEDYTIVEQYERAEGRHSTGGMDELYKGGTGGSMVS
KGEENNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEAFQTAKLKVTKGGPLPFAWDILSPQF
MYGSKAYIKHPADIPDYFKLSFPEGFRWERVMNFEDGGIIHVNQDSSLQDGVFIYKVKLRGTNFPSD
GPVMQKKTMG*

ySTEP0

MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRLENVYITADKQKNGIKANFKIRHNIEDGGV
QLADHYQQNTPIGDGPVLLPDNHYLSYQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGG
SMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKLICTTGKLPVPWPTLVTTLG
YGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEY*

cSTEP0

MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRLENVYIKADKQKNGIKANFKIRHNIEDGGV
QLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGG
SMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLT
WGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKE
DGNILGHKLEY*

4.6.2 Sample SDS-PAGE Gels

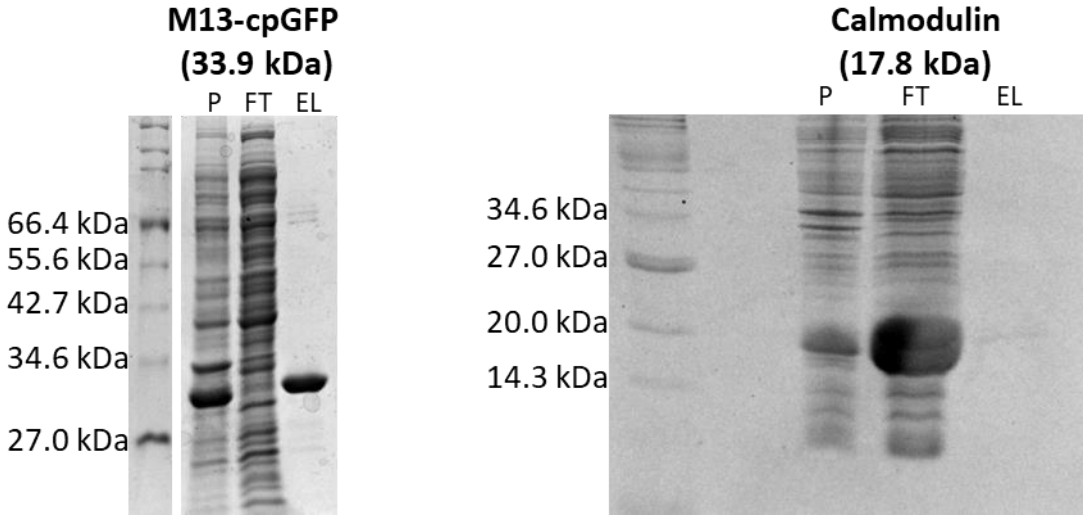


Figure S4.1 Sample SDS-PAGE gels of purification steps of M13-cpGFP and Calmodulin. Samples taken during protein purification were mixed 1:1 by volume with SDS-PAGE loading dye, heated to 95°C for 5 minutes, then loaded into either 10% (M13-cpGFP) or 15% (Calmodulin) acrylamide SDS-PAGE gels. The ladder used is the P7702 Unstained Protein Marker, Broad Range (NEB). Electrophoresis was performed at 120 volts until the dye front had migrated to the end of the gel. Abbreviations are as follows: P – pellet, FT – flowthrough, EL – eluted (purified). Note the low yield of calmodulin, leading to the inability to saturate M13-cpGFP.

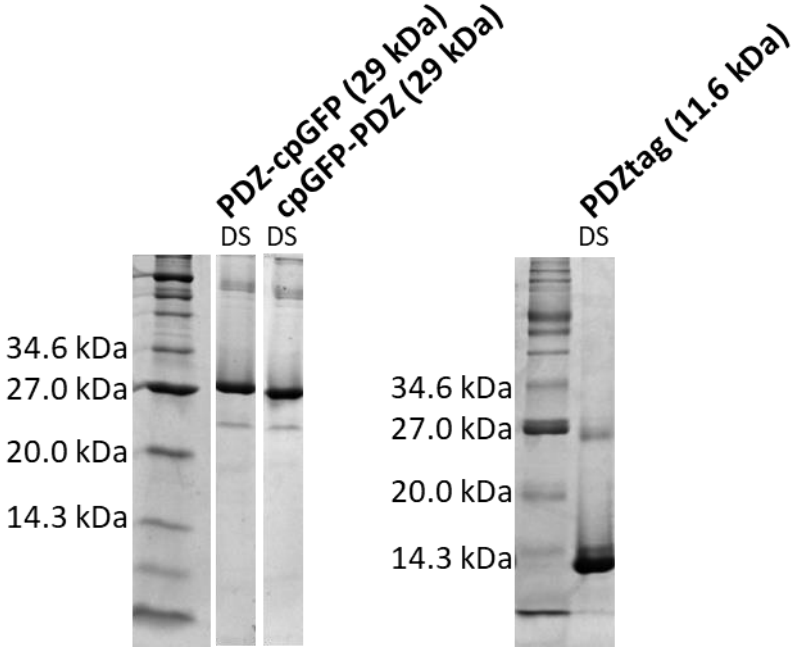


Figure S4.2 Sample SDS-PAGE gels of purified PDZ-based sensors. Samples were mixed 1:1 by volume with SDS-PAGE loading dye, heated to 95°C for 5 minutes, then loaded into 15% acrylamide SDS-PAGE gels. The ladder used is the P7702 Unstained Protein Marker, Broad Range (NEB). Electrophoresis was performed at 120 volts until the dye front had migrated to the end of the gel. Abbreviations are as follows: DS – desalted

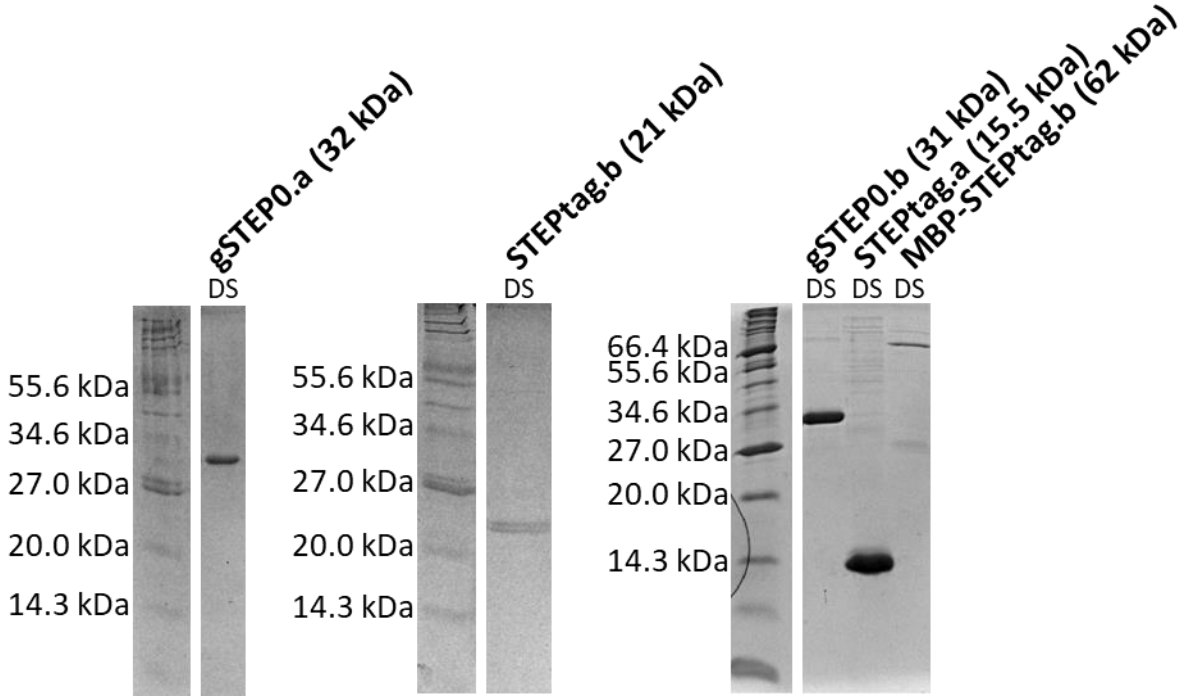


Figure S4.3 Sample SDS-PAGE gels of purified gSTEP.a and gSTEP.b sensors. Samples were mixed 1:1 by volume with SDS-PAGE loading dye, heated to 95°C for 5 minutes, then loaded into 15% acrylamide SDS-PAGE gels. The ladder used is the P7702 Unstained Protein Marker, Broad Range (NEB). Electrophoresis was performed at 120 volts until the dye front had migrated to the end of the gel. Abbreviations are as follows: DS – desalted

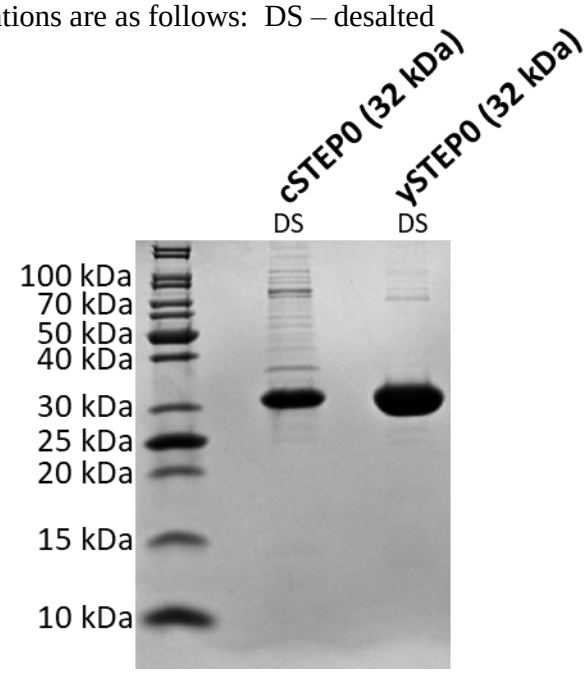


Figure S4.4 Sample SDS-PAGE gels of purified cSTEP0 and ySTEP0 sensors. Samples were mixed 1:1 by volume with SDS-PAGE loading dye, heated to 95°C for 5 minutes, then loaded into 15% acrylamide SDS-PAGE gels. The ladder used is the P7717 Unstained Protein Standard, Broad Range (NEB). Electrophoresis was performed at 120 volts until the dye front had migrated to the end of the gel. Abbreviations are as follows: DS – desalted

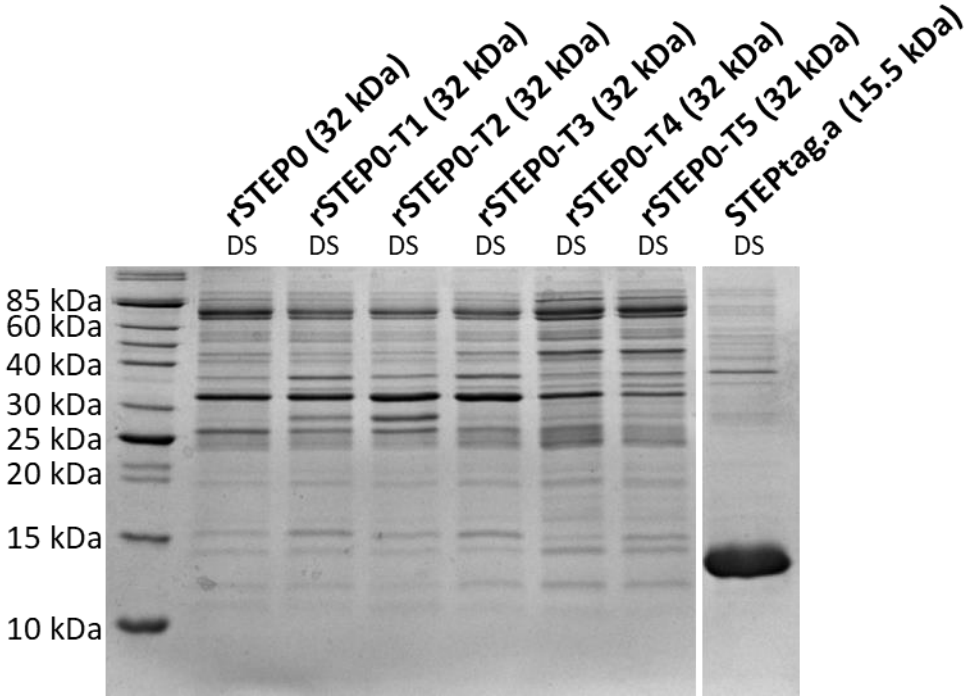


Figure S4.5 Sample SDS-PAGE gels of purified rSTEP0 and rSTEP0 truncated sensors. Samples were mixed 1:1 by volume with SDS-PAGE loading dye, heated to 95°C for 5 minutes, then loaded into 15% acrylamide SDS-PAGE gels. The ladder used is the P7717 Unstained Protein Standard, Broad Range (NEB). Electrophoresis was performed at 120 volts until the dye front had migrated to the end of the gel. Abbreviations are as follows: DS – desalted. Note that red fluorescent proteins undergo cleavage at the acylimine double bond of the chromophore when heat denatured in preparation for SDS-PAGE.[7] In rSTEP0, this cleavage should occur at position M202, thereby producing fragments of 23 kDa and 9.5 kDa. The truncation mutants should have the same 23 kDa fragment, with the smaller fragment ranging from 9.3 kDa to 8.8 kDa.

4.6.3 Raw Fluorescence Data

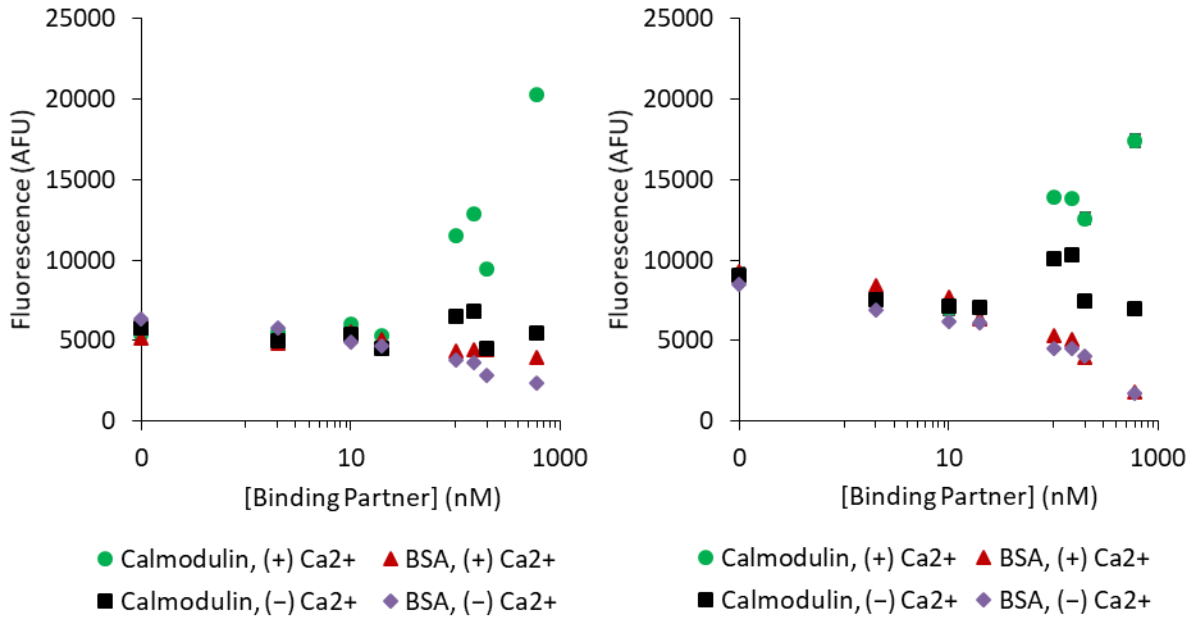


Figure S4.6 M13-cpGFP fluorescence in the presence of calmodulin and BSA. Two biological replicates are shown here. Fluorescence shown is the maximum emission at 550 nm when exciting around 500 nm, with measurements performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). The (+) Ca2+ curves have 1 mM of $CaCl_2$ added, while the (−) Ca2+ curves have 1 mM of EGTA added. Because the signal could not be saturated, no fit of the Hill equation is included. Error bars are the standard deviation of duplicate measurements.

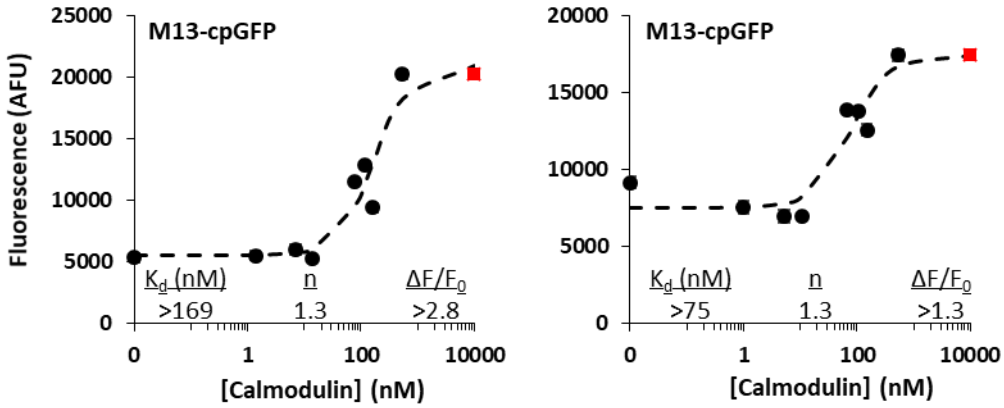


Figure S4.7 Binding curves of M13-cpGFP. Excitation scans of 75 nM of biological replicates of M13-cpGFP, mixed with varying concentrations of calmodulin, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl, and 1 mM $CaCl_2$ (pH 7.4). Each experiment was performed with a fresh purification of Calmodulin. The fluorescence signal is the emission at 550 nm, when exciting at 500 nm. As 600 nM was the maximum concentration of Calmodulin tested and the signal was not saturated, a point at 10 000 nM of the same fluorescence was added (red square), to allow a minimum value to be estimated for the parameters of the Hill equation. The dashed lines are the results of a fit of the Hill equation to the values, and K_d, $\Delta F/F_0$ and the value of n calculated from the fit are reported for each curve. Error bars are the standard deviation of duplicate measurements.

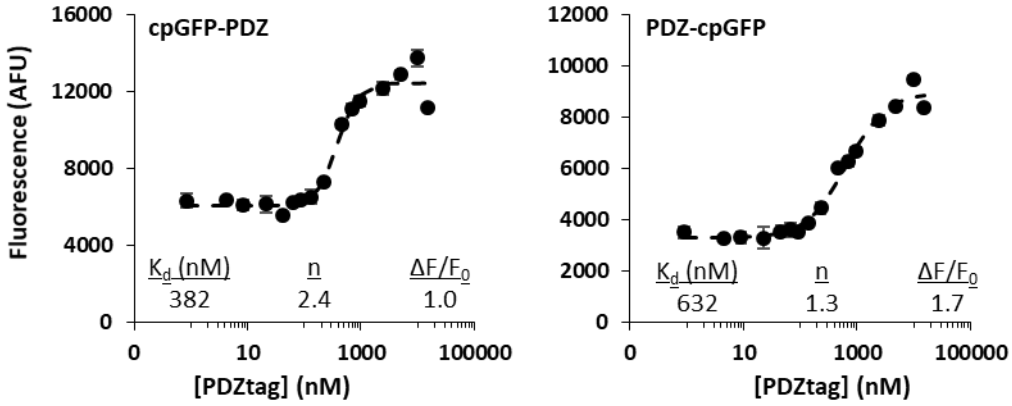


Figure S4.8 Binding curves of PDZ-based sensors. Emission scans of 75 nM of purified cpGFP-PDZ and PDZ-cpGFP, mixed with varying concentrations of PDZ-tag, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). Both experiments were performed with the same purification of PDZtag. The fluorescence signal is the emission at 515 nm, when exciting at 485 nm. The dashed lines are the results of a fit of the Hill equation to the values, and K_d, $\Delta F/F_0$ and the value of n calculated from the fit are reported for each curve. Error bars are the standard deviation of triplicate measurements.

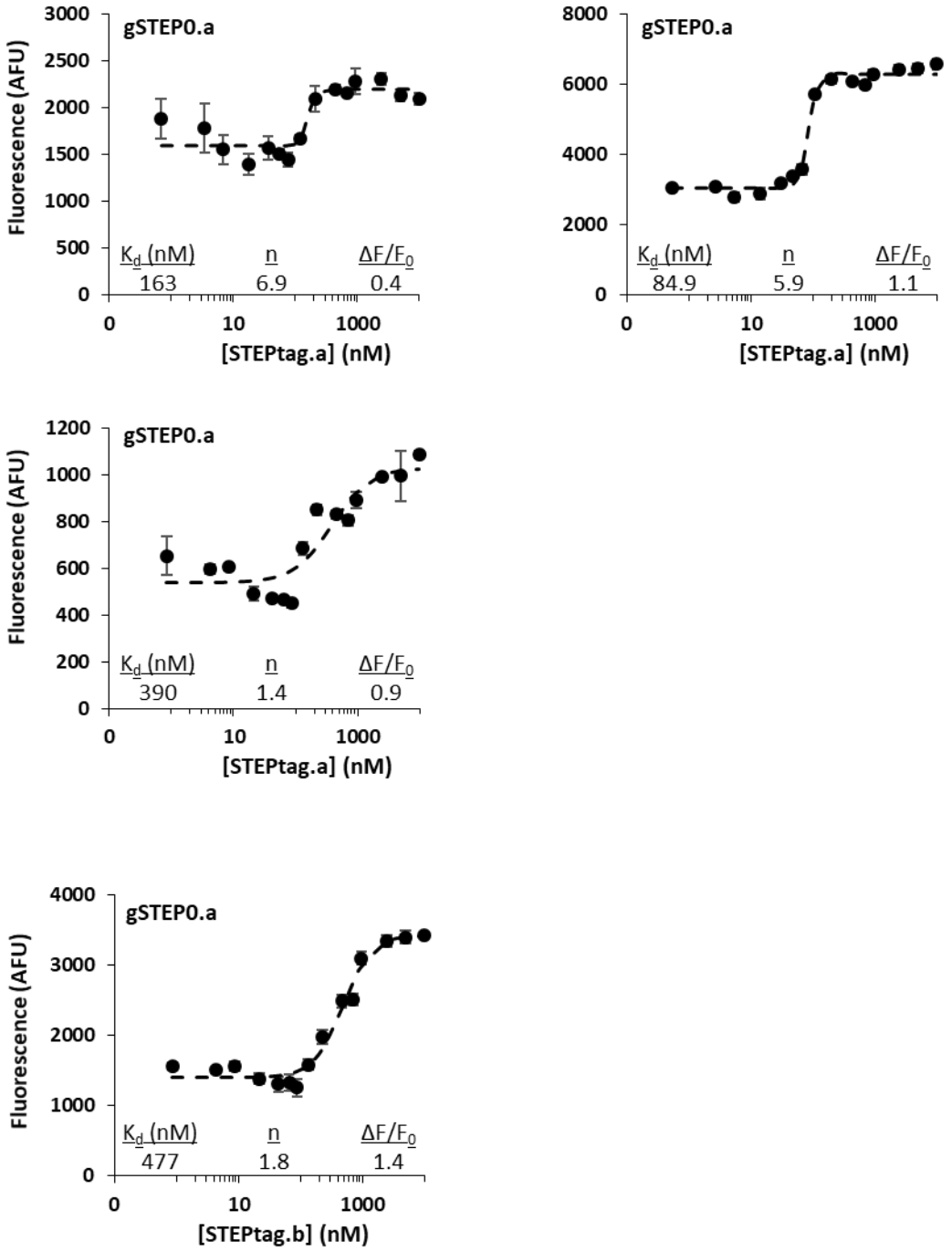


Figure S4.9 Binding curves of gSTEP0.a and STEPtags. Emission scans of 75 nM of biological replicates of gSTEP0.a, mixed with varying concentrations of STEPtag.a or STEPtag.b, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). Each experiment was performed with a fresh purification of the associated STEPtag. The fluorescence signal is the emission at 515 nm, when exciting at 485 nm. The dashed lines are the results of a fit of the Hill equation to the values, and K_d, $\Delta F/F_0$ and the value of n calculated from the fit are reported for each curve. Error bars are the standard deviation of triplicate measurements.

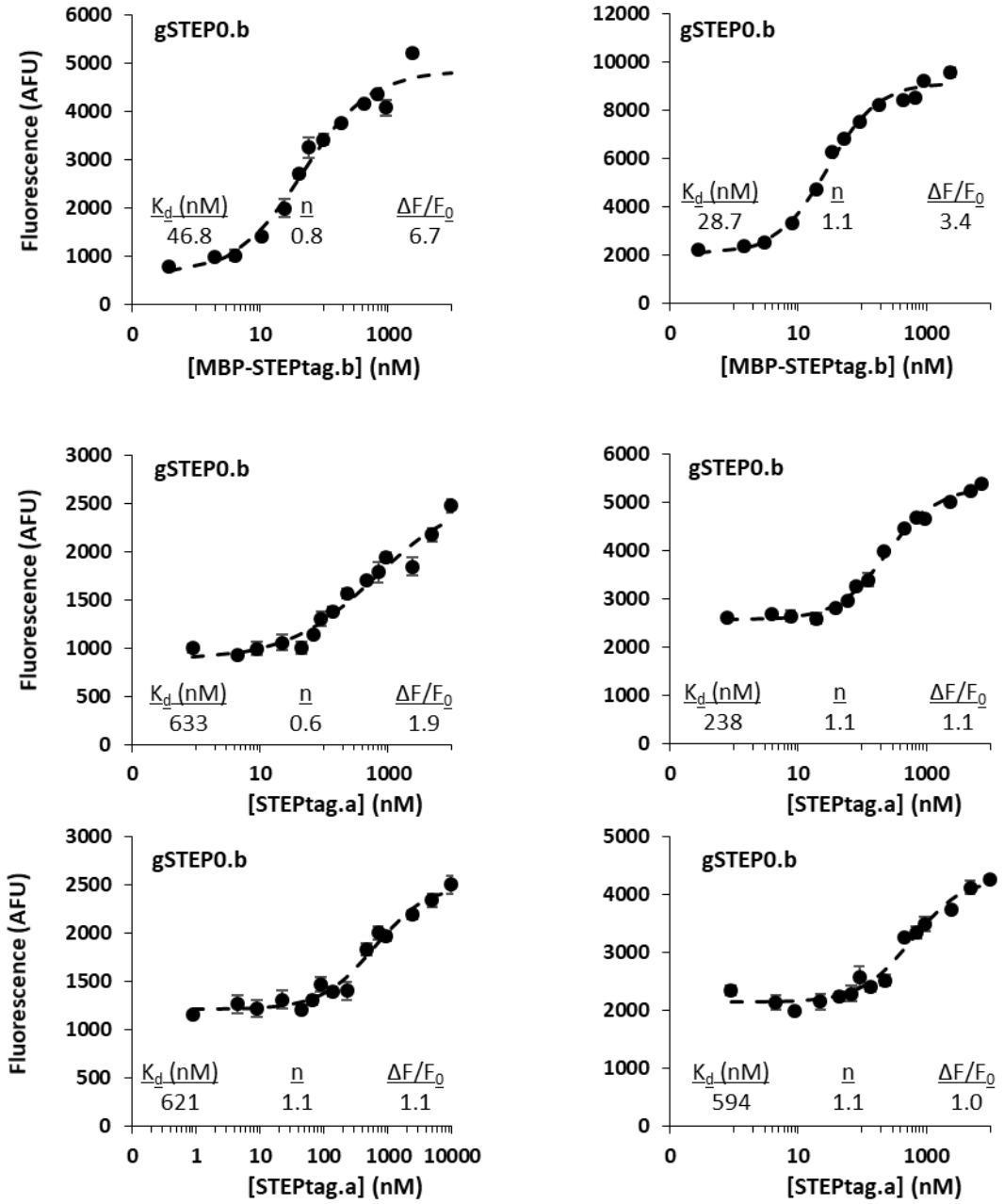


Figure S4.10 Binding curves of gSTEP0.b and STEPtags. Emission scans of 75 nM of biological replicates of gSTEP0.b, mixed with varying concentrations of MBP-STEPtag.b or STEPtag.a, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). Each experiment was performed with a fresh purification of the associated STEPtag. The fluorescence signal is the emission at 515 nm, when exciting at 485 nm. The dashed lines are the results of a fit of the Hill equation to the values, and K_d, $\Delta F/F_0$ and the value of n calculated from the fit are reported for each curve. Error bars are the standard deviation of triplicate measurements.

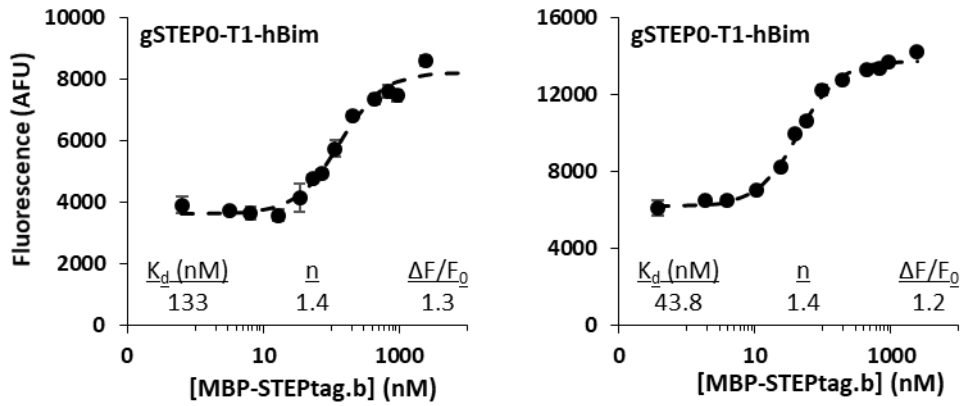


Figure S4.11 Binding curves of gSTEP0-T1-hBim and MBP-STEPtag.b. Emission scans of 75 nM of biological replicates of gSTEP0-T1-hBim, mixed with varying concentrations of MBP-STEPtag.b, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). Both experiments were performed with fresh purifications of MBP-STEPtag.b. The fluorescence signal is the emission at 515 nm, when exciting at 485 nm. The dashed lines are the results of a fit of the Hill equation to the values, and K_d, $\Delta F/F_0$ and the value of n calculated from the fit are reported for each curve. Error bars are the standard deviation of triplicate measurements.

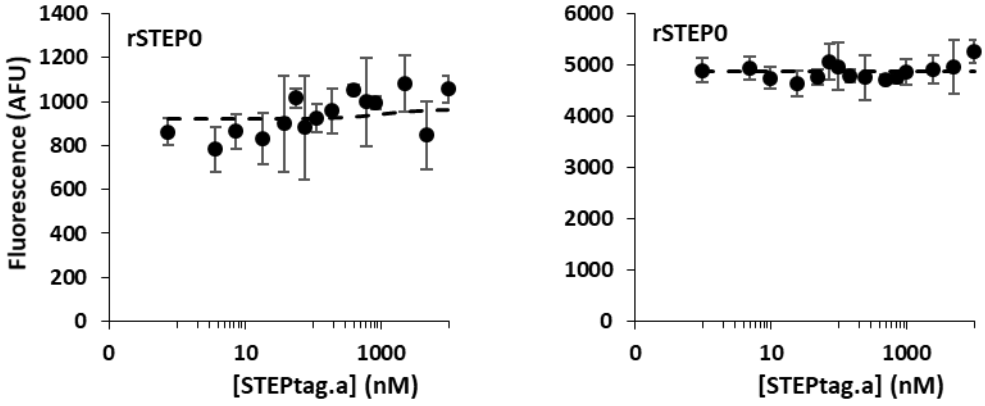


Figure S4.12 Binding curves of rSTEP0. Emission scans of 350 nM (left) or 1475 nM (right) of biological replicates of rSTEP0, mixed with varying concentrations of STEPtag, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). Each experiment was performed with a fresh purification of STEPtag.a. The fluorescence signal is the emission at 592 nm, when exciting at 540 nm. The dashed lines are the results of a fit of the Hill equation to the data, although in this case the fit indicates no binding, and so no parameters are reported. Error bars are the standard deviation of triplicate measurements.

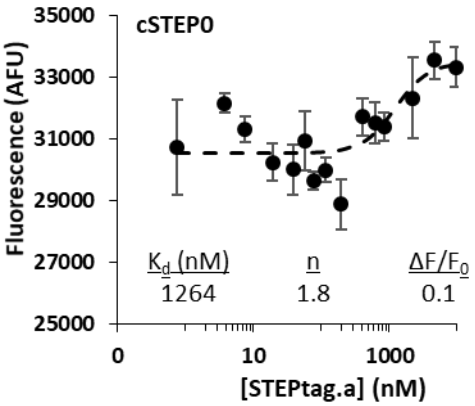


Figure S4.13 Binding curve of cSTEP0. Emission scans of 375 nM of purified cSTEP0, mixed with varying concentrations of STEPtag.a, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). The fluorescence signal is the emission at 490 nm, when exciting at 430 nm. The dashed lines are the results of a fit of the Hill equation to the data, and K_d, $\Delta F/F_0$ and the value of n calculated from the fit are reported. Error bars are the standard deviation of triplicate measurements.

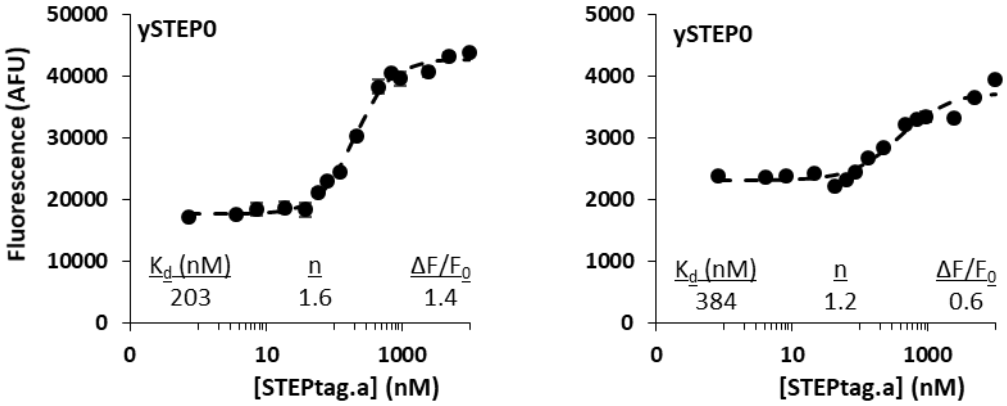


Figure S4.14 Binding curve of ySTEP0. Emission scans of 75 nM of biological replicates of ySTEP0, mixed with varying concentrations of STEPtag.a, were performed in 20 mM sodium phosphate buffer containing 50 mM NaCl (pH 7.4). Each experiment was performed with a fresh purification of STEPtag.a. The fluorescence signal is the emission at 530 nm, when exciting at either 490 nm (left) or 512 nm (right). Note that the gain used differed between these two experiments, so the fluorescence intensities are not directly comparable. The dashed lines are the results of a fit of the Hill equation to the data, and K_d, $\Delta F/F_0$ and the value of n calculated from the fit are reported. Error bars are the standard deviation of triplicate measurements.

4.6.4 Supplementary Images

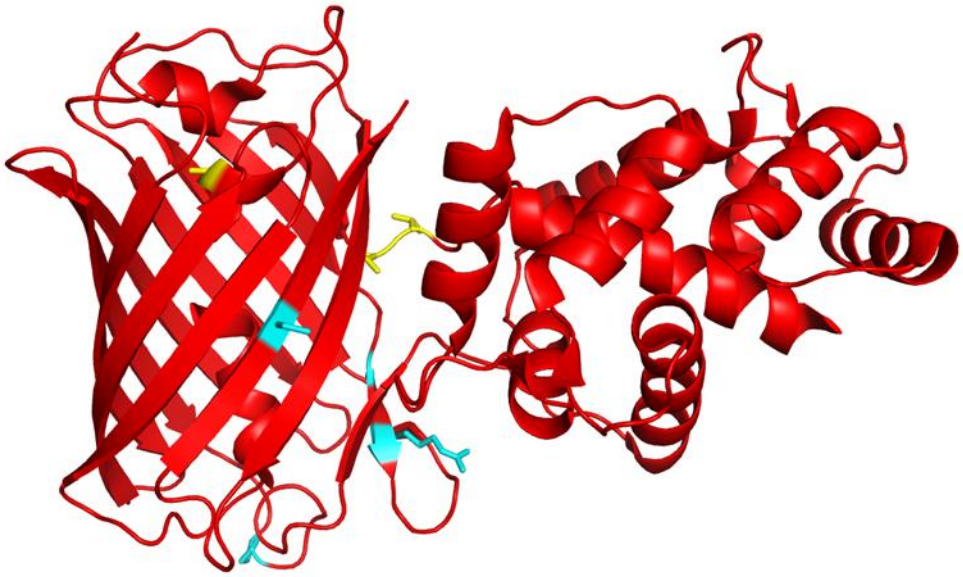


Figure S4.15 Model of RGECO1 used to design rSTEP0. This figure shows the positions of the mutations made to the RFP barrel during the evolution of RGECO1 (mutations made to the calmodulin and M13 domains are not shown). Mutations that were kept in rSTEP0 are coloured in yellow, while those that were reverted to the corresponding mApple amino acids are coloured in teal. Note that two other positions that are not visible in the crystal structure due to lack of electron density were also mutated: M158L and N164D (based on the numbering relative to the mApple-derived analogue of GCamP3). These positions are both on a loop at the top of the barrel, and were reverted in rSTEP0. Figure based on the RGECO1 crystal structure (PDB: 4I2Y)

4.6.5 Multiple Sequence Alignment

```
CLUSTAL O(1.2.4) multiple sequence alignment

rSTEP0      MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRPVVSER-MYPEDGALKSEIKKRL       59
ySTEP0      MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRLENVYITADKQKNGIKANFKIRH       60
cSTEP0      MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRLENVYIKADKQKNGIKANFKIRH       60
gSTEP1      MHHHHHHDLRPEIWIAQELRRIGDEFNAYYARRHKIRLENVYIKADKQKNGIKANFKIRH       60
            ************************************         :...:*:::* *

rSTEP0      KLKDGGHYAAEVK--TTYKAKKPVQLPGAYIVDIKLDIVS-HNEDYTIVEQYERAEGRHS       116
ySTEP0      NIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSYQSKLSKDPNEKRDHMVLLEFVTAAGI       120
cSTEP0      NIEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGI       120
gSTEP1      NIEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLEFVTAAGI       120
            :::***   *     .*  .. ** **. : :. : .: .  **.   :   * . .

rSTEP0      TGGMDELYKGGTGGSMVSKGEENNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEA       176
ySTEP0      TLGMDELYKGGTGGSMVSKGEELFTGV----VPILVELDGDVNGHKFSVSGEGEGDATYG       176
cSTEP0      TLGMDELYKGGTGGSMVSKGEELFTGV----VPILVELDGDVNGHKFSVSGEGEGDATYG       176
gSTEP1      TLGMDELYKGGTGGSMVSKGEELFTGV----VPILVELDGDVNGHKFSVSGEGEGDATYG       176
            * ********************   .:    : : *.::*.****:*.:.*****    .

rSTEP0      FQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYIKHPAD--IPDYFKLSFPEGFRWERVMN       234
ySTEP0      KLTLKL-ICTTGKLPVPWPTLVTTLGYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIF       235
cSTEP0      KLTLKF-ICTTGKLPVPWPTLVTTLTWGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIF       235
gSTEP1      KLTLKF-ICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIF       235
              * *: : . * **. *  *   : :* :.: ::* .    *:** ::***:  **.:

rSTEP0      FEDGGIIHVNQDSSLQDGVFIYKVKLRGTNFPSDGPVMQKKTMGWEATR*                 283
ySTEP0      FKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEY*-----                 279
cSTEP0      FKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEY*-----                 279
gSTEP1      FKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEY*-----                 279
            *:*.*  :.. : .::...:: :::*:* :* .** :: :*
```

Figure S4.16 Multiple sequence alignment of the different colours of STEP. A multiple sequence alignment was generated using Clustal Omega. Characters used to denote the degree of conservation are: * positions that have a single fully conserved residue, : positions that have conservation between groups of strongly similar properties, . positions that have conservation between groups of weakly similar properties.

Chapter 5: Discussion and Perspectives

5.1 Summary

Fluorescent proteins are indispensable tools for studying biology at the molecular level. Coupled with advances in microscopy techniques, their use as fusion tags now allows proteins to be tracked at the single-molecule level,[233-235] and they form the basis for a number of biosensors,[79, 236, 237] measuring parameters including calcium ion levels,[86, 207] voltage across a membrane,[238] kinase activity,[239] and protease activity.[240] Yet before any fluorescence can be detected, the proteins must first mature to form the chromophore. This introduces a delay of tens of minutes between expression of a fluorescent protein and the detection of fluorescence,[92] during which time no information can be obtained about its localization or mobility. In order to avoid this problem, and allow detection of an expressed protein within seconds after it has folded, we have developed a Sensor for Transiently Expressed Proteins, or STEP, which is a two part biosensor. The first part consists of a circularly permuted fluorescent protein tagged with a peptide, arranged such that the chromophore matures fully, but is adjacent to a pore in the β-barrel structure, resulting in quenched, dim fluorescence. This part is named xSTEP, where x is the colour of fluorescent protein used. The second part consists of a protein tag, named STEPtag, which can be fused to a protein of interest. STEPtag binds specifically to the peptide on the xSTEP, restoring bright fluorescence, resulting in a signal that can be visualised on the timescale of binding, rather than maturation. This thesis presents the initial development and testing of the STEP, using a circularly permuted green fluorescent protein with a Bim peptide as the gSTEP, and a truncated Bcl-x_L protein as the STEPtag. A series of rational improvements were introduced, mainly based around the pore and the linker between the Bim peptide and the cpGFP, as these regions have been shown to be important in the GCaMP family of calcium ion sensors upon which we based our design for the STEP.[11, 82, 86] These designs, coupled with the FACS-based screening of a combinatorial saturation library, led us to our current best sensor, known as gSTEP1. We characterized the

equilibrium binding of gSTEP1 to STEPtag, as well as the binding kinetics *in vitro*, confirming that binding occurs on the order of seconds, with a K_d of 120 ± 30 nM, and showing an increase in fluorescence ($\Delta F/F_0$) of 3.4 ± 0.5 times the baseline signal. We also demonstrated that the induction of STEPtag expression in *E. coli* cells leads to an increase in the green fluorescence of gSTEP1, and that this increase in fluorescence can be detected roughly 10 minutes before the fluorescence of induced EGFP can be seen. Finally, we attempted to create orthogonal STEPs that could be used for multiplexing, by changing both the binding pairs and colours of fluorescent protein used. In these experiments, the binding pairs tested either did not perform as well as gSTEP1 and STEPtag, such as the PDZ-based tags tested, or were not orthogonal, such as the gSTEP0.a/b sensors, based on different peptides homologous to Bim, which both bound STEPtag.a/b tags. The different colours of the sensors also require optimization, as the red and cyan versions showed effectively no binding, although the yellow sensor, ySTEP0, performs reasonably. Thus, although we do not currently have multiple orthogonal STEPs available, gSTEP1 is functional both *in vitro* and *in cellulo*, and scaffolds have been prepared to begin optimizing the other colours and binding partners.

5.2 Future Directions

5.2.1 Improving the STEP by Rational Design and Directed Evolution

While gSTEP1 is functional, there is still room to improve it. A greater change in fluorescence could be beneficial, as many biosensors show a $\Delta F/F_0$ of 10 or more,[84, 238, 241] and variants with different affinity could also be useful to detect proteins at different concentrations in the cell, in particular higher affinity sensors with a lower K_d, to ensure that even proteins of interest that are expressed at below average concentrations can be detected.[209, 212, 213] While the detection of fluorophores at sub-micromolar concentrations can be challenging, it is possible, for instance by using total internal reflection fluorescence microscopy.[235]

In order to facilitate the rational design of the STEP, we have attempted to obtain crystal structures of a number of constructs, including uni-gSTEP0, gSTEP0-T1, gSTEP1, ySTEP0, rSTEP0, and a unimolecular version of gSTEP1 with STEPtag fused to its C-terminus. To date, none of these constructs have yielded diffractable crystals, so attempts are ongoing. High-resolution structures have proven invaluable in the understanding and engineering of improved fluorescent proteins and GCaMPs.[11, 22, 42, 50, 84] Ideally, crystal structures of the STEPs would be used to identify the interaction surface between STEPtag and the barrel of the fluorescent protein, which could then be targeted for the rational design of interactions that would increase affinity, for instance by introducing pairs of hydrogen-bonding residues.

For rSTEP in particular, crystal structures could aid in understanding why no change in fluorescence is seen, as it has been shown that RCaMP and R-GECO red fluorescent calcium ion sensors have different orientations of the fluorescent protein and the bound peptide-protein complex, leading to different residues interacting with the chromophore. If a similar case is seen with rSTEP, it could be possible that binding is occurring, but that it does not influence any of the residues interacting with the chromophore. An alternative explanation for not seeing a change when STEPtag is added would simply be that no binding is occurring. To confirm whether binding occurs independently of a fluorescent signal, we could use a biophysical method to detect protein-protein binding, such as isothermal titration calorimetry (ITC),[242] or surface plasmon resonance (SPR).[243] If binding occurs, then an experiment similar to the linker saturation library used to identify gSTEP1 could be used to identify combinations of residues that would interact with the chromophore and change the fluorescence.

If the crystal structures continue to be elusive, we can also attempt to improve the STEP through directed evolution, initially by randomly mutating both gSTEP1 and STEPtag independently. We have already shown that FACS can be used to detect bright mutants of gSTEP, and our flow

cytometry results suggest that the pZA23/pBAD expression system should properly express both the on state and off state of the sensor, allowing us to perform the two-step sort that was attempted with the linker saturation library, to screen for mutants with an increase in $\Delta F/F_0$. With this expression system, the cloning step to remove STEPtag from the Duet vector will be unnecessary, so we will simply begin by sorting the cells with baseline gSTEP expression from pZA23, to select the population that has an unbound, baseline signal similar to that of the parent. The selected cells will then be returned to growth media, transcription of STEPtag will be induced with arabinose, and the cells in the on state will be sorted again, this time selecting for the brightest population of cells, as the sensor should be in its bound, bright state. Unlike the previous attempt, this iteration of the two-step screen will test the on state as part of the final sort, thereby confirming that any mutants picked for the plate-based testing will be fluorescent. This should prevent the reoccurrence of what was seen with the negative population in the original sort, where cells that we hoped would contain low-baseline mutants ended up simply being nonfluorescent. Alternatively, if the FACS screen continues to have difficulty detecting mutants with improved sensor activity, we could simply perform a positive sort to select bright mutants, and apply our 96-well plate-based bacterial kinetic assay (Chapter 3) to identify mutants that rapidly increase in fluorescence when STEPtag is added. In either case, a two-step screening method should allow us to identify mutants that are not just bright in the bound state, but that also maintain their dim off state, resulting in gSTEP1 mutants with improved $\Delta F/F_0$.

Additionally, the same system could be used to screen for mutants of gSTEP1 or STEPtag with improved K_d. As the level of induction of the P_{BAD} promoter can be modulated by the concentration of arabinose added,[153] the resulting concentration of STEPtag produced can be changed. By inducing gSTEP1 as normal, and STEPtag with lower concentrations of arabinose, the FACS screen for fluorescence can be made more stringent for tight-binding interactions, as cells where gSTEP1 is saturated with STEPtag will be brighter than those where gSTEP1 is only partially bound. Due to leaky

expression from the pBAD vector, there is a lower limit to the concentration of STEPtag produced in the cell, as even the fully repressed vector was found to produce roughly 60 molecules/cell, while the fully induced vector produced roughly 20,000 molecules/cell.[153] By approximating the volume of an *E. coli* cell to 1 μm^3,[211, 244] this leads to an approximate concentration of STEPtag of 100 nM, even when pBAD is fully repressed. Thus, this method will have a lower limit of K_d discrimination around 100 nM, where the cells both before and after induction of STEPtag will show similar levels of gSTEP1 fluorescence. This should still allow some improvement in K_d to be detected, and it could be possible to further lower the baseline STEPtag expression by using a lower copy number plasmid, as it has been found that the copy number of the plasmid affects the number of protein molecules produced in the repressed state.[223] We have therefore identified methods to further improve the STEP, by rational design if crystal structures can be obtained, or by directed evolution if not.

In general, the process of developing the STEP has demonstrated the interconnectivity of the individual structural components that make up the biosensor. This is most apparent in the modifications made to the linker length and peptide sequence as described in Chapter 3, where the combination of the best linker and binding partner led to a mutant with a chromophore environment that favoured the neutral form of the chromophore, unlike what was observed in the two parents. This epistasis means that we cannot treat a protein as a collection of independent, modular domains, and thus a more holistic approach to design would be beneficial, where all aspects are considered simultaneously. Practically, this may not be entirely feasible, due to the combinatorial increase in library size with each introduced factor, but minimally it means that positions that were optimized in the context of one STEP scaffold may need to be revisited as the structure of the biosensor changes. For instance, a linker that was optimized in the context of a GFP-based sensor is unlikely to be ideal for the RFP-based version, and vice-versa. On a positive note, the development of the STEP has also shown that lessons learned in the development of the GECIs can be applied to other biosensors based on similar scaffolds. Aspects such

as linker sequence and length that have been shown to affect fluorescence in prior biosensors also did so in the STEP, so targeting these same locations when developing new cpFP-based biosensors is likely to bear fruit.

5.2.2 Applying the STEP: Transiently Expressed Proteins and Beyond

The overall goal of the STEP remains the detection of transiently expressed proteins. One of the original targets for the project was tracking the establishment of the Bicoid gradient in *Drosophila*,[245] where Little *et al.* noticed that the maturation time of their fluorescent protein tags could have an effect on their measured concentration gradient.[10] This target remains one of the reasons why higher affinity versions of the STEP are desired, as the Bicoid gradient ranges from roughly 55 nM to 0 nM in the nucleus, based on the position along the anterior-posterior axis.[246] It is therefore unlikely that much signal will be obtained from gSTEP1, with a K_d of 120 nM, if it is expressed alongside a Bicoid-STEPtag fusion. Engineering a STEP with a K_d in the low nanomolar range is thus a challenge that remains for this application.

Yet the Bicoid gradient is not the only example of a transient cellular process. Another potential use of the STEP would be the real-time tracking of the vertebrate somitogenesis clock, an oscillatory system that dictates the segmentation of the embryo as it grows.[247] In zebrafish, this oscillator has been found to have a period of 25-30 minutes,[248, 249] and the Her1 and Her7 regulatory proteins involved were found to degrade rapidly, on the 10-minute scale.[250] Thus, these proteins cannot be tracked in real-time as GFP fusions, as some of the proteins will have degraded before the fusion tag has fully matured, whereas the STEP will not be limited by this factor. In fact, any study interested in comparing transcriptional and translational dynamics could potentially be an application for the STEP, such as the correlation of mRNA to protein levels in the cell,[9] or characterization of promoter kinetics.[92] As we discussed in the Introduction (Section 1.1.4), the delay required for maturation can introduce errors when determining protein concentration shortly after translation, thereby hiding information that can

be important to modelling biological processes. By reducing this delay from the minute timescale to the second timescale, the STEP can improve the temporal resolution of the data acquired.

Another potential use for the STEP would be to track the mixing of two cellular compartments. If a population of gSTEP and STEPtag are separated by a membrane, membrane degradation could be visualized by the increase in fluorescence as the two halves of the STEP mix. Tracking this increase in signal could be simpler than following a fluorescent protein from one compartment to the other, as this would be observed as a dilution of the concentration of fluorescent proteins, and thus a loss in signal. The STEP could therefore be used for applications such visualizing extracellular vesicle trafficking between different cells,[251] or tracking the release of infectious bacteria into the cytoplasm of host cells,[252] where small compartments containing STEPtag could be visualized as they mix with the cytoplasm of gSTEP-containing host cells. These are examples of applications suggested to us by collaborators after hearing about the STEP, and we were excited to discover that the rapid binding kinetics of the sensor can potentially solve other research problems, outside of our initial goal of visualizing proteins immediately post-translation.

In each of these applications, it will be important to perform control experiments to verify cellular localization of the STEP and STEPtag. As the binding pair used for molecular recognition in gSTEP1 is the human Bim peptide and a truncated version of the human Bcl-xL protein, it is possible that binding of sensor components to endogenous equivalents of Bim and Bcl-xL will occur when the sensor is expressed in human cells. Although we have removed the membrane-binding domains from these proteins (see Chapter 1.2) to reduce the likelihood that STEP expression will affect the cellular apoptosis pathway, it will be important to verify that this is the case, and ensure that the STEP components do not localize to the mitochondrial membranes as a result of off-target binding.

5.3 Perspective on the STEP: Adding to the Toolbox of Fluorescent Protein-based Biosensors

The general concept of a fluorescent reporter that is separate from the protein of interest, of which the STEP is an example, has been used to solve a number of problems in biological imaging, and each system comes with advantages and disadvantages.[253] For example, in cases where the size of the fluorescent protein label (≈27 kDa) perturbs the activity or expression of the POI, peptide motifs that bind to fluorophores have been used, either by direct labelling of the peptide with a small molecule fluorophore, as in the FLASH labelling developed by the Tsien lab,[254] and the more recent FlARe technique,[255] or by binding the peptide with a fluorescently labelled protein, such as a nanobody,[98] or the engineered covalently-binding Spycatcher protein.[256] The major disadvantage to these methods is that the fluorophores are not genetically encodable, and thus must be synthesized and injected into the organism of interest.

In order to maintain the genetically encoded fluorophore associated with fluorescent proteins while using a small peptide tag, a number of split fluorescent proteins have been created, where one or more of the β-strands of the FP barrel are used as a peptide tag, and the remainder of the fluorescent protein is expressed separately. Examples of these systems include the split GFP system based on superfolder GFP missing strands 10 and 11, which uses those missing strands as the peptide tag or tags,[257-259] as well as the leave-one-out GFPs,[260, 261] where each possible construct missing a single secondary structure element, including all of the β-strands and the α-helix, were tested for fluorescence when the missing element was added, to varying degrees of success. The split-GFP constructs in particular are widely used to detect protein-protein interactions and test the solubility of tagged proteins, but they are not suitable for the detection of transiently expressed proteins, as it has been found that the chromophore maturation process does not begin until after complementation of the protein, leading to the same delay as if a GFP fusion was used.

With that delay in mind, there are also tags that have been specifically developed to detect protein translation. *Gaussia* luciferase has been used as a tag that generates a luminescent signal within seconds of protein translation, however it requires the addition of exogenous coelenterazine as a substrate,[262] which poses additional challenges for use in live animal embryos. Furthermore, it exhibits flash kinetics that rapidly decay, preventing continued tracking of the synthesized protein. The SunTag[263] and MoonTag[264] systems consist of peptide sequences that can be bound by fluorescent protein-tagged antibodies. By binding up to 24 copies of the fluorescent protein tag, the nascent protein can be detected by the concentration of fluorescent signal. As with the STEP, this system allows the fluorescent protein to mature before the protein of interest is expressed, but it is a very large tag (1400 kDa once 24 antibody-GFP fusions have bound), which has the potential to interfere with activity and mobility of the POI. Other protein expression sensors rely on the translocation of mature fluorescent proteins to provide the signal. The dynamic protein synthesis translocation reporter (dPSTR) is one such sensor,[265] where a constitutively expressed fluorescent protein is attached to a synthetic SynZip coiled-coil peptide,[107] and freely diffuses between the cytoplasm and the nucleus. Another construct consisting of the cognate SynZip and nuclear localization signals is placed under the control of the promoter driving the process of interest. Translation occurring from the promoter is detected by concentration of the fluorescent protein signal in the nucleus. This is therefore an indirect method to detect protein translation, as the protein of interest remains untagged, and so localization information needs to be obtained by another method. Bothma *et al.*[266] realized that such a method could be used to directly detect the POI in the specific case of studying transcription factors, which already localize to the nucleus. They developed the LlamaTag system, where the transcription factor to be studied is fused with a nanobody that binds GFP.[267] EGFP expressed in advance can therefore be bound by the transcription factor once it is expressed, and will be imported to the nucleus as part of its native activity, allowing cells with active translation of the transcription factor to be detected by nuclear concentration

of the fluorescent signal. This is an effective, direct system with rapid binding kinetics, but is limited to proteins of interest that localize outside of the cytosol.

As was discussed in Chapter 2, the existing sensors most resembling the STEP are the dimerization-dependent fluorescent proteins,[217, 218] and the Flashbody.[219] Each of these systems consist of a quenched, mature fluorescent protein, which increases in signal upon binding of a protein-based tag, either a fluorescent protein without a chromophore, or a peptide antigen. Thus, the signal-to-noise ratio and affinity are the main distinctions between these sensors. The dimerization-dependent fluorescent proteins have excellent signal-to-noise ratios, with 12- to 60-fold changes in fluorescence intensity, depending on the colour of sensor used, but were designed to have K_d values in the micromolar range, and thus would not be suitable for detecting proteins at low concentrations. The Flashbody has a lower K_d of 400 nM, with a $\Delta F/F_0$ of 3, which is comparable to gSTEP1, with its K_d of 120 ± 30 nM and $\Delta F/F_0$ of 3.4 ± 0.5. The k_{on} of the Flashbody, interestingly, was measured to be 3.4 x 10^3 M^{-1} s^{-1}, considerably slower than the 1.7 x 10^5 $M^{-1}s^{-1}$ measured for gSTEP1. Thus, the point where the STEP currently shines is in its high affinity for STEPtag, allowing it to detect targets at a lower concentration, as well as its rapid on-rate, allowing it to detect the target POI with as little delay as possible.

Thus, a niche is available for the STEP to provide new insight and understanding into protein expression in living cells. With the advent of CRISPR/Cas9 genome editing and the creation of a synthetic bacterial genome,[268] the tools required to edit DNA sequences in living organisms are more accessible than ever before. In order to wisely use these tools, our understanding of how the DNA sequences dictate the transcriptional and translational patterns required to maintain a whole organism needs to keep pace with them. By expanding the toolbox of biosensors, we hope that the STEP will prove of some use in this goal, by providing spatial and temporal information about protein translation at a timescale that standard fluorescent protein tags cannot reach.

References

1. Lambert, T. J., FPbase: a community-editable fluorescent protein database. *Nat Methods* **2019,** *16* (4), 277-278.
2. Givan, A. L., *Flow Cytometry: First Principles*. 2 ed.; Wiley-Liss: New York, 2001; p 59-80.
3. Grigorenko, B. L.; Krylov, A. I.; Nemukhin, A. V., Molecular Modeling Clarifies the Mechanism of Chromophore Maturation in the Green Fluorescent Protein. *J Am Chem Soc* **2017,** *139* (30), 10239-10249.
4. Lee, E. F.; Fairlie, W. D., The Structural Biology of Bcl-xL. *Int J Mol Sci* **2019,** *20* (9), 1-18.
5. Hulme, E. C.; Trevethick, M. A., Ligand binding assays at equilibrium: validation and interpretation. *Br J Pharmacol* **2010,** *161* (6), 1219-37.
6. Hochreiter, B.; Garcia, A. P.; Schmid, J. A., Fluorescent proteins as genetically encoded FRET biosensors in life sciences. *Sensors (Basel)* **2015,** *15* (10), 26281-314.
7. Gross, L. A.; Baird, G. S.; Hoffman, R. C.; Baldridge, K. K.; Tsien, R. Y., The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc Natl Acad Sci U S A* **2000,** *97* (22), 11990-5.
8. Novick, R. P., Plasmid incompatibility. *Microbiol Rev* **1987,** *51* (4), 381-395.
9. Taniguchi, Y.; Choi, P. J.; Li, G. W.; Chen, H.; Babu, M.; Hearn, J.; Emili, A.; Xie, X. S., Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **2010,** *329* (5991), 533-8.
10. Little, S. C.; Tkačik, G.; Kneeland, T. B.; Wieschaus, E. F.; Gregor, T., The formation of the Bicoid morphogen gradient requires protein movement from anteriorly localized mRNA. *PLoS Biol* **2011,** *9* (3), e1000596.
11. Akerboom, J.; Rivera, J. D.; Guilbe, M. M.; Malavé, E. C.; Hernandez, H. H.; Tian, L.; Hires, S. A.; Marvin, J. S.; Looger, L. L.; Schreiter, E. R., Crystal structures of the GCaMP calcium sensor reveal the mechanism of fluorescence signal change and aid rational design. *J Biol Chem* **2009,** *284* (10), 6455-64.
12. Gedeon, T.; Bokes, P., Delayed protein synthesis reduces the correlation between mRNA and protein fluctuations. *Biophys J* **2012,** *103* (3), 377-85.
13. Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*. 3 ed.; Springer: New York, 2006; p 1-24.
14. Stokes, G. G., XVI. On the change of refrangibility of light. —No. II. *Phil. Trans. R. Soc.* **1853,** *143*.
15. Braslavsky, S., Glossary of terms used in photochemistry, 3rd edition (IUPAC Recommendations 2006). *Pure Appl Chem* **2009,** *79* (3), 293-465.
16. Englman, R.; Jortner, J., The energy gap law for radiationless transitions in large molecules. *Mol Phys* **1970,** *18* (2), 145-164.
17. Dobretsov, G. E.; Syrejschikova, T. I.; Smolina, N. V., On mechanisms of fluorescence quenching by water. *Biophysics* **2014,** *59*, 183-188.
18. Majumdar, D. K.; Basu, S., Charge Transfer and Fluorescence Quenching. *J Chem Phys* **1960,** *33*, 1199-201.
19. Hawe, A.; Sutter, M.; Jiskoot, W., Extrinsic fluorescent dyes as tools for protein characterization. *Pharm Res* **2008,** *25* (7), 1487-99.
20. Shimomura, O.; Johnson, F. H.; Saiga, Y., Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J Cell Comp Physiol* **1962,** *59*, 223-39.
21. Johnson, F. H.; Shimomura, O.; Saiga, Y.; Gershman, L. C.; Reynolds, G. T.; Waters, J. R., Quantum efficiency of Cypridina luminescence, with a note on that of Aequorea. *J Cell Comp Physiol* **1962,** *60* (1), 85-103.
22. Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J., Crystal structure of the Aequorea victoria green fluorescent protein. *Science* **1996,** *273* (5280), 1392-5.

23. Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormö, M.; Remington, S. J., Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. *Proc Natl Acad Sci U S A* **1997,** *94* (6), 2306-11.
24. Prasher, D. C.; Eckenrode, V. K.; Ward, W. W.; Prendergast, F. G.; Cormier, M. J., Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* **1992,** *111* (2), 229-33.
25. Heim, R.; Prasher, D. C.; Tsien, R. Y., Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci U S A* **1994,** *91* (26), 12501-4.
26. Barondeau, D. P.; Putnam, C. D.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D., Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proc Natl Acad Sci U S A* **2003,** *100* (21), 12111-6.
27. Wood, T. I.; Barondeau, D. P.; Hitomi, C.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D., Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis. *Biochemistry* **2005,** *44* (49), 16211-20.
28. Sniegowski, J. A.; Lappe, J. W.; Patel, H. N.; Huffman, H. A.; Wachter, R. M., Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein. *J Biol Chem* **2005,** *280* (28), 26248-55.
29. Niwa, H.; Inouye, S.; Hirano, T.; Matsuno, T.; Kojima, S.; Kubota, M.; Ohashi, M.; Tsuji, F. I., Chemical nature of the light emitter of the Aequorea green fluorescent protein. *Proc Natl Acad Sci U S A* **1996,** *93* (24), 13617-22.
30. Altoe', P.; Bernardi, F.; Garavelli, M.; Orlandi, G.; Negri, F., Solvent effects on the vibrational activity and photodynamics of the green fluorescent protein chromophore: a quantum-chemical study. *J Am Chem Soc* **2005,** *127* (11), 3952-63.
31. Labas, Y. A.; Gurskaya, N. G.; Yanushevich, Y. G.; Fradkov, A. F.; Lukyanov, K. A.; Lukyanov, S. A.; Matz, M. V., Diversity and evolution of the green fluorescent protein family. *Proc Natl Acad Sci U S A* **2002,** *99* (7), 4256-61.
32. Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A., Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat Biotechnol* **1999,** *17* (10), 969-73.
33. Wiedenmann, J.; Schenk, A.; Röcker, C.; Girod, A.; Spindler, K. D.; Nienhaus, G. U., A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from Entacmaea quadricolor (Anthozoa, Actinaria). *Proc Natl Acad Sci U S A* **2002,** *99* (18), 11646-51.
34. Merzlyak, E. M.; Goedhart, J.; Shcherbo, D.; Bulina, M. E.; Shcheglov, A. S.; Fradkov, A. F.; Gaintzeva, A.; Lukyanov, K. A.; Lukyanov, S.; Gadella, T. W.; Chudakov, D. M., Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat Methods* **2007,** *4* (7), 555-7.
35. Shagin, D. A.; Barsova, E. V.; Yanushevich, Y. G.; Fradkov, A. F.; Lukyanov, K. A.; Labas, Y. A.; Semenova, T. N.; Ugalde, J. A.; Meyers, A.; Nunez, J. M.; Widder, E. A.; Lukyanov, S. A.; Matz, M. V., GFP-like proteins as ubiquitous metazoan superfamily: evolution of functional features and structural complexity. *Mol Biol Evol* **2004,** *21* (5), 841-50.
36. Hunt, M. E.; Scherrer, M. P.; Ferrari, F. D.; Matz, M. V., Very bright green fluorescent proteins from the Pontellid copepod Pontella mimocerami. *PLoS One* **2010,** *5* (7), e11517.
37. Masuda, H.; Takenaka, Y.; Yamaguchi, A.; Nishikawa, S.; Mizuno, H., A novel yellowish-green fluorescent protein from the marine copepod, Chiridius poppei, and its use as a reporter protein in HeLa cells. *Gene* **2006,** *372*, 18-25.
38. Deheyn, D. D.; Kubokawa, K.; McCarthy, J. K.; Murakami, A.; Porrachia, M.; Rouse, G. W.; Holland, N. D., Endogenous green fluorescent protein (GFP) in amphioxus. *Biol Bull* **2007,** *213* (2), 95-100.
39. Bomati, E. K.; Manning, G.; Deheyn, D. D., Amphioxus encodes the largest known family of green fluorescent proteins, which have diversified into distinct functional classes. *BMC Evol Biol* **2009,** *9*, 77.

40. Tomosugi, W.; Matsuda, T.; Tani, T.; Nemoto, T.; Kotera, I.; Saito, K.; Horikawa, K.; Nagai, T., An ultramarine fluorescent protein with increased photostability and pH insensitivity. *Nat Methods* **2009,** *6* (5), 351-3.
41. Ai, H. W.; Shaner, N. C.; Cheng, Z.; Tsien, R. Y.; Campbell, R. E., Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. *Biochemistry* **2007,** *46* (20), 5904-10.
42. Goedhart, J.; von Stetten, D.; Noirclerc-Savoye, M.; Lelimousin, M.; Joosen, L.; Hink, M. A.; van Weeren, L.; Gadella, T. W.; Royant, A., Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat Commun* **2012,** *3*, 751.
43. Zhang, G.; Gurtu, V.; Kain, S. R., An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. *Biochem Biophys Res Commun* **1996,** *227* (3), 707-11.
44. Nagai, T.; Ibata, K.; Park, E. S.; Kubota, M.; Mikoshiba, K.; Miyawaki, A., A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol* **2002,** *20* (1), 87-90.
45. Shaner, N. C.; Lin, M. Z.; McKeown, M. R.; Steinbach, P. A.; Hazelwood, K. L.; Davidson, M. W.; Tsien, R. Y., Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat Methods* **2008,** *5* (6), 545-51.
46. Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N.; Palmer, A. E.; Tsien, R. Y., Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* **2004,** *22* (12), 1567-72.
47. Li, Z.; Zhang, Z.; Bi, L.; Cui, Z.; Deng, J.; Wang, D.; Zhang, X. E., Mutagenesis of mNeptune Red-Shifts Emission Spectrum to 681-685 nm. *PLoS One* **2016,** *11* (4), e0148749.
48. Cubitt, A. B.; Woollenweber, L. A.; Heim, R., Understanding structure-function relationships in the Aequorea victoria green fluorescent protein. *Methods Cell Biol* **1999,** *58*, 19-30.
49. Strack, R. L.; Strongin, D. E.; Mets, L.; Glick, B. S.; Keenan, R. J., Chromophore formation in DsRed occurs by a branched pathway. *J Am Chem Soc* **2010,** *132* (24), 8496-505.
50. Campbell, R. E.; Tour, O.; Palmer, A. E.; Steinbach, P. A.; Baird, G. S.; Zacharias, D. A.; Tsien, R. Y., A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* **2002,** *99* (12), 7877-82.
51. Bindels, D. S.; Haarbosch, L.; van Weeren, L.; Postma, M.; Wiese, K. E.; Mastop, M.; Aumonier, S.; Gotthard, G.; Royant, A.; Hink, M. A.; Gadella, T. W. J., mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat Methods* **2017,** *14* (1), 53-56.
52. Chudakov, D. M.; Matz, M. V.; Lukyanov, S.; Lukyanov, K. A., Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol Rev* **2010,** *90* (3), 1103-63.
53. Tsien, R. Y., The green fluorescent protein. *Annu Rev Biochem* **1998,** *67*, 509-44.
54. Rizzo, M. A.; Davidson, M. W.; Piston, D. W., Fluorescent protein tracking and detection: fluorescent protein structure and color variants. *Cold Spring Harb Protoc* **2009,** *4* (12), 1-21.
55. Costantini, L. M.; Snapp, E. L., Fluorescent proteins in cellular organelles: serious pitfalls and some solutions. *DNA Cell Biol* **2013,** *32* (11), 622-7.
56. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C., Green fluorescent protein as a marker for gene expression. *Science* **1994,** *263* (5148), 802-5.
57. Wang, S.; Hazelrigg, T., Implications for bcd mRNA localization from spatial distribution of exu protein in Drosophila oogenesis. *Nature* **1994,** *369* (6479), 400-03.
58. Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y., Understanding, improving and using green fluorescent proteins. *Trends Biochem Sci* **1995,** *20* (11), 448-55.
59. Cormack, B. P.; Valdivia, R. H.; Falkow, S., FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **1996,** *173* (1 Spec No), 33-8.
60. Huh, W. K.; Falvo, J. V.; Gerke, L. C.; Carroll, A. S.; Howson, R. W.; Weissman, J. S.; O'Shea, E. K., Global analysis of protein localization in budding yeast. *Nature* **2003,** *425* (6959), 686-91.
61. Harikumar, A.; Edupuganti, R. R.; Sorek, M.; Azad, G. K.; Markoulaki, S.; Sehnalová, P.; Legartová, S.; Bártová, E.; Farkash-Amar, S.; Jaenisch, R.; Alon, U.; Meshorer, E., An

Endogenously Tagged Fluorescent Fusion Protein Library in Mouse Embryonic Stem Cells. *Stem Cell Rep* **2017,** *9* (4), 1304-1314.
62. Cohen, A. A.; Geva-Zatorsky, N.; Eden, E.; Frenkel-Morgenstern, M.; Issaeva, I.; Sigal, A.; Milo, R.; Cohen-Saidon, C.; Liron, Y.; Kam, Z.; Cohen, L.; Danon, T.; Perzov, N.; Alon, U., Dynamic proteomics of individual cancer cells in response to a drug. *Science* **2008,** *322* (5907), 1511-6.
63. Frenkel-Morgenstern, M.; Cohen, A. A.; Geva-Zatorsky, N.; Eden, E.; Prilusky, J.; Issaeva, I.; Sigal, A.; Cohen-Saidon, C.; Liron, Y.; Cohen, L.; Danon, T.; Perzov, N.; Alon, U., Dynamic Proteomics: a database for dynamics and localizations of endogenous fluorescently-tagged proteins in living human cells. *Nucleic Acids Res* **2010,** *38* (Database issue), D508-12.
64. Sarov, M.; Barz, C.; Jambor, H.; Hein, M. Y.; Schmied, C.; Suchold, D.; Stender, B.; Janosch, S.; K J, V. V.; Krishnan, R. T.; Krishnamoorthy, A.; Ferreira, I. R.; Ejsmont, R. K.; Finkl, K.; Hasse, S.; Kämpfer, P.; Plewka, N.; Vinis, E.; Schloissnig, S.; Knust, E.; Hartenstein, V.; Mann, M.; Ramaswami, M.; VijayRaghavan, K.; Tomancak, P.; Schnorrer, F., A genome-wide resource for the analysis of protein localisation in Drosophila. *Elife* **2016,** *5*, e12068.
65. Hoffman, R. M., The multiple uses of fluorescent proteins to visualize cancer in vivo. *Nat Rev Cancer* **2005,** *5* (10), 796-806.
66. Hoffman, R. M., Recent advances on in vivo imaging with fluorescent proteins. *Methods Cell Biol* **2008,** *85*, 485-95.
67. Livet, J.; Weissman, T. A.; Kang, H.; Draft, R. W.; Lu, J.; Bennis, R. A.; Sanes, J. R.; Lichtman, J. W., Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **2007,** *450* (7166), 56-62.
68. Weissman, T. A.; Pan, Y. A., Brainbow: new resources and emerging biological applications for multicolor genetic labeling and analysis. *Genetics* **2015,** *199* (2), 293-306.
69. Yu, Y. A.; Szalay, A. A.; Wang, G.; Oberg, K., Visualization of molecular and cellular events with green fluorescent proteins in developing embryos: a review. *Luminescence* **2003,** *18* (1), 1-18.
70. Zenker, J.; White, M. D.; Gasnier, M.; Alvarez, Y. D.; Lim, H. Y. G.; Bissiere, S.; Biro, M.; Plachta, N., Expanding Actin Rings Zipper the Mouse Embryo for Blastocyst Formation. *Cell* **2018,** *173* (3), 776-791.e17.
71. Witte, R.; Andriasyan, V.; Georgi, F.; Yakimovich, A.; Greber, U. F., Concepts in Light Microscopy of Viruses. *Viruses* **2018,** *10* (4), 202.
72. Costantini, L. M.; Snapp, E. L., Going Viral with Fluorescent Proteins. *J Virol* **2015,** *89* (19), 9706-8.
73. Xia, T.; Li, N.; Fang, X., Single-molecule fluorescence imaging in living cells. *Annu Rev Phys Chem* **2013,** *64*, 459-80.
74. Mishin, A. S.; Belousov, V. V.; Solntsev, K. M.; Lukyanov, K. A., Novel uses of fluorescent proteins. *Curr Opin Chem Biol* **2015,** *27*, 1-9.
75. Wiedenmann, J.; Oswald, F.; Nienhaus, G. U., Fluorescent proteins for live cell imaging: opportunities, limitations, and challenges. *IUBMB Life* **2009,** *61* (11), 1029-42.
76. Heim, R.; Tsien, R. Y., Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol* **1996,** *6* (2), 178-82.
77. Mitra, R. D.; Silva, C. M.; Youvan, D. C., Fluorescence resonance energy transfer between blue-emitting and red-shifted excitation derivatives of the green fluorescent protein. *Gene* **1996,** *173* (1 Spec No), 13-7.
78. Greenwald, E. C.; Mehta, S.; Zhang, J., Genetically Encoded Fluorescent Biosensors Illuminate the Spatiotemporal Regulation of Signaling Networks. *Chem Rev* **2018,** *118* (24), 11707-11794.
79. Kostyuk, A. I.; Demidovich, A. D.; Kotova, D. A.; Belousov, V. V.; Bilan, D. S., Circularly Permuted Fluorescent Protein-Based Indicators: History, Principles, and Classification. *Int J Mol Sci* **2019,** *20* (17).
80. Wang, H.; Nakata, E.; Hamachi, I., Recent Progress in Strategies for the Creation of Protein-Based Fluorescent Biosensors. *ChemBioChem* **2009,** *10* (16), 2560-2577.

81. Pérez Koldenkova, V.; Nagai, T., Genetically encoded Ca(2+) indicators: properties and evaluation. *Biochim Biophys Acta* **2013,** *1833* (7), 1787-97.
82. Nakai, J.; Ohkura, M.; Imoto, K., A high signal-to-noise Ca(2+) probe composed of a single green fluorescent protein. *Nat Biotechnol* **2001,** *19* (2), 137-41.
83. Wang, Q.; Shui, B.; Kotlikoff, M. I.; Sondermann, H., Structural basis for calcium sensing by GCaMP2. *Structure* **2008,** *16* (12), 1817-27.
84. Akerboom, J.; Carreras Calderón, N.; Tian, L.; Wabnig, S.; Prigge, M.; Tolö, J.; Gordus, A.; Orger, M. B.; Severi, K. E.; Macklin, J. J.; Patel, R.; Pulver, S. R.; Wardill, T. J.; Fischer, E.; Schüler, C.; Chen, T. W.; Sarkisyan, K. S.; Marvin, J. S.; Bargmann, C. I.; Kim, D. S.; Kügler, S.; Lagnado, L.; Hegemann, P.; Gottschalk, A.; Schreiter, E. R.; Looger, L. L., Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. *Front Mol Neurosci* **2013,** *6*, 2.
85. Mütze, J.; Iyer, V.; Macklin, J. J.; Colonell, J.; Karsh, B.; Petrášek, Z.; Schwille, P.; Looger, L. L.; Lavis, L. D.; Harris, T. D., Excitation spectra and brightness optimization of two-photon excited probes. *Biophys J* **2012,** *102* (4), 934-44.
86. Tian, L.; Hires, S. A.; Mao, T.; Huber, D.; Chiappe, M. E.; Chalasani, S. H.; Petreanu, L.; Akerboom, J.; McKinney, S. A.; Schreiter, E. R.; Bargmann, C. I.; Jayaraman, V.; Svoboda, K.; Looger, L. L., Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nat Methods* **2009,** *6* (12), 875-81.
87. Ohkura, M.; Sasaki, T.; Sadakari, J.; Gengyo-Ando, K.; Kagawa-Nagamura, Y.; Kobayashi, C.; Ikegaya, Y.; Nakai, J., Genetically encoded green fluorescent Ca2+ indicators with improved detectability for neuronal Ca2+ signals. *PLoS One* **2012,** *7* (12), e51286.
88. Wu, J.; Liu, L.; Matsuda, T.; Zhao, Y.; Rebane, A.; Drobizhev, M.; Chang, Y. F.; Araki, S.; Arai, Y.; March, K.; Hughes, T. E.; Sagou, K.; Miyata, T.; Nagai, T.; Li, W. H.; Campbell, R. E., Improved orange and red Ca^{2+} indicators and photophysical considerations for optogenetic applications. *ACS Chem Neurosci* **2013,** *4* (6), 963-72.
89. Carlson, H. J.; Campbell, R. E., Mutational analysis of a red fluorescent protein-based calcium ion indicator. *Sensors (Basel)* **2013,** *13* (9), 11507-21.
90. Zhao, Y.; Araki, S.; Wu, J.; Teramoto, T.; Chang, Y. F.; Nakano, M.; Abdelfattah, A. S.; Fujiwara, M.; Ishihara, T.; Nagai, T.; Campbell, R. E., An expanded palette of genetically encoded Ca^{2+} indicators. *Science* **2011,** *333* (6051), 1888-91.
91. Evdokimov, A. G.; Pokross, M. E.; Egorov, N. S.; Zaraisky, A. G.; Yampolsky, I. V.; Merzlyak, E. M.; Shkoporov, A. N.; Sander, I.; Lukyanov, K. A.; Chudakov, D. M., Structural basis for the fast maturation of Arthropoda green fluorescent protein. *EMBO Rep* **2006,** *7* (10), 1006-12.
92. Balleza, E.; Kim, J. M.; Cluzel, P., Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat Methods* **2018,** *15* (1), 47-51.
93. Subach, F. V.; Subach, O. M.; Gundorov, I. S.; Morozova, K. S.; Piatkevich, K. D.; Cuervo, A. M.; Verkhusha, V. V., Monomeric fluorescent timers that change color from blue to red report on cellular trafficking. *Nat Chem Biol* **2009,** *5* (2), 118-26.
94. Iizuka, R.; Yamagishi-Shirasaki, M.; Funatsu, T., Kinetic study of de novo chromophore maturation of fluorescent proteins. *Anal Biochem* **2011,** *414* (2), 173-8.
95. Moore, M. M.; Oteng-Pabi, S. K.; Pandelieva, A. T.; Mayo, S. L.; Chica, R. A., Recovery of red fluorescent protein chromophore maturation deficiency through rational design. *PLoS One* **2012,** *7* (12), e52463.
96. Heim, R.; Cubitt, A. B.; Tsien, R. Y., Improved green fluorescence. *Nature* **1995,** *373* (6516), 663-4.
97. Sanchez, M. I.; Ting, A. Y., Directed evolution improves the catalytic efficiency of TEV protease. *Nat Methods* **2019**, 167-174.
98. Virant, D.; Traenkle, B.; Maier, J.; Kaiser, P. D.; Bodenhöfer, M.; Schmees, C.; Vojnovic, I.; Pisak-Lukáts, B.; Endesfelder, U.; Rothbauer, U., A peptide tag-specific nanobody enables high-quality labeling for dSTORM imaging. *Nat Commun* **2018,** *9* (1), 930.

99. Khan, T.; Kandola, T. S.; Wu, J.; Venkatesan, S.; Ketter, E.; Lange, J. J.; Rodríguez Gama, A.; Box, A.; Unruh, J. R.; Cook, M.; Halfmann, R., Quantifying Nucleation In Vivo Reveals the Physical Basis of Prion-like Phase Behavior. *Mol Cell* **2018,** *71* (1), 155-168.e7.
100. Canty, L.; Hariharan, S.; Liu, Q.; Haney, S. A.; Andrews, D. W., Peak emission wavelength and fluorescence lifetime are coupled in far-red, GFP-like fluorescent proteins. *PLoS One* **2018,** *13* (11), e0208075.
101. Wacker, S. A.; Oswald, F.; Wiedenmann, J.; Knöchel, W., A green to red photoconvertible protein as an analyzing tool for early vertebrate development. *Dev Dynam* **2007,** *236* (2), 473-480.
102. Yu, J.; Xiao, J.; Ren, X.; Lao, K.; Xie, X. S., Probing gene expression in live cells, one protein molecule at a time. *Science* **2006,** *311* (5767), 1600-3.
103. Edfors, F.; Danielsson, F.; Hallström, B. M.; Käll, L.; Lundberg, E.; Pontén, F.; Forsström, B.; Uhlén, M., Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol Syst Biol* **2016,** *12* (10), 883-883.
104. Schwanhäusser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M., Global quantification of mammalian gene expression control. *Nature* **2011,** *473* (7347), 337-342.
105. Romei, M. G.; Boxer, S. G., Split Green Fluorescent Proteins: Scope, Limitations, and Outlook. *Annu Rev Biophys* **2019,** *48*, 19-44.
106. Chen, Z.; Boyken, S. E.; Jia, M.; Busch, F.; Flores-Solis, D.; Bick, M. J.; Lu, P.; VanAernum, Z. L.; Sahasrabuddhe, A.; Langan, R. A.; Bermeo, S.; Brunette, T. J.; Mulligan, V. K.; Carter, L. P.; DiMaio, F.; Sgourakis, N. G.; Wysocki, V. H.; Baker, D., Programmable design of orthogonal protein heterodimers. *Nature* **2019,** *565* (7737), 106-111.
107. Reinke, A. W.; Grant, R. A.; Keating, A. E., A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J Am Chem Soc* **2010,** *132* (17), 6025-31.
108. Smith-Gill, S. J., Protein-protein interactions: structural motifs and molecular recognition. *Curr Opin Biotechnol* **1991,** *2* (4), 568-75.
109. Heyduk, E.; Hickey, R.; Pozzi, N.; Heyduk, T., Peptide ligand-based ELISA reagents for antibody detection. *Anal Biochem* **2018,** *559*, 55-61.
110. Dueber, J. E.; Yeh, B. J.; Chak, K.; Lim, W. A., Reprogramming control of an allosteric signaling switch through modular recombination. *Science* **2003,** *301* (5641), 1904-8.
111. Ernst, A.; Gfeller, D.; Kan, Z.; Seshagiri, S.; Kim, P. M.; Bader, G. D.; Sidhu, S. S., Coevolution of PDZ domain–ligand interactions analyzed by high-throughput phage display and deep sequencing. *Mol Biosyst* **2010,** *6* (10), 1782-1790.
112. Lee, H.-J.; Zheng, J. J., PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun Signal* **2010,** *8*, 8-8.
113. Kaneko, T.; Li, L.; Li, S. S., The SH3 domain--a family of versatile peptide- and protein-recognition module. *Front Biosci* **2008,** *13*, 4938-52.
114. Teyra, J.; Huang, H.; Jain, S.; Guan, X.; Dong, A.; Liu, Y.; Tempel, W.; Min, J.; Tong, Y.; Kim, P. M.; Bader, G. D.; Sidhu, S. S., Comprehensive Analysis of the Human SH3 Domain Family Reveals a Wide Variety of Non-canonical Specificities. *Structure* **2017,** *25* (10), 1598-1610.e3.
115. Macias, M. J.; Wiesner, S.; Sudol, M., WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Letters* **2002,** *513* (1), 30-37.
116. Ingham, R. J.; Colwill, K.; Howard, C.; Dettwiler, S.; Lim, C. S. H.; Yu, J.; Hersi, K.; Raaijmakers, J.; Gish, G.; Mbamalu, G.; Taylor, L.; Yeung, B.; Vassilovski, G.; Amin, M.; Chen, F.; Matskova, L.; Winberg, G.; Ernberg, I.; Linding, R.; O'Donnell, P.; Starostine, A.; Keller, W.; Metalnikov, P.; Stark, C.; Pawson, T., WW domains provide a platform for the assembly of multiprotein networks. *Mol Cell Biol* **2005,** *25* (16), 7092-7106.
117. DeBartolo, J.; Dutta, S.; Reich, L.; Keating, A. E., Predictive Bcl-2 family binding models rooted in experiment or structure. *J Mol Biol* **2012,** *422* (1), 124-144.

118. Huang, D. C.; Strasser, A., BH3-Only proteins-essential initiators of apoptotic cell death. *Cell* **2000,** *103* (6), 839-42.
119. Chen, L.; Willis, S. N.; Wei, A.; Smith, B. J.; Fletcher, J. I.; Hinds, M. G.; Colman, P. M.; Day, C. L.; Adams, J. M.; Huang, D. C., Differential targeting of prosurvival Bcl-2 proteins by their BH3-only ligands allows complementary apoptotic function. *Mol Cell* **2005,** *17* (3), 393-403.
120. Dutta, S.; Ryan, J.; Chen, T. S.; Kougentakis, C.; Letai, A.; Keating, A. E., Potent and specific peptide inhibitors of human pro-survival protein Bcl-xL. *J Mol Biol* **2015,** *427* (6 Pt B), 1241-1253.
121. Adams, J. M.; Cory, S., The Bcl-2 protein family: arbiters of cell survival. *Science* **1998,** *281* (5381), 1322-6.
122. Hardwick, J. M.; Youle, R. J., SnapShot: BCL-2 proteins. *Cell* **2009,** *138* (2), 404, 404.e1.
123. O'Connor, L.; Strasser, A.; O'Reilly, L. A.; Hausmann, G.; Adams, J. M.; Cory, S.; Huang, D. C., Bim: a novel member of the Bcl-2 family that promotes apoptosis. *EMBO J* **1998,** *17* (2), 384-95.
124. Petros, A. M.; Olejniczak, E. T.; Fesik, S. W., Structural biology of the Bcl-2 family of proteins. *Biochim Biophys Acta* **2004,** *1644* (2-3), 83-94.
125. Kale, J.; Osterlund, E. J.; Andrews, D. W., BCL-2 family proteins: changing partners in the dance towards death. *Cell Death Differ* **2018,** *25* (1), 65-80.
126. Chi, X.; Kale, J.; Leber, B.; Andrews, D. W., Regulating cell death at, on, and in membranes. *Biochim Biophys Acta* **2014,** *1843* (9), 2100-13.
127. Muchmore, S. W.; Sattler, M.; Liang, H.; Meadows, R. P.; Harlan, J. E.; Yoon, H. S.; Nettesheim, D.; Chang, B. S.; Thompson, C. B.; Wong, S. L.; Ng, S. L.; Fesik, S. W., X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. *Nature* **1996,** *381* (6580), 335-41.
128. Petros, A. M.; Medek, A.; Nettesheim, D. G.; Kim, D. H.; Yoon, H. S.; Swift, K.; Matayoshi, E. D.; Oltersdorf, T.; Fesik, S. W., Solution structure of the antiapoptotic protein bcl-2. *Proc Natl Acad Sci U S A* **2001,** *98* (6), 3012-3017.
129. Petros, A. M.; Nettesheim, D. G.; Wang, Y.; Olejniczak, E. T.; Meadows, R. P.; Mack, J.; Swift, K.; Matayoshi, E. D.; Zhang, H.; Thompson, C. B.; Fesik, S. W., Rationale for Bcl-xL/Bad peptide complex formation from structure, mutagenesis, and biophysical studies. *Protein Sci* **2000,** *9* (12), 2528-34.
130. Sattler, M.; Liang, H.; Nettesheim, D.; Meadows, R. P.; Harlan, J. E.; Eberstadt, M.; Yoon, H. S.; Shuker, S. B.; Chang, B. S.; Minn, A. J.; Thompson, C. B.; Fesik, S. W., Structure of Bcl-xL-Bak Peptide Complex: Recognition Between Regulators of Apoptosis. *Science* **1997,** *275* (5302), 983.
131. Puthalakath, H.; Huang, D. C. S.; O'Reilly, L. A.; King, S. M.; Strasser, A., The Proapoptotic Activity of the Bcl-2 Family Member Bim Is Regulated by Interaction with the Dynein Motor Complex. *Mol Cell* **1999,** *3* (3), 287-296.
132. Weber, A.; Paschen, S. A.; Heger, K.; Wilfling, F.; Frankenberg, T.; Bauerschmitt, H.; Seiffert, B. M.; Kirschnek, S.; Wagner, H.; Häcker, G., BimS-induced apoptosis requires mitochondrial localization but not interaction with anti-apoptotic Bcl-2 proteins. *J Cell Biol* **2007,** *177* (4), 625-36.
133. Clem, R. J.; Cheng, E. H.; Karp, C. L.; Kirsch, D. G.; Ueno, K.; Takahashi, A.; Kastan, M. B.; Griffin, D. E.; Earnshaw, W. C.; Veliuona, M. A.; Hardwick, J. M., Modulation of cell death by Bcl-XL through caspase interaction. *Proc Natl Acad Sci U S A* **1998,** *95* (2), 554-9.
134. Jonas, E. A.; Hickman, J. A.; Chachar, M.; Polster, B. M.; Brandt, T. A.; Fannjiang, Y.; Ivanovska, I.; Basañez, G.; Kinnally, K. W.; Zimmerberg, J.; Hardwick, J. M.; Kaczmarek, L. K., Proapoptotic N-truncated BCL-xL protein activates endogenous mitochondrial channels in living synaptic terminals. *Proc Natl Acad Sci U S A* **2004,** *101* (37), 13590-5.
135. Rosano, G. L.; Ceccarelli, E. A., Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **2014,** *5*, 172.
136. Novick, R. P., Extrachromosomal inheritance in bacteria. *Bacteriol Rev* **1969,** *33* (2), 210-63.

137. Hershfield, V.; Boyer, H. W.; Yanofsky, C.; Lovett, M. A.; Helinski, D. R., Plasmid ColEl as a molecular vehicle for cloning and amplification of DNA. *Proc Natl Acad Sci U S A* **1974,** *71* (9), 3455-3459.
138. Bolivar, F.; Rodriguez, R. L.; Greene, P. J.; Betlach, M. C.; Heyneker, H. L.; Boyer, H. W.; Crosa, J. H.; Falkow, S., Construction and characterization of new cloning vehicle. II. A multipurpose cloning system. *Gene* **1977,** *2* (2), 95-113.
139. Camps, M., Modulation of ColE1-like plasmid replication for recombinant gene expression. *Recent Pat DNA Gene Seq* **2010,** *4* (1), 58-73.
140. del Solar, G.; Giraldo, R.; Ruiz-Echevarría, M. J.; Espinosa, M.; Díaz-Orejas, R., Replication and control of circular bacterial plasmids. *Microbiol Mol Biol R* **1998,** *62* (2), 434-464.
141. Wang, Z.; Jin, L.; Yuan, Z.; Wegrzyn, G.; Wegrzyn, A., Classification of plasmid vectors using replication origin, selection marker and promoter as criteria. *Plasmid* **2009,** *61* (1), 47-51.
142. Chang, A. C.; Cohen, S. N., Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol* **1978,** *134* (3), 1141-1156.
143. Som, T.; Tomizawa, J., Origin of replication of Escherichia coli plasmid RSF 1030. *Mol Gen Genet* **1982,** *187* (3), 375-83.
144. Veltkamp, E.; Stuitje, A. R., Replication and structure of the bacteriocinogenic plasmids Clo DF13 and Col E1. *Plasmid* **1981,** *5* (1), 76-99.
145. Selzer, G.; Som, T.; Itoh, T.; Tomizawa, J., The origin of replication of plasmid p15A and comparative studies on the nucleotide sequences around the origin of related plasmids. *Cell* **1983,** *32* (1), 119-29.
146. Churchward, G.; Linder, P.; Caro, L., The nucleotide sequence of replication and maintenance functions encoded by plasmid pSC101. *Nucleic Acids Res* **1983,** *11* (16), 5645-5659.
147. Russell, D. R.; Bennett, G. N., Characterization of the beta-lactamase promoter of pBR322. *Nucleic Acids Res* **1981,** *9* (11), 2517-2533.
148. Yansura, D. G.; Henner, D. J., *Use of Escherichia coli trp promoter for direct expression of proteins*. Academic Press: 1990; Vol. 185, p 54-60.
149. Yansura, D. G.; Bass, S. H., *Application of the E. coli trp Promoter*. Humana Press: Totowa, NJ, 1997; p 55-62.
150. Silverstone, A. E.; Arditti, R. R.; Magasanik, B., Catabolite-Insensitive Revertants of Lac Promoter Mutants. *Proc Natl Acad Sci U S A* **1970,** *66* (3), 773.
151. de Boer, H. A.; Comstock, L. J.; Vasser, M., The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc Natl Acad U S A* **1983,** *80* (1), 21.
152. Brosius, J.; Erfle, M.; Storella, J., Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity. *J Biol Chem* **1985,** *260* (6), 3539-41.
153. Guzman, L. M.; Belin, D.; Carson, M. J.; Beckwith, J., Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **1995,** *177* (14), 4121-30.
154. O'Gorman, R. B.; Rosenberg, J. M.; Kallai, O. B.; Dickerson, R. E.; Itakura, K.; Riggs, A. D.; Matthews, K. S., Equilibrium binding of inducer to lac repressor.operator DNA complex. *J Biol Chem* **1980,** *255* (21), 10107-14.
155. Lee, N.; Francklyn, C.; Hamilton, E. P., Arabinose-induced binding of AraC protein to araI2 activates the araBAD operon promoter. *Proc Natl Acad Sci U S A* **1987,** *84* (24), 8814-8818.
156. Pollard, T. D., A guide to simple and informative binding assays. *Mol Biol Cell* **2010,** *21* (23), 4061-7.
157. Johnson, M. L.; Frasier, S. G., *Nonlinear least-squares analysis*. Academic Press: 1985; Vol. 117, p 301-342.
158. Goutelle, S.; Maurin, M.; Rougier, F.; Barbaut, X.; Bourguignon, L.; Ducher, M.; Maire, P., The Hill equation: a review of its capabilities in pharmacological modelling. *Fundam Clin Pharmacol* **2008,** *22* (6), 633-48.

159. Pollard, T. D.; De La Cruz, E. M., Take advantage of time in your experiments: a guide to simple, informative kinetics assays. *Mol Biol Cell* **2013,** *24* (8), 1103-10.
160. Givan, A. L., Flow cytometry: an introduction. *Methods Mol Biol* **2011,** *699*, 1-29.
161. Shapiro, H. M., *Practical flow cytometry*. 4 ed.; Wiley-Liss: Hoboken, NJ, 2003; p 166-170.
162. Shapiro, H. M., *Practical flow cytometry*. 4 ed.; Wiley-Liss: Hoboken, NJ, 2003; p 275-279.
163. Waggoner, A., Optical Filter Sets for Multiparameter Flow Cytometry. *Curr Protoc Cytom* **1997,** *00* (1), 1.5.1-1.5.8.
164. Hoffman, R. A., Pulse Width for Particle Sizing. *Curr Protoc Cytom* **2009,** *50* (1), 1.23.1-1.23.17.
165. Gustafsson, Å.; Krais, A. M.; Gorzsás, A.; Lundh, T.; Gerde, P., Isolation and characterization of a respirable particle fraction from residential house-dust. *Environ Res* **2018,** *161*, 284-290.
166. Shapiro, H. M., *Practical flow cytometry*. 4 ed.; Wiley-Liss: Hoboken, NJ, 2003; p 290.
167. Ou, F.; McGoverin, C.; Swift, S.; Vanholsbeeck, F., Absolute bacterial cell enumeration using flow cytometry. *J Appl Microbiol* **2017,** *123* (2), 464-477.
168. Adan, A.; Alizada, G.; Kiraz, Y.; Baran, Y.; Nalbant, A., Flow cytometry: basic principles and applications. *Crit Rev Biotechnol* **2017,** *37* (2), 163-176.
169. Davey, H. M.; Kell, D. B., Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. *Microbiol Rev* **1996,** *60* (4), 641-696.
170. Arnold, F. H., When blind is better: protein design by evolution. *Nat Biotechnol* **1998,** *16* (7), 617-8.
171. Morozova, K. S.; Piatkevich, K. D.; Gould, T. J.; Zhang, J.; Bewersdorf, J.; Verkhusha, V. V., Far-red fluorescent protein excitable with red lasers for flow cytometry and superresolution STED nanoscopy. *Biophys J* **2010,** *99* (2), L13-5.
172. Strack, R. L.; Hein, B.; Bhattacharyya, D.; Hell, S. W.; Keenan, R. J.; Glick, B. S., A rapidly maturing far-red derivative of DsRed-Express2 for whole-cell labeling. *Biochemistry* **2009,** *48* (35), 8279-8281.
173. Nguyen, A. W.; Daugherty, P. S., Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat Biotechnol* **2005,** *23* (3), 355-360.
174. Nadler, D. C.; Morgan, S.-A.; Flamholz, A.; Kortright, K. E.; Savage, D. F., Rapid construction of metabolite biosensors using domain-insertion profiling. *Nat Commun* **2016,** *7*, 12266-12266.
175. Galbraith, D. W.; Anderson, M. T.; Herzenberg, L. A., *Flow Cytometric Analysis and FACS Sorting of Cells Based on GFP Accumulation*. Academic Press: 1998; Vol. 58, p 315-341.
176. Telford, W. G.; Hawley, T.; Subach, F.; Verkhusha, V.; Hawley, R. G., Flow cytometry of fluorescent proteins. *Methods* **2012,** *57* (3), 318-330.
177. Gurskaya, N. G.; Staroverov, D. B.; Zhang, L.; Fradkov, A. F.; Markina, N. M.; Pereverzev, A. P.; Lukyanov, K. A., Analysis of alternative splicing of cassette exons at single-cell level using two fluorescent proteins. *Nucleic Acids Res* **2012,** *40* (8), e57-e57.
178. Thyrock, A.; Stehling, M.; Waschbüsch, D.; Barnekow, A., Characterizing the interaction between the Rab6 GTPase and Mint3 via flow cytometry based FRET analysis. *Biochem Biophys Res Commun* **2010,** *396* (3), 679-83.
179. Nedosekin, D. A.; Verkhusha, V. V.; Melerzanov, A. V.; Zharov, V. P.; Galanzha, E. I., In vivo photoswitchable flow cytometry for direct tracking of single circulating tumor cells. *Chemistry & biology* **2014,** *21* (6), 792-801.
180. Nienhaus, K.; Nienhaus, G. U., Fluorescent proteins for live-cell imaging with super-resolution. *Chem Soc Rev* **2014,** *43* (4), 1088-106.
181. Wang, S.; Moffitt, J. R.; Dempsey, G. T.; Xie, X. S.; Zhuang, X., Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proc Natl Acad Sci U S A* **2014,** *111* (23), 8452-7.
182. El Khatib, M.; Martins, A.; Bourgeois, D.; Colletier, J. P.; Adam, V., Rational design of ultrastable and reversibly photoswitchable fluorescent proteins for super-resolution imaging of the bacterial periplasm. *Sci Rep* **2016,** *6*, 18459.

183. Altinoglu, I.; Merrifield, C. J.; Yamaichi, Y., Single molecule super-resolution imaging of bacterial cell pole proteins with high-throughput quantitative analysis pipeline. *Sci Rep* **2019,** *9* (1), 6680.
184. Kremers, G. J.; Goedhart, J.; van Munster, E. B.; Gadella, T. W., Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Förster radius. *Biochemistry* **2006,** *45* (21), 6570-80.
185. Ay, A.; Knierer, S.; Sperlea, A.; Holland, J.; Özbudak, E. M., Short-lived Her proteins drive robust synchronized oscillations in the zebrafish segmentation clock. *Development* **2013,** *140* (15), 3244-53.
186. Chen, W.; Smeekens, J. M.; Wu, R., Systematic study of the dynamics and half-lives of newly synthesized proteins in human cells. *Chem Sci* **2016,** *7* (2), 1393-1400.
187. Belle, A.; Tanay, A.; Bitincka, L.; Shamir, R.; O'Shea, E. K., Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci U S A* **2006,** *103* (35), 13004-9.
188. Wang, X.; Errede, B.; Elston, T. C., Mathematical analysis and quantification of fluorescent proteins as transcriptional reporters. *Biophys J* **2008,** *94* (6), 2017-26.
189. Wilson-Annan, J.; O'Reilly, L. A.; Crawford, S. A.; Hausmann, G.; Beaumont, J. G.; Parma, L. P.; Chen, L.; Lackmann, M.; Lithgow, T.; Hinds, M. G.; Day, C. L.; Adams, J. M.; Huang, D. C., Proapoptotic BH3-only proteins trigger membrane integration of prosurvival Bcl-w and neutralize its activity. *J Cell Biol* **2003,** *162* (5), 877-87.
190. Chao, D. T.; Korsmeyer, S. J., BCL-2 family: regulators of cell death. *Annu Rev Immunol* **1998,** *16*, 395-419.
191. Dutta, S.; Gullá, S.; Chen, T. S.; Fire, E.; Grant, R. A.; Keating, A. E., Determinants of BH3 binding specificity for Mcl-1 versus Bcl-xL. *J Mol Biol* **2010,** *398* (5), 747-62.
192. Lee, E. F.; Sadowsky, J. D.; Smith, B. J.; Czabotar, P. E.; Peterson-Kaufman, K. J.; Colman, P. M.; Gellman, S. H.; Fairlie, W. D., High-resolution structural characterization of a helical alpha/beta-peptide foldamer bound to the anti-apoptotic protein Bcl-xL. *Angew Chem Int Ed* **2009,** *48* (24), 4318-22.
193. Hinds, M. G.; Lackmann, M.; Skea, G. L.; Harrison, P. J.; Huang, D. C.; Day, C. L., The structure of Bcl-w reveals a role for the C-terminal residues in modulating biological activity. *EMBO J* **2003,** *22* (7), 1497-507.
194. Oberstein, A.; Jeffrey, P. D.; Shi, Y., Crystal structure of the Bcl-XL-Beclin 1 peptide complex: Beclin 1 is a novel BH3-only protein. *J Biol Chem* **2007,** *282* (17), 13123-32.
195. Kvansakul, M.; Yang, H.; Fairlie, W. D.; Czabotar, P. E.; Fischer, S. F.; Perugini, M. A.; Huang, D. C.; Colman, P. M., Vaccinia virus anti-apoptotic F1L is a novel Bcl-2-like domain-swapped dimer that binds a highly selective subset of BH3-containing death ligands. *Cell Death Differ* **2008,** *15* (10), 1564-71.
196. Chen-Levy, Z.; Cleary, M. L., Membrane topology of the Bcl-2 proto-oncogenic protein demonstrated in vitro. *J Biol Chem* **1990,** *265* (9), 4929-33.
197. Nguyen, M.; Millar, D. G.; Yong, V. W.; Korsmeyer, S. J.; Shore, G. C., Targeting of Bcl-2 to the mitochondrial outer membrane by a COOH-terminal signal anchor sequence. *J Biol Chem* **1993,** *268* (34), 25265-8.
198. Akerboom, J.; Chen, T. W.; Wardill, T. J.; Tian, L.; Marvin, J. S.; Mutlu, S.; Calderón, N. C.; Esposti, F.; Borghuis, B. G.; Sun, X. R.; Gordus, A.; Orger, M. B.; Portugues, R.; Engert, F.; Macklin, J. J.; Filosa, A.; Aggarwal, A.; Kerr, R. A.; Takagi, R.; Kracun, S.; Shigetomi, E.; Khakh, B. S.; Baier, H.; Lagnado, L.; Wang, S. S.; Bargmann, C. I.; Kimmel, B. E.; Jayaraman, V.; Svoboda, K.; Kim, D. S.; Schreiter, E. R.; Looger, L. L., Optimization of a GCaMP calcium indicator for neural activity imaging. *J Neurosci* **2012,** *32* (40), 13819-40.
199. Weisemann, J. M.; Weinstock, G. M., The promoter of the recA gene of Escherichia coli. *Biochimie* **1991,** *73* (4), 457-70.
200. Rosenberg, A. H.; Lade, B. N.; Chui, D. S.; Lin, S. W.; Dunn, J. J.; Studier, F. W., Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **1987,** *56* (1), 125-35.

201. Studier, F. W.; Moffatt, B. A., Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **1986,** *189* (1), 113-30.
202. Studier, F. W.; Rosenberg, A. H.; Dunn, J. J.; Dubendorff, J. W., Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **1990,** *185*, 60-89.
203. Lo Leggio, L.; Kalogiannis, S.; Eckert, K.; Teixeira, S. C.; Bhat, M. K.; Andrei, C.; Pickersgill, R. W.; Larsen, S., Substrate specificity and subsite mobility in T. aurantiacus xylanase 10A. *FEBS Lett* **2001,** *509* (2), 303-8.
204. Lo Leggio, L.; Kalogiannis, S.; Bhat, M. K.; Pickersgill, R. W., High resolution structure and sequence of T. aurantiacus xylanase I: implications for the evolution of thermostability in family 10 xylanases and enzymes with (beta)alpha-barrel architecture. *Proteins* **1999,** *36* (3), 295-306.
205. Wierenga, R. K., The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett* **2001,** *492* (3), 193-8.
206. Natesh, R.; Bhanumoorthy, P.; Vithayathil, P. J.; Sekar, K.; Ramakumar, S.; Viswamitra, M. A., Crystal structure at 1.8 A resolution and proposed amino acid sequence of a thermostable xylanase from Thermoascus aurantiacus. *J Mol Biol* **1999,** *288* (5), 999-1012.
207. Chen, T. W.; Wardill, T. J.; Sun, Y.; Pulver, S. R.; Renninger, S. L.; Baohan, A.; Schreiter, E. R.; Kerr, R. A.; Orger, M. B.; Jayaraman, V.; Looger, L. L.; Svoboda, K.; Kim, D. S., Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **2013,** *499* (7458), 295-300.
208. Horne, W. S.; Boersma, M. D.; Windsor, M. A.; Gellman, S. H., Sequence-based design of alpha/beta-peptide foldamers that mimic BH3 domains. *Angew Chem Int Ed* **2008,** *47* (15), 2853-6.
209. Ghaemmaghami, S.; Huh, W. K.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S., Global analysis of protein expression in yeast. *Nature* **2003,** *425* (6959), 737-41.
210. Jorgensen, P.; Nishikawa, J. L.; Breitkreutz, B. J.; Tyers, M., Systematic identification of pathways that couple cell growth and division in yeast. *Science* **2002,** *297* (5580), 395-400.
211. Milo, R.; Jorgensen, P.; Moran, U.; Weber, G.; Springer, M., BioNumbers--the database of key numbers in molecular and cell biology. *Nucleic Acids Res* **2010,** *38* (Database issue), D750-3.
212. Beck, M.; Schmidt, A.; Malmstroem, J.; Claassen, M.; Ori, A.; Szymborska, A.; Herzog, F.; Rinner, O.; Ellenberg, J.; Aebersold, R., The quantitative proteome of a human cell line. *Mol Syst Biol* **2011,** *7*, 549.
213. Malmström, J.; Beck, M.; Schmidt, A.; Lange, V.; Deutsch, E. W.; Aebersold, R., Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **2009,** *460* (7256), 762-5.
214. Phillips, G. N., Structure and dynamics of green fluorescent protein. *Curr Opin Struct Biol* **1997,** *7* (6), 821-7.
215. Rajan, S.; Choi, M.; Nguyen, Q. T.; Ye, H.; Liu, W.; Toh, H. T.; Kang, C.; Kamariah, N.; Li, C.; Huang, H.; White, C.; Baek, K.; Grüber, G.; Yoon, H. S., Structural transition in Bcl-xL and its potential association with mitochondrial calcium ion transport. *Sci Rep* **2015,** *5*, 10609.
216. Lee, E. F.; Dewson, G.; Evangelista, M.; Pettikiriarachchi, A.; Gold, G. J.; Zhu, H.; Colman, P. M.; Fairlie, W. D., The functional differences between pro-survival and pro-apoptotic B cell lymphoma 2 (Bcl-2) proteins depend on structural differences in their Bcl-2 homology 3 (BH3) domains. *J Biol Chem* **2014,** *289* (52), 36001-36017.
217. Alford, S. C.; Abdelfattah, A. S.; Ding, Y.; Campbell, R. E., A fluorogenic red fluorescent protein heterodimer. *Chem Biol* **2012,** *19* (3), 353-60.
218. Alford, S. C.; Ding, Y.; Simmen, T.; Campbell, R. E., Dimerization-dependent green and yellow fluorescent proteins. *ACS Synth Biol* **2012,** *1* (12), 569-75.
219. Wongso, D.; Dong, J.; Ueda, H.; Kitaguchi, T., Flashbody: A Next Generation Fluobody with Fluorescence Intensity Enhanced by Antigen Binding. *Anal Chem* **2017,** *89* (12), 6719-6725.
220. Ernst, O.; Zor, T., Linearization of the bradford protein assay. *J Vis Exp* **2010,** (38).

221. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A., *Protein Identification and Analysis Tools on the ExPASy Server*. Humana Press: 2005; Vol. 1.
222. Patterson, G. H.; Knobel, S. M.; Sharif, W. D.; Kain, S. R.; Piston, D. W., Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys J* **1997,** *73* (5), 2782-90.
223. Lutz, R.; Bujard, H., Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **1997,** *25* (6), 1203-10.
224. Shen, Y.; Dana, H.; Abdelfattah, A. S.; Patel, R.; Shea, J.; Molina, R. S.; Rawal, B.; Rancic, V.; Chang, Y.-F.; Wu, L.; Chen, Y.; Qian, Y.; Wiens, M. D.; Hambleton, N.; Ballanyi, K.; Hughes, T. E.; Drobizhev, M.; Kim, D. S.; Koyama, M.; Schreiter, E. R.; Campbell, R. E., A genetically encoded Ca2+ indicator based on circularly permutated sea anemone red fluorescent protein eqFP578. *BMC Biol* **2018,** *16* (1), 9.
225. Horton, R. M.; Ho, S. N.; Pullen, J. K.; Hunt, H. D.; Cai, Z.; Pease, L. R., Gene splicing by overlap extension. *Methods Enzymol* **1993,** *217*, 270-9.
226. Myers, J. A.; Curtis, B. S.; Curtis, W. R., Improving accuracy of cell and chromophore concentration measurements using optical density. *BMC Biophys* **2013,** *6* (1), 4.
227. Rizzo, M. A.; Davidson, M. W.; Piston, D. W., Fluorescent Protein Tracking and Detection: Applications Using Fluorescent Proteins in Living Cells. *Cold Spring Harb Protoc* **2009,** *4* (12), 1-13.
228. Tallini, Y. N.; Ohkura, M.; Choi, B. R.; Ji, G.; Imoto, K.; Doran, R.; Lee, J.; Plan, P.; Wilson, J.; Xin, H. B.; Sanbe, A.; Gulick, J.; Mathai, J.; Robbins, J.; Salama, G.; Nakai, J.; Kotlikoff, M. I., Imaging cellular signals in the heart in vivo: Cardiac expression of the high-signal Ca2+ indicator GCaMP2. *Proc Natl Acad Sci U S A* **2006,** *103* (12), 4753-8.
229. Stevens, F. C., Calmodulin: an introduction. *Can J Biochem Cell Biol* **1983,** *61* (8), 906-10.
230. Romoser, V. A.; Hinkle, P. M.; Persechini, A., Detection in living cells of Ca2+-dependent changes in the fluorescence emission of an indicator composed of two green fluorescent protein variants linked by a calmodulin-binding sequence. A new class of fluorescent indicators. *J Biol Chem* **1997,** *272* (20), 13270-4.
231. Nagai, T.; Sawano, A.; Park, E. S.; Miyawaki, A., Circularly permuted green fluorescent proteins engineered to sense Ca2+. *Proc Natl Acad Sci U S A* **2001,** *98* (6), 3197-202.
232. Lelimousin, M.; Noirclerc-Savoye, M.; Lazareno-Saez, C.; Paetzold, B.; Le Vot, S.; Chazal, R.; Macheboeuf, P.; Field, M. J.; Bourgeois, D.; Royant, A., Intrinsic dynamics in ECFP and Cerulean control fluorescence quantum yield. *Biochemistry* **2009,** *48* (42), 10038-46.
233. Yasui, M.; Hiroshima, M.; Kozuka, J.; Sako, Y.; Ueda, M., Automated single-molecule imaging in living cells. *Nat Commun* **2018,** *9* (1), 3061.
234. Hibino, K.; Hiroshima, M.; Takahashi, M.; Sako, Y., *Single-Molecule Imaging of Fluorescent Proteins Expressed in Living Cells*. Humana Press: Totowa, NJ, 2009; p 451-460.
235. Douglass, A. D.; Vale, R. D., *Single-Molecule Imaging of Fluorescent Proteins*. Academic Press: 2008; Vol. 85, p 113-125.
236. Enterina, J. R.; Wu, L.; Campbell, R. E., Emerging fluorescent protein technologies. *Curr Opin Chem Biol* **2015,** *27*, 10-17.
237. Morris, M. C., Fluorescent Biosensors of Intracellular Targets from Genetically Encoded Reporters to Modular Polypeptide Probes. *Cell Biochem Biophys* **2009,** *56* (1), 19.
238. St-Pierre, F.; Marshall, J. D.; Yang, Y.; Gong, Y.; Schnitzer, M. J.; Lin, M. Z., High-fidelity optical reporting of neuronal electrical activity with an ultrafast fluorescent voltage sensor. *Nat Neurosci* **2014,** *17* (6), 884-889.
239. Schmitt, D. L.; Mehta, S.; Zhang, J., Illuminating the kinome: Visualizing real-time kinase activity in biological systems using genetically encoded fluorescent protein-based biosensors. *Curr Opin Chem Biol* **2020,** *54*, 63-69.

240. Ai, H.-w.; Hazelwood, K. L.; Davidson, M. W.; Campbell, R. E., Fluorescent protein FRET pairs for ratiometric imaging of dual biosensors. *Nat Methods* **2008,** *5* (5), 401-403.
241. Dana, H.; Sun, Y.; Mohar, B.; Hulse, B. K.; Kerlin, A. M.; Hasseman, J. P.; Tsegaye, G.; Tsang, A.; Wong, A.; Patel, R.; Macklin, J. J.; Chen, Y.; Konnerth, A.; Jayaraman, V.; Looger, L. L.; Schreiter, E. R.; Svoboda, K.; Kim, D. S., High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat Methods* **2019,** *16* (7), 649-657.
242. Leavitt, S.; Freire, E., Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr Opin Struct Biol* **2001,** *11* (5), 560-6.
243. Douzi, B., *Protein–Protein Interactions: Surface Plasmon Resonance*. Springer: New York, NY, 2017; p 257-275.
244. Phillips, R., *Physical biology of the cell*. 2 ed.; Taylor & Francis: New York, NY, 2013; p 31.
245. Mir, M.; Reimer, A.; Haines, J. E.; Li, X.-Y.; Stadler, M.; Garcia, H.; Eisen, M. B.; Darzacq, X., Dense Bicoid hubs accentuate binding along the morphogen gradient. *Gene Dev* **2017,** *31* (17), 1784-1794.
246. Gregor, T.; Tank, D. W.; Wieschaus, E. F.; Bialek, W., Probing the limits to positional information. *Cell* **2007,** *130* (1), 153-64.
247. Giudicelli, F.; Lewis, J., The vertebrate segmentation clock. *Curr Opin Genet Dev* **2004,** *14* (4), 407-414.
248. Lewis, J., Autoinhibition with Transcriptional Delay: A Simple Mechanism for the Zebrafish Somitogenesis Oscillator. *Curr Biol* **2003,** *13* (16), 1398-1408.
249. Liao, B.-K.; Jörg, D. J.; Oates, A. C., Faster embryonic segmentation through elevated Delta-Notch signalling. *Nat Commun* **2016,** *7* (1), 11861.
250. Giudicelli, F.; Ozbudak, E. M.; Wright, G. J.; Lewis, J., Setting the tempo in development: an investigation of the zebrafish somite clock mechanism. *PLoS Biol* **2007,** *5* (6), e150-e150.
251. Nguyen, M.-A.; Karunakaran, D.; Geoffrion, M.; Cheng, H. S.; Tandoc, K.; Perisic Matic, L.; Hedin, U.; Maegdefessel, L.; Fish, J. E.; Rayner, K. J., Extracellular Vesicles Secreted by Atherogenic Macrophages Transfer MicroRNA to Inhibit Cell Migration. *Arterioscl Throm Vas* **2018,** *38* (1), 49-63.
252. Campbell-Valois, F. X.; Pontier, S. M., Implications of Spatiotemporal Regulation of Shigella flexneri Type Three Secretion Activity on Effector Functions: Think Globally, Act Locally. *Front Cell Infect Microbiol* **2016,** *6*, 28.
253. Specht, E. A.; Braselmann, E.; Palmer, A. E., A Critical and Comparative Review of Fluorescent Tools for Live-Cell Imaging. *Annu Rev Physiol* **2017,** *79*, 93-117.
254. Griffin, B. A.; Adams, S. R.; Tsien, R. Y., Specific Covalent Labeling of Recombinant Protein Molecules Inside Live Cells. *Science* **1998,** *281* (5374), 269.
255. Chen, Y.; Tsao, K.; Acton, S. L.; Keillor, J. W., A Green BODIPY-Based, Super-Fluorogenic, Protein-Specific Labelling Agent. *Angew Chem Int Edit* **2018,** *57* (38), 12390-12394.
256. Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M., Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc Natl Acad Sci U S A* **2012,** *109* (12), E690-E697.
257. Cabantous, S.; Terwilliger, T. C.; Waldo, G. S., Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol* **2005,** *23* (1), 102-7.
258. Cabantous, S.; Nguyen, H. B.; Pedelacq, J. D.; Koraïchi, F.; Chaudhary, A.; Ganguly, K.; Lockard, M. A.; Favre, G.; Terwilliger, T. C.; Waldo, G. S., A new protein-protein interaction sensor based on tripartite split-GFP association. *Sci Rep* **2013,** *3*, 2854.
259. Pedelacq, J.-D.; Cabantous, S., Development and Applications of Superfolder and Split Fluorescent Protein Detection Systems in Biology. *Int J Mol Sci* **2019,** *20* (14), 3479.
260. Huang, Y.-M.; Bystroff, C., Complementation and Reconstitution of Fluorescence from Circularly Permuted and Truncated Green Fluorescent Protein. *Biochemistry* **2009,** *48* (5), 929-940.

261. Huang, Y.-M.; Nayak, S.; Bystroff, C., Quantitative in vivo solubility and reconstitution of truncated circular permutants of green fluorescent protein. *Protein Science* **2011,** *20* (11), 1775-1780.
262. Na, Y.; Park, S.; Lee, C.; Kim, D.-K.; Park, J. M.; Sockanathan, S.; Huganir, R. L.; Worley, P. F., Real-Time Imaging Reveals Properties of Glutamate-Induced Arc/Arg 3.1 Translation in Neuronal Dendrites. *Neuron* **2016,** *91* (3), 561-573.
263. Tanenbaum, M. E.; Gilbert, L. A.; Qi, L. S.; Weissman, J. S.; Vale, R. D., A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **2014,** *159* (3), 635-646.
264. Boersma, S.; Khuperkar, D.; Verhagen, B. M. P.; Sonneveld, S.; Grimm, J. B.; Lavis, L. D.; Tanenbaum, M. E., Multi-Color Single-Molecule Imaging Uncovers Extensive Heterogeneity in mRNA Decoding. *Cell* **2019,** *178* (2), 458-472.e19.
265. Aymoz, D.; Wosika, V.; Durandau, E.; Pelet, S., Real-time quantification of protein expression at the single-cell level via dynamic protein synthesis translocation reporters. *Nat Commun* **2016,** *7*, 11304-11304.
266. Bothma, J. P.; Norstad, M. R.; Alamos, S.; Garcia, H. G., LlamaTags: A Versatile Tool to Image Transcription Factor Dynamics in Live Embryos. *Cell* **2018,** *173* (7), 1810-1822.e16.
267. Kirchhofer, A.; Helma, J.; Schmidthals, K.; Frauer, C.; Cui, S.; Karcher, A.; Pellis, M.; Muyldermans, S.; Casas-Delucchi, C. S.; Cardoso, M. C.; Leonhardt, H.; Hopfner, K.-P.; Rothbauer, U., Modulation of protein properties in living cells using nanobodies. *Nat Struct Mol Biol* **2010,** *17* (1), 133-138.
268. Gibson, D. G.; Glass, J. I.; Lartigue, C.; Noskov, V. N.; Chuang, R.-Y.; Algire, M. A.; Benders, G. A.; Montague, M. G.; Ma, L.; Moodie, M. M.; Merryman, C.; Vashee, S.; Krishnakumar, R.; Assad-Garcia, N.; Andrews-Pfannkoch, C.; Denisova, E. A.; Young, L.; Qi, Z.-Q.; Segall-Shapiro, T. H.; Calvey, C. H.; Parmar, P. P.; Hutchison, C. A.; Smith, H. O.; Venter, J. C., Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **2010,** *329* (5987), 52.