Winter 1-23-2017

Engineering a Fluorescent Protease Sensor for in vivo Protein Detection

Thomas C. Kinard
thomas.kinard@uconn.edu

Follow this and additional works at: https://opencommons.uconn.edu/srhonors_theses

Part of the Biochemical and Biomolecular Engineering Commons, Molecular Biology Commons, and the Molecular, Cellular, and Tissue Engineering Commons

Recommended Citation
https://opencommons.uconn.edu/srhonors_theses/591
Engineering a Fluorescent Protease Sensor for *in vivo* Protein Detection

*For application in Investigation of Ovulation in Drosophila*

Thomas Kinard  
Advisor: Dr. Yongku Cho  
Research Partners: Kathryn Frances, Caroline Olesiak, Hamze Mah, Krystyna Kapaczynski,  
Steven Rodriguez

An Honors Thesis

Department of Chemical and Biomolecular Engineering  
University of Connecticut  
June 10, 2016
Abstract
This report details the results of an ongoing project to engineer a mutant form of Red Fluorescent Protein (RFP) variant mCherry that acts as a real-time in vivo protease sensor. The sought-after mutant only becomes fluorescent when exposed to Tobacco Etch Virus (TEV) Protease, this system’s model protease. This will be accomplished via the insertion of the TEV Protease Recognition Site (TEV-PRS) in such a position that, before cleavage, will prevent the protein from folding to fluorescent conformation, but upon cleavage, will allow for fluorescent conformation to occur. The cylindrical structure of the protein, composed of beta-pleated sheets, contains “loops” connecting these sheets, was analyzed, and sixteen sites were identified within these loops as candidates for insertion sites of the TEV-PRS. Using Site-Directed Mutagenesis (SDM) and Overlap-Extension Polymerase Chain-Reaction (OE-PCR), insertion at twelve of the identified sites has been attempted, two of which have been expressed in surface display yeast, digitally imaged, and their initial (pre-cleavage) fluorescence has been quantified. Four of the intended sites were successfully mutated, in addition to one accidental site. While the sought after mutant has not yet been identified, the results at this stage indicate promise, and further study is recommended and will be carried on in the future by other Research Assistants.
# Table of Contents

Abstract ................................................................................................................................. 1  
Introduction and Theory ........................................................................................................ 3  
Methods and Materials ......................................................................................................... 5  
  Enzymes ............................................................................................................................... 5  
  Cells ..................................................................................................................................... 6  
  DNA/Genetic Data .............................................................................................................. 6  
  Mutation .............................................................................................................................. 7  
    Site Directed Mutagenesis (SDM) ....................................................................................... 7  
    Overlap Extension PCR (OE-PCR) .................................................................................... 8  
  Heat-Shock Plasmid Transformation into Bacteria ............................................................. 9  
  Yeast Growth and Induction for Surface Display ............................................................... 9  
Results .................................................................................................................................... 10  
  Insertion Site Identification ............................................................................................... 10  
  Primer Design ..................................................................................................................... 12  
  Mutation .............................................................................................................................. 13  
    Site Directed Mutagenesis (SDM) ....................................................................................... 13  
    Overlap Extension PCR (OE-PCR) .................................................................................... 14  
Expression ............................................................................................................................. 19  
Discussion .............................................................................................................................. 21  
  Mutation .............................................................................................................................. 21  
  Expression ........................................................................................................................... 22  
Conclusion .............................................................................................................................. 23  
Appendix ................................................................................................................................ 24  
References .............................................................................................................................. 24  
  Figures .................................................................................................................................. 25  
  Tables .................................................................................................................................. 27  
  Nomenclature ....................................................................................................................... 27
Introduction and Theory

Sensing protein activity in vivo is a great challenge. While protein assays can be performed upon the spilled contents of lysed cells to find concentrations, this requires that the cells die. These assays, as a result, only roughly reveal the protein concentration at the time of lysis, not over time. In the case of prokaryotes or single-celled eukaryotes, groups of cells can be carried through the same conditions and made to lyse at different times, generating a profile of concentration over time. This process, however, is cumbersome, and introduces a possible source of error in the presence of a lytic agent; it may cause degradation or misfolding of the proteins of interest. Additionally, in the case of multicellular eukaryotes such as animals, this strategy becomes increasingly difficult to implement due to the complex nature of the organisms. Therefore, a non-invasive method of detecting protein activity in real time would be a highly useful tool for investigation of gene expression and protein function, especially in plants and animals. Perhaps the most significant advantage, however, would be the ability to detect protein activity in living organisms, either unicellular or multicellular. One such signaling method would be fluorescence. Light, unlike other signaling agents, can travel harmlessly through cells and can be easily detected, and even quantified, by photovoltaic sensors.

One particular example of a use for these sorts of sensors would be in detecting the real-time activity of certain proteases in Drosophila melanogaster, a common fruit fly. This multicellular organism is of great interest in the study of multicellular genetics, as, similar to Escherichia coli among bacteria, Drosophila has been studied in great detail relative to other animals, and the understanding of its gene expression is relatively deep. One major gap in understanding, however, is the genetic and physiological mechanism behind ovulation [1]. Certain proteases are speculated to be involved [2]. The ability to sense the presence of these proteins in real time could potentially elucidate their role in the ovulation process, shedding great light on the species, and multicellular organisms in general [3]. Due to the suspected involvement of a protease in Drosophila ovulation, this mystery could be a prime candidate for the application of a fluorescent protease sensor system in vivo.

![Figure 1](image_url); Visualization of the mCherry secondary and tertiary structure generated in PyMOL with an Educational License, with the color gradient from N to C terminus.
In designing such a system, the fluorescent molecule selected should be optimized for the application; in this case, fluorescence must be detectible from within the organism. A few varieties of fluorescent proteins occur in nature. The most well-known is likely Green Fluorescent Protein (GFP), derived from jellyfish. The protein is made up of a “barrel” shape constructed from beta-pleated sheets, as well as a fluorescent core made up of several alpha helices. Another variety is known as dsRed, derived from coral-reef organisms [4]. While similar in structure, the wavelengths of light emitted are at the lower-frequency end of the visible spectrum. This is ideal for in vivo detection within organisms, as longer wavelengths of light can more effectively pass through tissue. This heterodimeric protein has been mutated, [4] generating variants with different fluorescent wavelengths, one of which is a monomer called mCherry. mCherry is a monomeric protein capable of emitting peak fluorescence at a wavelength of 610 nm [5]. Due its exceptionally high photostability and maturation rate, and low bleaching rate, is an excellent candidate for use as a signaling agent. [6]

An important factor for use as a signaling agent is that the mechanism of initiating signaling should not be transcriptional. The reason for this is that, were this the case, sensing would depend upon gene expression, which is very difficult to predict in multicellular eukaryotes. Additionally, a very high concentration of the protein being monitored would need to accumulate in order for detection to be possible, meaning that if the protein is only active at very low concentrations, the fluorescence would not be sufficient. A more significant issue with expressional signaling is that expression can take hours to occur at a level sufficient for detection. This time delay makes real time detection impossible.

As an alternative, the mCherry protein will be engineered into a protease sensor. The goal of this project will be to create a protein which is naturally non-fluorescent, but becomes fluorescent in the presence of a protease via recognition-site-cleavage. This will be accomplished by inserting DNA coding for recognition site for the Tobacco Etch Virus Protease, or TEV Protease at various spots or “Insertion Sites”. Criteria for the site are as follows:

- Must be between codons, such that all codons in the mCherry gene and TEV Protease Recognition Site (TEV-PRS) remain intact and no frameshift mutations occur.
- Must not disturb the secondary structure of the mCherry protein, by breaking up any beta-sheets or alpha-helices, as this would irreparably damage the structure.
- Must be present on the surface of the roughly cylindrical protein structure, such that the protease will be able to access the recognition site.

mCherry is also optimal for TEV-PRS insertion for the reason that it is a monomeric protein, so impacting oligomerization behavior with insertions is not a concern.
As a result of these restrictions, the sites will be present in the loops which connect the beta-sheets with each other and also the fluorescent core at the top and bottom of the protein. The behavior of the protein and the function of the inserted recognition site will be as follows: the inserted sequence will initially cause the protein to be unable to fold into a fluorescent conformation by “blocking” this formation. The presence of TEV, even at low concentrations, will lead to virtually instantaneous cleavage of these sites, allowing the proteins to come together and fluoresce, without otherwise impacting the behavior or metabolism of the cell.

Methods and Materials

Various kits, enzymes, cells, DNA and plasmids, and lab equipment were employed throughout the lab work that went into this thesis.

Enzymes

Phusion High-Fidelity DNA polymerase from Thermo-Scientific was used for all PCR reactions, and performed all in vitro DNA amplification in-lab. Similarly to Taq polymerase, Phusion operates optimally at high temperatures (72°C), but has 52x fidelity.

T4 DNA Ligase was used for ligation reactions to “repair nicks” in DNA; in other words, the ligase catalyzes the formation of phosphodiester bonds to join adjacent nucleotides (connected by complementary “sticky ends at restriction sites, for example).

DpnI [7] is a restriction enzyme which is used for digesting methylated DNA. Methylation, being a tool used for in vivo gene expression, would only be present in DNA produced in vivo, but not synthetic DNA produced via PCR.

The High-Fidelity versions of NheI and BsrGI restriction enzymes were used to cleave an insert containing the yo-mCherry gene from plasmids.

Figure 3; Cleavage sites for BsrGI-HF (left) and NheI-HF (right) from New England Bio Labs
Cells

The bacterial cells used in plasmid amplification were TOP10 strain *E. coli* cells from Invitrogen. These cells were selected for their high transformation efficiency (1x10^9 CFU/µg plasmid) and chemical competence, which make them ideal for high efficiency cloning and plasmid propagation. Chemical competence is achieved by treating the cells with calcium chloride to facilitate plasmid DNA attachment to the cell membrane, increasing transformation efficiency. The cells lack antibiotic resistance, and allow for stable replication of plasmids with high copy number.

The yeast cells used in fluorescence expression tests for the mutated mCherry genes were EBY100 strain *Saccharomyces cerevisiae*, a strain of yeast optimized for surface display. The strain contains the GAL operon, and also an insertion of AGA1 regulated by GAL. AGA1 codes for the anchorage subunit of a-agglutinin, which embeds in the cell wall of molecules, and allows for the surface presentation of AGA2 linked. This means that in the presence of galactose as a carbon source, AGA1 and other GAL-linked genes are expressed.

DNA/Genetic Data

The gene coding for the protein of interest, mCherry, was a preexisting yeast-optimized variant. While the resulting amino-acid sequence is the same, the codons are optimized based upon the codon tRNA bias within the yeast genome.

The two plasmids employed in this research project were pN3 [8] and pCT302 [9].

![Plasmid maps of pCT302 and pN3](https://example.com/plasmid_maps.png)

*pCT302* is a yeast expression plasmid which is just over 6.1 kb in size without the insertion. It contains the gene for Ampicillin resistance, as well as that for yeast Tryptophan synthase. pCT302 also contains a GAL promoter and AGA2 linked to the insertion site, so when transformed into EBY100 yeast and induced by presence of galactose, the insert-aga2 complex disulfide bonds to the aga1 complex on the yeast surface and can be detected outside the cell.

pN3 is a mammalian expression plasmid just under 4kb in size without an insert. It contains the gene for Kanamycin resistance.
Mutation

Two methods, Site Directed Mutagenesis (SDM), and Overlap Extension Polymerase-Chain Reaction (OE-PCR) were employed to perform the insertion mutation. Both methods used the same plasmids, and had the same end result, but each had different requirements and procedures.

Site Directed Mutagenesis (SDM)

The first of two mutagenesis techniques employed was Site-Directed Mutagenesis (SDM). The technique requires 2 complementary oligonucleotides to act as primers. Starting from the 5’ end, the oligonucleotides must include roughly 20 base pairs upstream of the insertion point, followed by the insertion sequence, followed by roughly 20 base pairs downstream. The reaction solution is prepared to contain 10 pmol of each oligonucleotide primer, 50 ng template plasmid, 10 μL 5X reaction buffer (containing various buffers, BSA, and detergent to ensure proper polymerase function), 1 μL dNTP to fuel synthesis of DNA, and 0.5 μL Phusion DNA polymerase, made up to a volume of 50 μL with ddH2O. It is important to keep the solution and all components on ice until it is placed into the thermal cycler to prevent denaturation of the polymerase.

The thermal cycle, as depicted in Figure 5B, begins with 1 minute of denaturation phase at 95°C before beginning the 280 second cycle of denaturation-annealing-extension. The denaturation phase is also run at 95°C so that all dsDNA separates to allow primers to bond during the annealing phase. The 50 second annealing phase is run at 65°C, allowing primers to bond to the template plasmids (as shown in the first phase of Figure 5A). The 180 second extension phase is run at 72°C, and at this point the Phusion DNA polymerase attaches to the primers which are bonded to the template plasmid and extends the plasmids in the 3’ direction around the circular plasmid until it encounters the 5’ end of the plasmid being extended, where it detaches. The extension phase for this SDM cycle is longer than that for a standard PCR cycle. This is because the polymerase must extend the primer around the entire template plasmid, more than 4kb. The SDM reaction cycles 18 times, and does not propagate exponentially in the way that PCR does. At the conclusion of the 18th cycle, 5 additional minutes are allowed for extension, and then the products are held indefinitely at 4°C, where they can be safely stored for 1-2 days.

Figure 5; A; Visual representation of Site-Directed Mutagenesis used for insertion, derived from Geiser et al. [10]. The segmented loop represents the template plasmid (which is methylated, as it was synthesized in vivo), while the thick black segment represents the insertion. The primers are extended from the 3’ end along the template DNA until termination at their own 5’ end, resulting in a complete, nicked plasmid with the insertion, which is unmethylated since it was synthesized in vitro. Overnight digestion in DpnI restricts only the methylated DNA, leaving only the mutated plasmid to be transformed into E. coli and amplified. B; Thermal cycle diagram of “sdm phusion” cycle used for Site-Directed Mutagenesis. The cycle begins with a 60 second denaturation phase.
Following the thermal cycling, the SDM products must undergo *DpnI* restriction digest to remove all methylated template DNA. Each sample must receive 1 μL of *DpnI*, and must incubate overnight (for at least 8 hours) at 37°C. The digested products can then be transformed into TOP10 Competent *E. coli* cells for amplification, and the cells miniprepped to yield mutant plasmid.

**Overlap Extension PCR (OE-PCR)**

The second Mutagenesis technique employed was Overlap Extension Polymerase Chain Reaction, (OE-PCR). The technique requires 4 oligonucleotide primers; 2 mutagenic primers (same as those used for SDM) and additionally 2 flanking primers. Additionally, the technique requires 2 separate PCR reactions carried out in parallel, followed by a second round PCR reaction. The reacting solution is prepared in much the same way was described for SDM; the key variables are the primers used and the template dsDNA used. 10 pmol of each primer, 50 ng template plasmid, 10 μL 5X reaction buffer, 1 μL dNTP, and 0.5 μL Phusion DNA polymerase are made up to a volume of 50μL. The thermal cycle used is that illustrated in Figure 6B.

The first round of PCR pairs the forward flanking primer with the reverse mutagenic primer (Figure 6A a + b respectively) for one reaction, and the forward mutagenic primer with the reverse flanking primer (Figure 6A c + d respectively). These PCR reactions are independently carried out with the template DNA to produce two independent dsDNA products (Figure 6A AB and CD), which overlap with each other over the region of the mutagenic primers. The products must undergo gel purification and recovery to isolate the dsDNA PCR products. In the second round of PCR, the products for both previous PCR reactions must be included as template DNA. 1 μL of each product, in addition to both flanking primers, are included in this reacting solution. The remainder of the solution should be prepared in the same way as the first PCR. This second round of PCR should be performed in duplicate in order to ensure enough product is produced.

---

**Figure 6; A; Visual representation of OE-PCR used for insertion, derived from Heckman et al. [11].** The double stranded DNA at the top represents the template DNA, a + d are flanking primers, and c + b are mutagenic primers. Flanking primer a pairs with mutagenic b, which are used together in the first round PCR reaction to make double stranded DNA fragment AB. Likewise, c + d pair to make CD. For the second round PCR reaction, dsDNA fragments AB and CD are combined along with flanking primers a + d. Denatured AB ssDNA anneals to the complementary CD region containing the insertion, and the fragments are extended to produce the final product of AD with the
Following gel purification of the second round PCR products, the double stranded DNA can be sequenced, ligated into a backbone, and expressed.

**Heat-Shock Plasmid Transformation into Bacteria**

For bacterial amplification of plasmids, Invitrogen TOP10 *E. coli* competent cells were used. The cells were first removed from the -80°C freezer and allowed to thaw on ice. 4 μL β-mercaptoethanol were added per 100 μL cells in order to increase transformation efficiency by degrading cell surface hydrocarbons [12]. The cells were allowed to incubate for 10 minutes on ice. For each transformation, 20 μL of cells were added to 2 μL of *DpnI* digested plasmid product in chilled microfuge tubes, which were then gently mixed and incubated on ice for 30 minutes. For the actual heat shock, the tubes were submerged in a 42°C water bath for 30 seconds and then placed back on ice for 2 minutes. In a BioSafety cabinet, 300 μL 37°C preheated SOC media were added to each tube, which were then incubated in a shaker at 37°C for 40 minutes. Again in the Biosafety cabinet, 150 μL of each tube were added to a tube containing 3-5 mL LB media with antibiotics (depending upon plasmid; ampicillin for pCT, kanamycin for pN3). This ensures that only transformed cells can grow. These tubes were then shaken and incubated overnight at 37°C. The following day, 20 μL were plated under the Biosafety hood on LB with the same antibiotic as before, and incubated upside down overnight. More than 100 colonies are expected from this procedure, and less than 5 constitutes cause for concern. Colonies were picked and miniprepped to yield amplified primers.

**Yeast Growth and Induction for Surface Display**

Successfully transformed surface display yeast (EBY100) with pCT302 was inoculated in the Biosafety Cabinet into 3-5 mL SD-CAA media (a synthetic, minimal medium with glucose as a carbon source, casamino acids, and deficient in both tryptophan and uracil). TRP1 acts as the selection marker, since tryptophan synthetase allows growth on a tryptophan-deficient marker. This inoculated sample was incubated for 24 hours at 30°C (no more than 72 hours). (The cell density should be between 10^7 and 10^8 cells mL^-1 after growth, corresponding to an optical density of around OD_{600} ~ 1 to 10.) These cells were centrifuged and the supernatant discarded. The grown cells, now ready for induction, were diluted with SG-CAA to an optical density of OD_{600} ~ 1 and resuspended. SG-CAA media is identical to SD-CAA, except galactose replaces glucose as the carbon source. These samples was incubated at 20°C for at least 20 hours, but no more than 48. After a brief diauxic shift, the GAL operon will become active, and the GAL 1-10 promoter on the pCT302 present will be transcribed. Being that yeast are eukaryotes capable of post-transcriptional control of gene expression, the genes present in pCT are transcribed and translated to produce a hybrid protein of mutant mCherry and aga2. At the same time, the GAL-linked AGA1 gene in the yeast DNA is also expressed. The aga1 complexes via disulfide bonds with the aga2-mCherry complex, which then embeds in the cell wall of the cells, and can be detected and measured.
Results

Insertion Site Identification

Drawing from Pavoor et al., successful insertion mutations can be made between residues interconnecting the \( \beta \)-pleated sheets comprising the barrel-structure [13]. The two insertion sites utilized in Pavoor et al. are present at Asp-102/Asp-103 and at Glu-172/Asp-173, as illustrated in Figure 5B, and it has been demonstrated that GFP can retain its fluorescence when insertions of various lengths can be made here.

While GFP and mCherry have notably different amino acid sequences (refer to Figure 5A), their barrel-and-core structures are very similar (refer to Figure 5B and 5C). The crystal structure of mCherry was analyzed in PyMol, and new insertion sites were sought between residues that:

a) Connect \( \beta \)-pleated sheets comprising the barrel-structure with each other as well as the interconnecting the \( \alpha \)-helices comprising the fluorescent core, as all of these sites occur on the edges of the barrel structure similarly to those employed in Pavoor et al., and

b) Are not themselves a part of any secondary structure, so as not to disturb fluorescence or overall structure irreparably.

Figure 7; A; an amino acid sequence alignment of GFP to mCherry performed with EBI’s webPRANK, B; Visual map of secondary structure for Green Fluorescent Protein annotated with residue numbers at identifying points in secondary structure [14] and Pavoor et al. insertion sites at Asp-102/Asp-103 and at Glu-172/Asp-173 indicated with an ‘@’ and a ‘#’, respectively, C; Visual map of secondary structure for mCherry Protein annotated with numbered and labeled (plain black text) residues at identifying points in secondary structure as well as with the numbered (in black circles) and labeled (red text) original sixteen insertion sites.
The destruction of a beta-pleated-sheet would make proper tertiary conformation impossible, and that of an alpha helix could fundamentally alter or destroy fluorescence. Sixteen such sites were identified. They are indicated in Figure 5C, as well as tabulated in Table 1. For shorthand, each of the 16 sites was called MC??, wherein ?? was the 2-digit form of the insertion site number.

Table 1: Insertion sites MC01-MC16 paired with their positions, derived from the map in Figure 6

<table>
<thead>
<tr>
<th>Insertion</th>
<th>Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC01</td>
<td>Y53/E53</td>
</tr>
<tr>
<td>MC02</td>
<td>K65/G66</td>
</tr>
<tr>
<td>MC03</td>
<td>E104/G105</td>
</tr>
<tr>
<td>MC04</td>
<td>D116/G117</td>
</tr>
<tr>
<td>MC05</td>
<td>D130/G131</td>
</tr>
<tr>
<td>MC06</td>
<td>S146/D147</td>
</tr>
<tr>
<td>MC07</td>
<td>D147/G148</td>
</tr>
<tr>
<td>MC08</td>
<td>G149/P149</td>
</tr>
<tr>
<td>MC09</td>
<td>K153/K154</td>
</tr>
<tr>
<td>MC10</td>
<td>E168/D169</td>
</tr>
<tr>
<td>MC11</td>
<td>D184/G185</td>
</tr>
<tr>
<td>MC12</td>
<td>A198/K199</td>
</tr>
<tr>
<td>MC13</td>
<td>K199/K200</td>
</tr>
<tr>
<td>MC14</td>
<td>Q203/L204</td>
</tr>
<tr>
<td>MC15</td>
<td>N220/E221</td>
</tr>
<tr>
<td>MC16</td>
<td>E221/D222</td>
</tr>
</tbody>
</table>

Figure 8; a hand-annotated map of amino acid sequence for the mCherry amino acid sequence translated from the gene in pCT from Geneious. Annotated with insertions sites (shaded arrow and vertical line), β-sheet (horizontal bracket and S#), and α-helix (horizontal bracket and α#) locations

Having identified the desired insertion sites for the TEV Protease recognition sequence, work in the nucleic acid domain could now begin. It is interesting and worth noting that Pavoor et al. insertion sites Asp-102/Asp-103 and at Glu-172/Asp-173 are roughly analogous to mCherry
insertion sites MC04 and MC11 respectively, with respect to the secondary and tertiary structure of the protein.

**Primer Design**

All primer design activity took place in Geneious. The TEV-PRS, having an amino acid sequence 7 residues in length, was coded for by a nucleic acid sequence 21 base pairs long. Site-Directed Mutagenesis was selected as the first method to insert the sequence into the yo-mCherry gene. This method requires a pair of complementary primers, comprised of a segment of the sequence immediately upstream of the insertion site, the site itself, and a segment immediately downstream of the insertion site.

![Figure 9](image1.png)

*Figure 9: Amino acid (bottom) sequence for the TEV-PRS to be inserted into the mCherry protein, as well as the amino acid sequence that codes for it.*

In generating the forward primers, 20 bp upstream of the insertion site, followed by the insertion, followed by 20 bp downstream were initially entered into Geneious. Since these primers will at no point be themselves translated, there is no need to maintain the codon reading frame, so any number of nucleotides can be present for the upstream and downstream binding regions. The lengths of these upstream and downstream binding regions were varied in order to achieve the desired predicted T_m value (melting temperature) of 55 °C, such that it would function optimally with the thermal cycle employed in the first mutagenesis approach; Site-Directed Mutagenesis. Other factors taken into consideration when designing the primers were minimizing ΔG and avoiding hairpin structures that could prevent annealing of primers to template DNA, which arise from complementary ends. Oligonucleotide primers were all ordered from Integrated DNA Technologies.

![Figure 10](image2.png)

*Figure 10: Forward (left) and Reverse (right) primers for MC03 in predicted thermostable conformation in Geneious; while there is a significant intrabonding occurring, the oligonucleotide conformation represented here have free ends, and can thus function as primers.*

The first four pairs primers ordered were the forward and reverse primers for insertion sites MC-[02,06,09,11].
The first mutation technique attempted was Site-Directed Mutagenesis (SDM) [15] [16]. This particular version of the process of mutation began with the yeast-optimized mCherry gene present in a pCT302 backbone. SDM using Phusion is more effective and less error-prone [17] on smaller plasmids, so transfer of the yeast-optimized mCherry gene from pCT (7 kb) to pN3 (4.5 kb) needed to be performed. Restriction digest was performed upon pCT containing the mCherry gene/insert from the pCT backbone and the same restriction sites were present in pN3, so the same restriction sites were used with pN3 to separate and purify the backbone. The isolated yo-mCherry gene was ligated into the pN3 backbone using T4 DNA Ligase.

The SDM was performed on the entire pN3 plasmid containing the yo-mCherry gene, and yielded the same plasmid with an added mutation. The primers used contain the sequence to be inserted, in this case the recognition site for TEV Protease, bracketed on either side by a sequence complimentary to the template DNA upstream and downstream of the insertion site. This primer and a complementary primer (referred to as Forward and Reverse primers) are used to

|| Primers | Sequence | F or R in front of insertion site name | Length (nm) |
|---|---|---|---|
| MC01 | AGGTTGACCTACGCTGGAAGGTAAGTATTTTTCGACGTTGACCCAAACGC | F | 25 nm |
| MC02 | GGGTTTGGTGGACCTTACCGTTGAATAACAACTTTTGGACGTTGACCCAAACGC | R | 25 nm |
| MC03 | GTATTTGATTTTTGACGTTGACCCAAACGC | F | 25 nm |
| MC04 | AGGTTGACCTACGCTGGAAGGTAAGTATTTTTCGACGTTGACCCAAACGC | R | 25 nm |
| MC05 | CCAAGTTGTTTATTTTTCGACGTTGACCCAAACGC | F | 25 nm |
| MC06 | GTATTTGATTTTTGACGTTGACCCAAACGC | R | 25 nm |
| MC07 | GTATTTGATTTTTGACGTTGACCCAAACGC | F | 25 nm |
| MC08 | GTATTTGATTTTTGACGTTGACCCAAACGC | R | 25 nm |
| MC09 | GTATTTGATTTTTGACGTTGACCCAAACGC | F | 25 nm |
| MC10 | GTATTTGATTTTTGACGTTGACCCAAACGC | R | 25 nm |
| MC11 | GTATTTGATTTTTGACGTTGACCCAAACGC | F | 25 nm |
| MC12 | GTATTTGATTTTTGACGTTGACCCAAACGC | R | 25 nm |
together. The complementary DNA bracketing the insertion allows for hybridization with the template DNA. Phusion DNA Polymerase is then used to extend the plasmid along the entire length of the template DNA. This is accomplished on both the forward and reverse strands, yielding complementary strands which are broken in different places. The template DNA can be eliminated by digestion using a DpnI restriction enzyme which specifically targets methylated DNA, so only DNA produced \textit{in vivo} is affected. This distinguishes the template from the mutated DNA because the template was originally produced in \textit{E. coli}, and will be methylated, while the newly synthesized DNA will not be. \cite{18} This can be easily amplified in \textit{E. coli} cells, wherein cloning of the plasmid will result in uncut circular dsDNA.

SDM was attempted using the primers ordered for insertion sites MC-[02,06,09,11]. The products were treated with \textit{DpnI} and heat-shock transformed into TOP10 competent \textit{E. coli} cells. These cells were plated on LB-Kanamycin. Since pN3 carries kanamycin resistance, colonies formed were assumed to carry the mutant plasmid. Two colonies were picked from each plate, incubated in liquid LB media, and the resulting cells miniprepped. The resulting plasmid DNA was sequenced at the UCONN Bioinformatics lab using Sanger sequencing. The results were analyzed in Geneious software.

The attempt at site MC06 failed, as unwanted mutations were found. That at site MC02 was lost due to low concentration of DNA. Sequencing of the mutated gene after attempting to insert at MC09 and MC11 showed the SDM to have successfully inserted the TEV-PRS.

\textbf{Figure 11; Visualization of Sanger sequencing results analyzed in Geneious software for attempted insertions at sites MC09 (top) and MC11 (bottom) using Site-Directed Mutagenesis. Visual is zoomed to the insertion region.}

Eight residues downstream, so instead of inserting at MC09 (K153/K154), the insertion was made at (S161/S162). This insertion site will be referred to as MC0X.

Due to the high failure rate of the attempted insertion mutations made using SDM, combined with the difficulty of involving another plasmid (pN3), a different mutagenesis method was chosen; OE-PCR.

\textbf{Overlap Extension PCR (OE-PCR)}

The second mutation method attempted was Overlap-Extension PCR (OE-PCR) \cite{11}. This method is significantly more laborious and resource-intensive, but more reliable, and can be performed successfully on genes in plasmids of greater size. The end product in this case is linear double stranded DNA, the limits of which are determined by the primers used. The same primers used in SDM can be used here, in conjunction with 2 more primers which flank the gene of interest on either side.
First, two PCR reactions are performed in parallel. Each consists of one of the SDM primers, and one of the bracketing, or “sequencing”, primers. The “pcr ykc” thermal cycle is carried out upon the template DNA along with DNA Polymerase and one of the two pairs of two plasmids, resulting in the exponential amplification of the sequence including and between each pair of primers. In this case, the forward sequencing primer for the pCT plasmid containing yeast optimized mCherry (yo-mCherry) was paired with the reverse primer for a specific insertion site, and the reverse sequencing primer with the forward primer for the insertion site. This produced 2 overlapping strands of DNA, which were complimentary only at the sites complimentary to the insertion site primers. Combining these overlapping DNA segments with DNA Polymerase and the forward and reverse sequencing primers in a “pcr ykc” thermal cycle results in the amplification of the entire double stranded region between the sequencing primers, including the mutation. The segment of DNA between the sequencing primer sites contained the protease recognitions sites for NheI and BSRGI, and was therefore suitable for ligation back into pCT.

OE-PCR was attempted at sites MC-[01,03,04,05,06,07,12,13,14]. Beginning with pCT-yo-mCherry, the first round of PCR was performed by pairing the forward pCT insert sequencing primer with the reverse mutagenic MC## primer, and the reverse pCT insert sequencing primer with the forward mutagenic MC## primer. The products of these respective PCR reactions were gel purified and recovered. The second round PCR was then conducted in duplicate, using the products of both first round PCR’s as well as the forward and reverse pCT insert sequencing primers. The resulting dsDNA was gel purified, recovered, and sequenced.
Sequencing reveals insertions at sites MC-[01,04,07,12,14] to have been successful without complication in either mutation or sequencing.

Figure 12: Visualization of Sanger sequencing results analyzed in Geneious software and aligned to the yo-mCherry gene for attempted insertions at sites (from top to bottom) MC-[01,04,07,12,14] using OE-PCR. Visuals are zoomed to the insertion region. Note that in the cases of MC-[04,05,07,12], the alignment of the last nuclotide before the insertion, a thymine, is misaligned to the last nucleotide of the insertion (also a thymine). This is only a software error, and does not reflect any problems with the sequencing.
Figure 13; Visualization of Sanger sequencing results analyzed in Geneious software and aligned to the yo-mCherry gene for attempted insertions at site MC06 using OE-PCR. A; The insertion site where a frameshift mutation appears to have occurred in the mutated sequence. B; the site of the apparent frameshift mutation. The presence of an extended, amorphous guanine peak indicates a sequencing assembly error. C; the site of the apparent frameshift mutation after the added guanine residue is manually removed. D; The insertion site where the apparent frameshift mutation has been repaired through manual deletion of the added guanine. (Note that in the cases of the MC06, the alignment of the last nucleotide before the insertion, a thymine, is misaligned to the last nucleotide of the insertion (also a thymine). This is only a software error, and does not reflect any problems with the sequencing.)

Figure 14; Visualization of Sanger sequencing results analyzed in Geneious software and aligned to the yo-mCherry gene for attempted insertions at site MC03 using OE-PCR. Top; TEV-PRS insertion site, and proximal location of an apparent unwanted insertion mutation has occurred, resulting in a frameshift mutation. Bottom; The same site following manual deletion of the unwanted 1bp insertion mutation, revealing an otherwise successful MC03 insertion.
Figure 15: Visualization of Sanger sequencing results analyzed in Geneious software and aligned to the yo-mCherry gene for attempted insertions at site MC05 using OE-PCR. While the mutation region itself (top) confirms a successful OE-PCR insertion, downstream on the gene, the results for region between residues I212 and T217 (inclusive) is extremely noisy. Based on the results, expected residue I216 was predicted by Geneious to be S216 because the second base in the codon was interpreted to be a guanine instead of a thymine. This base was the last one in a buildup of noise, and significant peaks were present for both thymine and guanine. This was likely a sequencing error due to contamination or presence of an air bubble.

Figure 16: Visualization of Sanger sequencing results analyzed in Geneious software and aligned to the yo-mCherry gene for attempted insertions at site MC13 using OE-PCR. While the mutation itself (top) confirms a successful OE-PCR insertion, just upstream of the insertion, a 2bp deletion mutation appears to have occurred. Additionally, upstream of the mutation, Geneious has predicted a thymine based upon sequencing results instead of the expected cytosine. There are peaks indicating both, and the base is flanked by two thymines on either side, so it is likely due to noise.
Expression

The successfully mutated plasmids needed to be expressed in surface display yeast to see if the mutant mCherry protein demonstrated fluorescence. The gene had been successfully inserted into pCT302, in which tryptophan synthase were also present. The purified plasmid was transformed into yeast EBY 100 surface display yeast using METHOD. Transformed yeast was incubated in SD-CAA media which is deficient in Tryptophan to select for successfully transformed cells. These cells were centrifuged to their media were switched to SG-CAA, which is identical except that it replaces glucose with galactose as a carbon source. The EBY100 yeast contains the GAL operon, as well as an inserted GAL-regulated AGA1 surface display anchor protein. The introduction of galactose induces the expression of this gene, as well as the expression of the GAL-linked aga2-mCherry complex in pCT. As the cells are incubated in the SG-CAA medium, the aga1 and aga2-mCherry complex are expressed, bond, and embed in the cell wall for surface display. These cells are imaged under a digital microscope and transferred to a computer where the fluorescence in the resulting images can be quantified and compared to a control.

Figure 17; Fluorescence imaging and analysis; A: fluorescence imaging for the accidental insertion that occurred in attempting MC09, now known as MC0X; B: fluorescence imaging from MC11; C: fluorescence imaging from the wild type yo-mCherry; D: digitally analyzed version of image 8.C, wherein a fluorescent expression value was determined for each cell.
Table 3; Results of fluorescence imaging and digital analysis, wherein samples 9A correspond to site MC0X

<table>
<thead>
<tr>
<th>Sample</th>
<th>Colony Count</th>
<th>Average</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>9A-1</td>
<td>72</td>
<td>209.331</td>
<td>60.265</td>
</tr>
<tr>
<td>9A-2</td>
<td>100</td>
<td>220.761</td>
<td>60.335</td>
</tr>
<tr>
<td>9A (avg)</td>
<td></td>
<td>215.046</td>
<td></td>
</tr>
<tr>
<td>11A-1</td>
<td>180</td>
<td>526.183</td>
<td>285.981</td>
</tr>
<tr>
<td>11A-2</td>
<td>156</td>
<td>467.523</td>
<td>245.713</td>
</tr>
<tr>
<td>11A (avg)</td>
<td></td>
<td>496.853</td>
<td></td>
</tr>
<tr>
<td>yoCh1</td>
<td>190</td>
<td>501.775</td>
<td>304.005</td>
</tr>
<tr>
<td>yoCH2</td>
<td>85</td>
<td>606.896</td>
<td>339.688</td>
</tr>
<tr>
<td>yoCh (avg)</td>
<td></td>
<td>554.3355</td>
<td></td>
</tr>
</tbody>
</table>

Figure 18; Graphical representation of the average fluorescence units per colony for the expressed yo-mCherry mutants and the control on 2 individual plates as well as the average for said plates.
Discussion

Mutation

Table 4: Mutation status for each site, wherein the numbers following “MC...” refer to the sites identified in Figure 7C

<table>
<thead>
<tr>
<th>Insertion</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC01</td>
<td>Successful mutation with OE-PCR</td>
</tr>
<tr>
<td>MC02</td>
<td>Progress lost before sequencing could be performed successfully</td>
</tr>
<tr>
<td>MC03</td>
<td>Possible Successful mutation with OE-PCR; Resequencing needed (see Figure 14)</td>
</tr>
<tr>
<td>MC04</td>
<td>Successful mutation with OE-PCR</td>
</tr>
<tr>
<td>MC05</td>
<td>Possible Successful mutation with OE-PCR; Resequencing needed (see Figure 15)</td>
</tr>
<tr>
<td>MC06</td>
<td>Possible Successful mutation with OE-PCR; Resequencing needed (see Figure 13)</td>
</tr>
<tr>
<td>MC07</td>
<td>Successful mutation with OE-PCR</td>
</tr>
<tr>
<td>MC08</td>
<td>Not attempted</td>
</tr>
<tr>
<td>MC09</td>
<td>Not attempted(^1)</td>
</tr>
<tr>
<td>MC10</td>
<td>Not attempted</td>
</tr>
<tr>
<td>MC11</td>
<td>Successful mutation with SDM</td>
</tr>
<tr>
<td>MC12</td>
<td>Successful mutation with OE-PCR</td>
</tr>
<tr>
<td>MC13</td>
<td>Apparent failed mutation with OE-PCR; Resequencing needed (see Figure 16)</td>
</tr>
<tr>
<td>MC14</td>
<td>Successful mutation with OE-PCR</td>
</tr>
<tr>
<td>MC15</td>
<td>Not attempted</td>
</tr>
<tr>
<td>MC16</td>
<td>Not attempted</td>
</tr>
<tr>
<td>MC0X</td>
<td>Successful mutation with SDM</td>
</tr>
</tbody>
</table>

![Figure 19; Visual map of mCherry structure annotated with the current status of each](image)

\(^1\) Sequencing revealed insertion to have been made at an incorrect site denoted with a “?” in Figure 19. The accidental insertion site was made due to a clerical error in primer design, and is present between two pleated sheets on the side of the cylindrical structure. This site is to be referred to as MC0X.
The mutations attempted were sequenced to find varying results. Site Directed Mutagenesis has yielded 2 confirmed successes and one confirmed failure. OE-PCR, however, has produced at least 5 confirmed successful mutations, likely as many as 8, and one apparent failure. Both methods, SDM and OE-PCR, have proven viable mutagenesis techniques. However, failures have been encountered in both cases. The failures may have occurred due to several causes. Occurrence of other mutations may be due to sequencing failure or due to the inherent likelihood of mutation during transcription at the low concentrations present during the beginning stages of thermal cycling. The actual presence of mutations can only be determined by resequencing.

Based upon the outcome of the experimentation described in this report, SDM has at best a much higher error rate than OE-PCR. This is due to the much larger amount of genetic data processed during the extension phase; while OE-PCR only requires the extension of a primer to an end point within the gene, in this case not more than 1 kb, SDM requires the primers to be extended around entire plasmid each cycle, over 4 kb.

With regard to procedure, both methods have significant advantages and disadvantages. SDM, in theory, does not require a ligation step; the entire mutation can be achieved through thermal cycling followed by amplification in bacteria and miniprep. In the case of this experiment, however, this advantage did not apply, due to constraints later on in the experiment. Both pCT and pN3 were needed, as the gene was initially present in pCT, and needed to be returned to pCT for expression in yeast, but SDM could not be performed directly on pCT due to its size. This added the double digest and ligation steps both before and after SDM was performed. pCT, on the other hand, is as labor intensive both in theory as in practice. The technique requires two independent PCR steps, each followed by gel purification and recovery, and finally double digest and ligation back into the pCT backbone.

Due to the significantly lower error rate and the similar amount of work required for both techniques under these circumstances, OE-PCR is the preferred technique for the insertion of the TEV-PRS into the to-mCherry gene followed by yeast surface display.

The insertion at site MC0X occurred due to a clerical error when designing primers. While not originally identified as a potential insertion site, the site satisfies some of the conditions laid out in the Theory section of this report, in that it does not disturb secondary structure. While it is not present in a loop connecting slats or helical core of the barrel-structure, it does connect two beta-sheets which make up a single slat, disrupting tertiary structure, but not secondary. Structurally, this mutation is very intriguing, as the conformation may be analogous to a broken slat in the barrel structure being propped outward, slightly warping the entire structure. The site appears plausible for further investigation.

**Expression**

The fluorescence results are the only quantitative results from the entire body of lab work. The sought after mutant, as described in the Theory section, would exhibit minimal or no fluorescence before cleavage, and significant fluorescence after cleavage. Of the two sites expressed, MC11 and MC0X, both exhibited fluorescence. MC11 displayed nearly fluorescence that was not statistically different from the wild-type, indicating that this insertion site is not
suitable for use as an *in vivo* protease sensor, as cleavage would at best result in a minor increase in fluorescence. The second site, MC0X, presented with a substantial reduction in fluorescence, of greater than 50%. While this is not as great a reduction as desired, it is currently the closest to the sought after mutant.

**Conclusion**

It is difficult to draw conclusions from an incomplete research project. Being that this thesis functions more as a progress report than a conclusive research paper, most of the take-away pertains to future steps. Regarding the *seventeen* identified sites, the current state of the research project is as follows:

- Neither of the *two* successfully expressed mutations exhibited the desired lack of fluorescence. One, however, which was not initially desired, exhibited a reduction in fluorescence of greater than 50%.
- At least *five* additional insertions have been successfully made and confirmed with sequencing, with an additional *three* that require resequencing to confirm sequencing error, and these are ready for expression in yeast.
- Progress on *two* insertion sites attempted has been lost due to procedural error. Primers for MC09 should be redesigned to target the intended site, and insertion MC02 should be attempted again using OE-PCR.
- Primers for the remaining *four* identified insertion sites have not been ordered, in addition to *one* redesigned pair of primers for site MC09.

In summary, moving forward, all successfully mutated sequences should be expressed in yeast. The remaining insertion sites should be attempted using OEPCR, as this technique requires comparable labor and is far more reliable. All successfully mutated versions of the yo-mCherry gene should be expressed in yeast, and then introduced to TEV protease to test for real-time *in vivo* viability as protease sensors. Finally, it may be interesting and useful to investigate the structure of the mutated mCherry-MC0x and others using X-Ray crystallography.

The results thus far indicate that the project shows promise, since the expressed mutations thus far have exhibited reductions in fluorescence. Other undergraduate research assistants in Dr. Yongku Cho’s lab will continue pursuit of the desired mutation. Research on the project thus far was performed under instruction of Dr. Cho by undergraduates Kathryn Frances, Krystyna Kapalczynski, Hamze Mah, Caroline Olesiak, Steven Rodriguez, Moe Uddin, and myself. Work will be carried forward by Hamze Mah, Kathryn Frances, and others in the lab, as all others once assigned to the project have since graduated.
Appendix

References


**Figures**

Figure 1; Visualization of the mCherry secondary and tertiary structure generated in PyMOL with an Educational License, with the color gradient from N to C terminus. .................................................................3

Figure 2; Visual map of mCherry structure wherein arrows represent β-pleated sheets and cylinders represent α-helices. ..............................................................................................................................................5

Figure 3; Cleavage sites for BsrGI-HF (left) and NheI-HF (right) from New England Bio Labs. .................................................5

Figure 4; Plasmid maps of pCT302 (left) and pN3 (right) from addgene .................................................................6

Figure 5; A; Visual representation of Site-Directed Mutagenesis used for insertion, derived from Geiser et al. [10]. The segmented loop represents the template plasmid (which is methylated, as it was synthesized in vivo), while the thick black segment represents the insertion. The primers are extended from the 3’ end along the template DNA until termination at their own 5’ end, resulting in a complete, nicked plasmid with the insertion, which is unmethylated since it was synthesized in vitro. Overnight digestion in DpnI restricts only the methylated DNA, leaving only the mutated plasmid to be transformed into E. coli and amplified. B; Thermal cycle diagram of “sdm phusion” cycle used for Site-Directed Mutagenesis. The cycle begins with a 60 second denaturation phase at 95°C, followed by a cycle of 50 second denaturation phase at 95°C, 50 second annealing phase at 65°C, and 180 second extension phase at 72°C. This 3-part cycle repeats itself 18 times, followed by an additional 300 second extension phase at 72°C, after which the cycler holds at 4°C indefinitely (not shown). ..............................................................................................................................................7

Figure 6; A; Visual representation of OE-PCR used for insertion, derived from Heckman et al. [11]. The double stranded DNA at the top represents the template DNA, a + d are flanking primers, and c + b are mutagenic primers. Flanking primer a pairs with mutagenic b, which are used together in the first round PCR reaction to make double stranded DNA fragment AB. Likewise, c + d pair to make CD. For the second round PCR reaction, dsDNA fragments AB and CD are combined along with flanking primers a + d. Denatured AB ssDNA anneals to the complementary CD region containing the insertion, and the fragments are extended to produce the final product of AD with the insertion.
B; Thermal cycle diagram of “pcr ykc” cycle used for both steps of Overlap-Extension PCR. The cycle begins with a 5 minute denaturation phase at 95°C, followed by a cycle of 30 second denaturation phase at 95°C, 30 second annealing phase at 55°C, and 45 second extension phase at 72°C. This 3-part cycle repeats itself 30 times, followed by an additional 300 second extension phase at 72°C, after which the cycler holds at 4°C indefinitely (not shown).

Figure 7; A; an amino acid sequence alignment of GFP to mCherry performed with EBI’s webPRANK, B; Visual map of secondary structure for Green Fluorescent Protein annotated with residue numbers at identifying points in secondary structure [14] and Pavoor et al. insertion sites at Asp-102/Asp-103 and at Glu-172/Asp-173 indicated with an ‘@’ and a ‘#’, respectively, C; Visual map of secondary structure for mCherry Protein annotated with numbered and labeled (plain black text) residues at identifying points in secondary structure as well as with the numbered (in black circles) and labeled (red text) original sixteen insertion sites.

Figure 8; a hand-annotated map of amino acid sequence for the mCherry amino acid sequence translated from the gene in pCT from Geneious. Annotated with insertions sites (shaded arrow and vertical line), β-sheet (horizontal bracket and S#), and α-helix (horizontal bracket and α#) locations.

Figure 9; Amino acid (bottom) sequence for the TEV-PRS to be inserted into the mCherry protein, as well as the amino acid sequence that codes for it.

Figure 10; Forward (left) and Reverse (right) primers for MC03 in predicted thermostable conformation in Geneious; while there is a significant intrabonding occurring, the oligonucleotide conformation represented here have free ends, and can thus function as primers.

Figure 11; Visualization of Sanger sequencing results analyzed in Geneious software for attempted insertions at sites MC09 (top) and MC11 (bottom) using Site-Directed Mutagenesis. Visual is zoomed to the insertion region.

Figure 12; Visualization of Sanger sequencing results analyzed in Geneious software and aligned to the yo-mCherry gene for attempted insertions at sites (from top to bottom) MC-[01,04,07,12,14] using OE-PCR. Visuals are zoomed to the insertion region. Note that in the cases of MC-[04,05,07,12], the alignment of the last nuclotide before the insertion, a thymine, is misaligned to the last nucleotide of the insertion (also a thymine). This is only a software error, and does not reflect any problems with the sequencing.

Figure 13; Visualization of Sanger sequencing results analyzed in Geneious software and aligned to the yo-mCherry gene for attempted insertions at site MC06 using OE-PCR. A; The insertion site where a frameshift mutation appears to have occurred in the mutated sequence. B; the site of the apparent frameshift mutation. The presence of an extended, amorphous guanine peak indicates a sequencing assembly error. C; the site of the apparent frameshift mutation after the added guanine residue is manually removed. D; The insertion site where the apparent frameshift mutation has been repaired through manual deletion of the added guanine. (Note that in the cases of the MC06, the alignment of the last nuclotide before the insertion, a thymine, is misaligned to the last nucleotide of the insertion (also a thymine). This is only a software error, and does not reflect any problems with the sequencing.)
Figure 14; Visualization of Sanger sequencing results analyzed in Geneious software and aligned to the yo-mCherry gene for attempted insertions at site MC03 using OE-PCR. Top; TEV-PRS Insertion site, and proximal location of an apparent unwanted insertion mutation has occurred, resulting in a frameshift mutation. Bottom; The same site following manual deletion of the unwanted 1bp insertion mutation, revealing an otherwise successful MC03 insertion.

Figure 15; Visualization of Sanger sequencing results analyzed in Geneious software and aligned to the yo-mCherry gene for attempted insertions at site MC05 using OE-PCR. While the mutation region itself (top) confirms a successful OE-PCR insertion, downstream on the gene, the results for region between residues I212 and T217 (inclusive) is extremely noisy. Based on the results, expected residue I216 was predicted by Geneious to be S216 because the second base in the codon was interpreted to be a guanine instead of a thymine. This base was the last one in a buildup of noise, and significant peaks were present for both thymine and guanine. This was likely a sequencing error due to contamination or presence of an air bubble.

Figure 16; Visualization of Sanger sequencing results analyzed in Geneious software and aligned to the yo-mCherry gene for attempted insertions at site MC13 using OE-PCR. While the mutation itself (top) confirms a successful OE-PCR insertion, just upstream of the insertion, a 2bp deletion mutation appears to have occurred. Additionally, upstream of the mutation, Geneious has predicted a thymine based upon sequencing results instead of the expected cytosine. There are peaks indicating both, and the base is flanked by two thymines on either side, so it is likely due to noise.

Figure 17; Fluorescence imaging and analysis; A: fluorescence imaging for the accidental insertion that occurred in attempting MC09, now known as MC0X; B: fluorescence imaging from MC11; C: fluorescence imaging from the wild type yo-mCherry; D: digitally analyzed version of image 8.C, wherein a fluorescent expression value was determined for each cell.

Figure 18; Graphical representation of the average fluorescence units per colony for the expressed yo-mCherry mutants and the control on 2 individual plates as well as the average for said plates.

Figure 19; Visual map of mCherry structure annotated with the current status of each.

Tables
Table 1; Insertion sites MC01-MC16 paired with their positions, derived from the map in Figure 6
Table 2; Forward and Reverse Primer Sequences for MC01-MC16 wherein the direction of the primer is indicated by the prefix F or R in front of the insertion site name
Table 3; Results of fluorescence imaging and digital analysis, wherein samples 9A correspond to site MC0X
Table 4; Mutation status for each site, wherein the numbers following “MC…” refer to the sites identified in Figure 7C.

Nomenclature
TEV – Tobacco Etch Virus
TEV-PRS - Tobacco Etch Virus Protease Recognition Site
SDM – Site Directed Mutagenesis
OE-PCR – Overlap Extension Polymerase Chain Reaction
PCR – Polymerase Chain Reaction
GFP – Green Fluorescent Protein
RFP – Red Fluorescent Protein
yo-mCherry – Yeast Optimized m(onomeric)Cherry