LUND UNIVERSITY

LTH

FACULTY OF ENGINEERING

Assessment of the Fluorescent Activities of Different Fluorescent Proteins for Potential Use as Biosensors in *Saccharomyces cerevisiae*

Johan Davidsson
2020-06-08

Master's Degree Project in Applied Microbiology
Course: KMBM05 (30 hp)
Spring 2020

Main supervisor: Marie Gorwa-Grauslund, Professor
Laboratory supervisor: Raquel Perruca Foncillas, Doctoral student
Examiner: Magnus Carlquist, Associate Professor

Division of Applied Microbiology, Department of Chemistry, Lund University

Abstract

Development of genetically modified microorganisms such as *Saccharomyces cerevisiae* for use as microbial cell factories to produce fine and bulk chemicals has lately proceeded very rapidly, much thanks to new gene editing methods based on a CRISPR/Cas9 system adapted from bacteria which increases the rate of genetic engineering. However, development of methods which enable fast and simple screening of the best performing strains with specific sought-after characteristics is not advancing at the same pace. Therefore, screening can be a bottleneck in the cell factory engineering process, slowing down the overall development time. In this context, biosensors where inducible promoters are coupled to reporter molecules are of interest because they allow for high throughput screening. As reporter molecules, fluorescent proteins are good candidates because of their non-invasiveness and easy detection through, for example, fluorescence imaging or flow cytometry. Biosensors for analyzing multiple properties in parallel require bright fluorescent proteins with several different colors. In this project, the fluorescent activities of fluorescent protein candidates were investigated by integrating the genes encoding the proteins into the genome of *S. cerevisiae* and analyzing the fluorescent activities using flow cytometry. The candidates yEGFP, mEGFP, CyOFP1, yeast codon optimized CyOFP1 and yeast codon optimized mBeRFP were bright enough to be distinguished from the background fluorescence, whereas the strain with the gene encoding smURFP could not be distinguished from the background fluorescence, not even after attempting to increase the fluorescence signals by supplying the strain with a cofactor necessary for smURFP fluorescence. Not all proteins were suitable to be combined due to fluorescence spillover. Two strains were generated, one with codon optimized CyOFP1 and mEGFP and one with codon optimized mBeRFP and mEGFP. The data for these strains suggested that bright signals derived from each fluorescent protein could be detected, showing promise for use as dual biosensors. More work is required to find suitable combinations of three or more fluorescent proteins for biosensor applications, and it is recommended that the method of color compensation is investigated to handle unavoidable spillover fluorescence.

Contents

1. Introduction 1
1.1 Project background and aim 1
1.2 Theoretical background 1
1.2.1 Using the CRISPR/Cas9 system for gene editing in yeast 1
1.2.2 Using flow cytometry to measure fluorescence intensity 2
1.2.3 Fluorescent protein candidates 4

2. Material and methods 7
2.1 Strains, cultivation media and storage 7
2.2 Construction of plasmids and DNA fragment 8
2.2.4 Construction of pRP006, pRP007, pRP008 and pRP009 9
2.2.5 Construction of pRP012 9
2.2.6 Construction of DNA fragment for yeast transformation with pRP012 10
2.3 Generation of single and double fluorescent protein strains 10
2.3.1 Generation of single fluorescent protein strains 10
2.3.2 Generation of double fluorescent protein strains 11
2.4 Generation of yeast strain expressing smURFP and HO-1 11
2.5 Aerobic cultivations in shake flasks 11

3. Results 13
3.1 Generation and evaluation of single fluorescent protein strains 13
3.1.1 Generation of single fluorescent protein strains 13
3.1.2 Growth characterization of single fluorescent protein strains in shake flasks 14
3.1.3 Fluorescent activities of fluorescent proteins expressed in *S. cerevisiae* 15
3.2 Generation and evaluation of strains carrying two fluorescent proteins 21
3.2.1 Strain construction 21
3.2.2 Growth characterization of the strains with double FPs shake flasks 23
3.2.3 Fluorescent activities of co-expressed fluorescent proteins in *S. cerevisiae* 23
3.3 Generation and evaluation of modified smURFP strain 24
3.3.1 Generation of modified smURFP strain 24
3.3.2 Growth characterization of modified smURFP strain 25
3.3.3 Fluorescent activity of modified smURFP strain 26
3.4 Difference in signal strength between different filter positions 27
3.5 Mixing samples of different strain populations 29

4. Discussion 31

Acknowledgements 35

References 36

Appendices 39

A. Primers used in the project 39

B. Yeast transformation protocol used 40

1. Introduction

1.1 Project background and aim

The yeast *Saccharomyces cerevisiae* is a microorganism utilized as a platform cell factory for industrial production of a range of different products including biofuels, chemicals and pharmaceuticals. There are a number of advantages to production in cell factories using *S. cerevisiae* as the host, one example being the reduced reliance on fossil resources to produce compounds of interest (Borodina and Nielsen, 2014). Genetic engineering of microbial cell factories is developing at a rapid pace, but the current methods to screen for the best performing cells and optimal genetic modifications can be costly and inefficient, slowing down the overall cell factory engineering process. One way to tackle this problem is to develop various types of biosensors by coupling relevant inducible promoters with reporter molecules, with fluorescent proteins being the major ones (D'Ambrosio and Jensen, 2017). The idea is that a specific promoter will be induced when a certain process occurs in a cell, activating the production of a fluorescent protein. This non-invasive method allows users to gain knowledge about what is going on in a cell by simply measuring the fluorescence of the proteins. Using several biosensors for simultaneously studying different properties or processes in a cell requires the use of several different colored fluorescent proteins. One of the challenges in developing biosensors with different colors is to find fluorescent proteins which can be integrated into organisms without losing their fluorescent properties. The need to test the properties of fluorescent proteins in the organisms in which they are intended to be used has been shown by Botman et al. (2019) who found that the practical *in vivo* brightness of a fluorescent protein can vary from the brightness determined *in vitro* or in other organisms. The differences in brightness may depend on how the proteins fold, mature and degrade *in vivo,* as this will have an impact on the overall levels of functional fluorescent proteins and therefore the levels of fluorescence.

In this project, the focus was to study several fluorescent proteins and assess if they had potential to be used in biosensors in *S. cerevisiae*. The main aim was to find candidate fluorescent proteins and test them by integrating the genes encoding the proteins into the genome of the yeast and assessing if, and potentially how strongly, the proteins fluoresce in the yeast. If two or more fluorescent proteins were found to work well, the next aim was to assess the possibilities of combining two or more proteins into one yeast strain and generating a yeast strain with the combination if suitable conditions were met. In doing so, it would mean taking another step towards the end-goal of having multiple biosensors with different colored fluorescent proteins in one yeast strain.

1.2 Theoretical background

1.2.1 Using the CRISPR/Cas9 system for gene editing in yeast

In this project, the CRISPR/Cas9 method developed by Jessop-Fabre et al. (2016) was used to integrate the fluorescent protein encoding genes into genomic sites of *S. cerevisiae*. This method is based on cleaving the genomic DNA and relying on the homology directed repair machinery in yeast to repair the cleavage using a DNA fragment with the gene to be integrated. The CRISPR/Cas9 system is originally a defense system for bacteria, which cleaves viral DNA, but has been adapted and engineered to make precise genome edits in other organisms, for example in yeast (Jinek et al., 2012). For the system to work for *S. cerevisiae*, the Cas9 protein needs to be expressed in the yeast. In this project, Cas9 was expressed on an episomal plasmid which replicates within the yeast cell. As Cas9 cannot target and cleave DNA at a specific site by itself, an RNA molecule termed single guide RNA (sgRNA) also needs to be expressed in the cell. The single guide RNA in

turn consists of two parts known as crRNA and tracrRNA (Jinek et al., 2012). crRNA is designed to recognize and base-pair with a specific DNA sequence where the user wants to cleave the DNA, while the tracrRNA acts like a scaffold holding together the gRNA and the Cas9. The sequence which crRNA is designed to recognize needs to be roughly 20 bp long and located right next to a PAM (protospacer adjacent motif) sequence which serves as a binding signal for the sgRNA-Cas9 complex (Jinek et al., 2012). The PAM sequence NGG, where N can be any nucleotide, is recognized by the Cas9 from *Streptococcus pyogenes,* which is the most commonly used Cas9 protein in CRISPR/Cas9 gene editing systems. After the sgRNA-Cas9 complex binds to the DNA strand, the crRNA sequence anneals to the targeted sequence and if there are sufficient matches Cas9 will cleave the DNA strand three base pairs upstream of the PAM sequence (Jinek et al., 2012). Cleavage results in a double stranded break (DSB), which the cell needs to repair to survive. There are two ways that yeast can do this. The first is by non-homologous end joining (NHEJ), which is error prone and the second is by homology directed repair (HDR) which instead is very precise (Ran et al., 2013). As errors are not desired during gene editing, the HDR mechanism is taken advantage of by providing the cell with donor DNA. As mentioned, this DNA fragment has the gene (or genes) to be integrated into the yeast genome, but it is also flanked by sequences homologous to the sequences upstream and downstream of where Cas9 has introduced a DSB. This allows the yeast to simply repair the break in its DNA by making use of the donor DNA with the homology regions as templates. In the process, the gene (or genes) of interest are integrated. Figure 1 shows a schematic overview of the sgRNA-Cas9 complex, where the DSB is introduced and the ways that the DNA can be repaired.

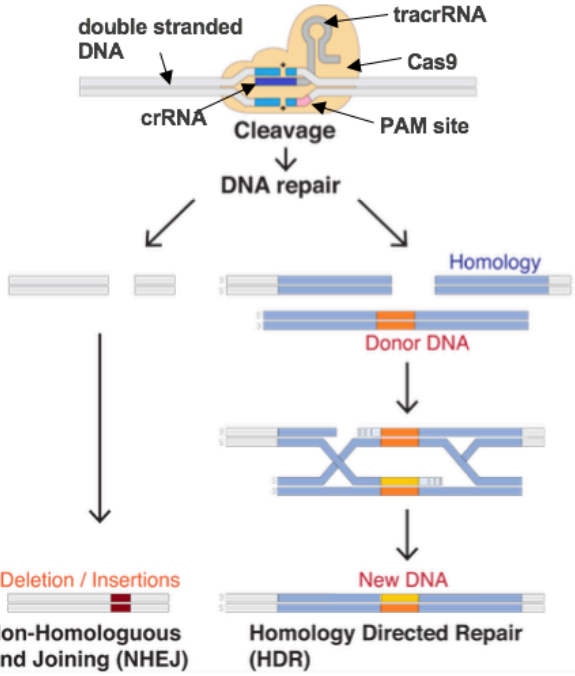

Figure 1. Overview of how the gRNA (consisting of crRNA and tracrRNA) binds to the Cas9 protein, where the cleavage occurs and the ways of repairing the double stranded break. The image is adapted from the image 'DNA Repair-colourfriendly.png' by Wikimedia Commons user Mariuswalter (2020).

1.2.2 Using flow cytometry to measure fluorescence intensity

Flow cytometry is a high-throughput technique for analysis of single cells in a sample. Flow cytometers are systems built using microfluidics in a specific way so that cells can be analyzed for different properties one by one. Briefly described, a sample inserted in a flow cytometer is transported through a series of tubes to a chamber where the cells in the sample are focused in a single file line through a phenomenon known as hydrodynamic focusing, before reaching a point

where they are analyzed. At this point, known as the interrogation point, cells are exposed to monochromatic light (lasers) and light is scattered in different directions, as well as absorbed by certain molecules in the cells. Absorbed light is emitted as fluorescence which is also scattered in different directions. Light and fluorescence which has been scattered is detected and quantified, giving information to the user about the properties of the analyzed sample. Since cells are analyzed at single cell-level, data is recorded for each cell in a sample (Givan, 2011).

In this project, samples of yeast cells expressing fluorescent proteins were used and therefore the analyzed property was the fluorescence intensity of the cells. For ease of data handling and analysis, the mean fluorescence intensities of all cells in the samples were analyzed.

The flow cytometer used in this project was equipped with two lasers, known as excitation lasers. One laser strikes samples with light at a wavelength of 488 nm (blue) and the other laser strikes samples with light at a wavelength of 640 nm (red)(BD Biosciences, 2015). Both of these lasers (also referred to as blue and red lasers) were used to excite the fluorescent proteins in the cells at two different interrogation points (Givan, 2011). The flow cytometer could detect the scattered fluorescence in four different positions, referred to as FL1, FL2, FL3 and FL4 (Figure 2A). A very important note is that fluorescence scattered as a result of excitation by the blue laser (488 nm) could only be detected in the FL1, FL2 and FL3 positions, whereas fluorescence scattered as a result of excitation by the red laser (640 nm) could only be detected in the FL4 position (BD Biosciences, 2014). This had certain implications when analyzing data from the different fluorescent proteins.

Although light of a specific wavelength (488 nm and 640 nm in this case) was used to excite the cells, the emitted fluorescence was not of just one wavelength. Fluorescence was instead emitted in a spectrum of wavelengths. However, when quantifying the fluorescence intensity of a protein, it was not practical to quantify the fluorescence of the entire spectrum. Therefore, so-called emission filters were used to filter out fluorescence which did not come from a peak in the spectrum (bandpass filters) or to filter out fluorescence of all wavelengths in a spectrum which were shorter than a specific wavelength (longpass filters). The idea was essentially to limit the detection to the most intense fluorescence signals. Using emission filters thereby allowed specification of in what range of wavelengths fluorescence should be detected. To exemplify, if a user wants to detect the fluorescence intensity of a green fluorescent protein, the user would use an emission filter which allows fluorescence in the green light range (wavelengths of roughly 520-560 nm) to pass. In the equipment used in this project, one emission filter could be placed in each of the four positions in the flow cytometer (Figure 2B). For example, one emission filter could be placed in the FL1 position to allow fluorescence in the green light range to pass and be detected in that position, another filter could be placed in the FL2 position allowing orange fluorescence to pass and a third filter could be placed in the FL3 position allowing red fluorescence to pass. In this way, it was for example possible to see if an analyzed green fluorescent protein firstly fluoresced green light as expected, but also if it fluoresced orange and red light as well. In the FL4 position, only filters allowing fluorescence in the red light range (roughly 635-700 nm) could be placed, since fluorescence scattered as a result of excitation by the red laser (640 nm) can never be of shorter wavelengths than 640 nm.

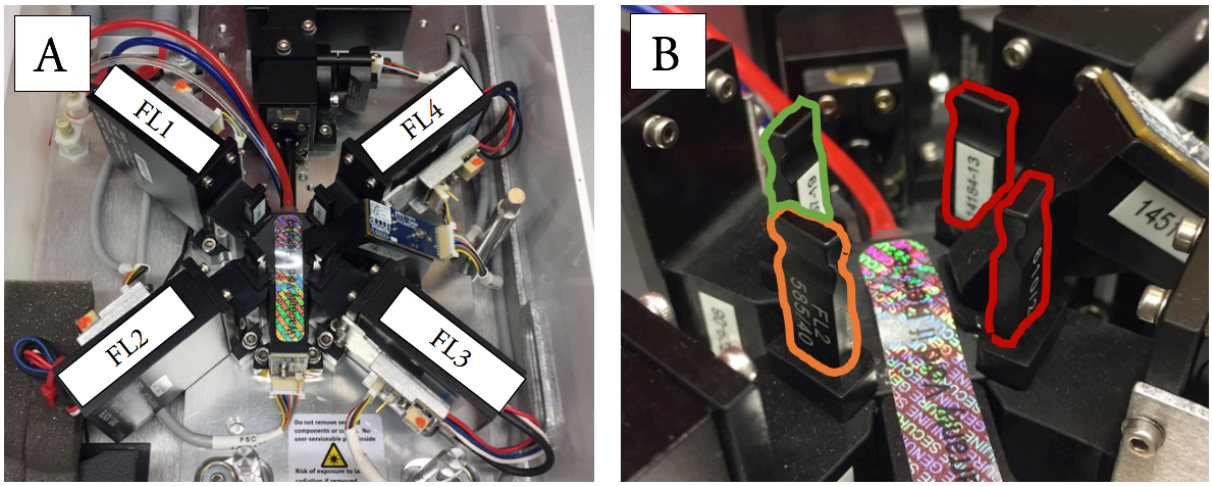

Figure 2. Pictures showing the inside of the flow cytometer used, with the FL1-FL4 positions marked in A and the filters outlined in green, orange and red in B. Pictures were taken by the author.

Fluorescence signals from a protein detected in a different light range than intended, for example detection of fluorescence from a green fluorescent protein in the orange range, is known as spillover. In a multicolor biosensor setup, spillover is undesired since it leads to interference between signals from different proteins. For example, if a green fluorescent protein is expressed under a certain condition, green fluorescence will be detected in the green light range; but orange fluorescence could also be detected in the orange range from the green fluorescent protein if the fluorescence from the green fluorescent protein is too strong. The user will then not know if the signals detected in the orange range are originating from the green fluorescent protein or from an orange fluorescent protein and will be unable to link the measured fluorescence to which biosensor is giving the signal. As exemplified, fluorescence spillover can make the biosensor system useless and therefore it is desired that fluorescence from only one protein is detected in each light range. It should be mentioned that a mathematical method known as color compensation can be applied to correct fluorescence signals by removing fluorescence spillover so that detected fluorescence in one light range only originates from one specific protein (Abcam plc and Davies, 2020). This is preferably done using software, however not enough time was available during the course of this project to fully explore this possibility.

1.2.3 Fluorescent protein candidates

When choosing fluorescent protein candidates for use in biosensors, the general idea is that the higher brightness, the better. When speaking about fluorescent proteins, brightness is a term for the product of the extinction coefficient (how much light the protein absorbs) and the quantum yield (the ratio of how many photons are emitted per photon absorbed) (Cranfill et al., 2016). However, the process is not as straightforward as just picking the brightest proteins. Each fluorescent protein has a different excitation and emission spectra. There are two ways in which these spectra have to be taken into consideration. Primarily, to theoretically reach as high fluorescence intensities as possible, the wavelengths at which the excitation spectra have their peaks should ideally be as close to the wavelengths of the excitation lasers in the equipment as possible. For the same reason, the peaks of each emission spectrum should ideally fit well with one of the available emission filters. Overlapping emission spectra of the fluorescent proteins is also something to take into consideration. It is desired to minimize overlaps to decrease the risk of fluorescence spillover and thereby increase the chance that two or more of the proteins can be used in combination. In this project however, the excitation laser and emission filter options were limited

to the equipment available in the lab, limiting the numbers of reasonable fluorescent proteins to choose from and thereby the possibilities of minimizing spectral overlaps.

Prior to the project, the database FPbase was searched based on these criteria and priorities, and the following fluorescent protein candidates were chosen: CyOFP1 (Cyan-excitable Orange Fluorescent Protein), smURFP (small Ultra-Red Fluorescent Protein), opt. CyOFP1 (yeast codon optimized CyOFP1) and opt. mBeRFP (yeast codon optimized monomeric Blue light-excited Red Fluorescent Protein) (Lambert, 2019). These proteins are listed in Table 1, in addition to selected relevant properties. Two more fluorescent proteins are listed as these had previously been integrated into the genome of *S. cerevisiae* and were also assessed in the project. These proteins were yEGFP (yeast-Enhanced Green Fluorescent Protein) and mEGFP (monomeric Enhanced Green Fluorescent Protein).

Table 1. List of investigated fluorescent proteins and selected properties. ***Note:*** *Data for yEGFP is taken from EGFP since no entries have been made for yEGFP in the database FPbase.*

Name	Max. excitation wavelength (nm)	Max. emission wavelength (nm)	Brightness	Reference
CyOFP1	497	589	30.4	(Chu et al., 2016)
smURFP	642	670	32.4	(Rodriguez et al., 2016)
opt. CyOFP1	497	589	30.4	(Chu et al., 2016)
opt. mBeRFP	446	611	17.55	(Yang et al., 2013)
yEGFP	488	507	33.54	(Cormack et al., 1997)
mEGFP	488	507	33.6	(Cormack, Valdivia and Falkow, 1996; Zacharias et al., 2002)

The excitation spectra of the candidate proteins are shown in Figure 3A with vertical lines indicating the blue and red excitation lasers used. The spectra for CyOFP1 and opt. CyOFP1 are assumed to be the same and the spectrum for EGFP is used to represent both yEGFP and mEGFP as they are also assumed to be the same. The emission spectra are shown in Figure 3B with different colored bars indicating the bandpass filters used in the project (510/15 nm, 585/40 nm, 610/20 nm and 675/25 nm from left to right) and a gray vertical line, connecting to a gray horizontal line at the top indicating the longpass filter used (670 LP, meaning all wavelengths over 670 nm). It should be clarified that the filter numbers indicate the center of the range of fluorescence wavelengths allowed to pass through the filter, along with the total width of the range. For example, the 510/15 nm filter allows fluorescence signals in the range of 510 ± 7.5 nm to pass through.

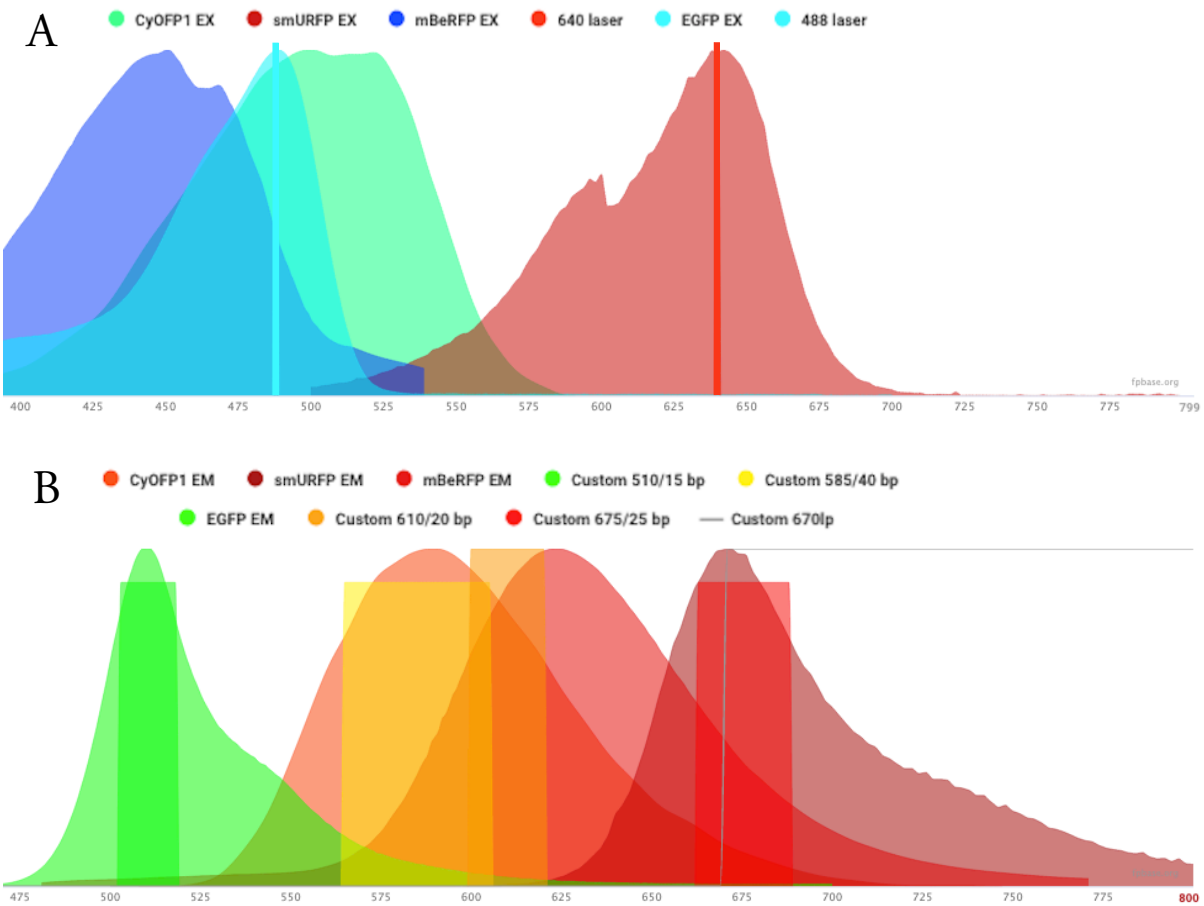

Figure 3. Excitation spectra along with excitation lasers used (A) and emission spectra along with emission filters used (B). The images are screenshots of graphs constructed using the spectra viewer tool on FPbase (Lambert, 2019).

The choices of filters to use for detection of the respective fluorescent proteins were based on the emission spectra. Signals from the green fluorescent proteins yEGFP and mEGFP were expected to be detected using the 510/15 nm filter, signals from the orange fluorescent proteins CyOFP1 and opt. CyOFP1 using the 585/40 nm filter and signals from the red fluorescent protein opt. mBeRFP mainly using the 610/20 nm filter. However, since the 610/20 nm filter did not align as well with the emission peak of opt. mBeRFP as the other filters aligned with the emission peaks of the green and orange fluorescent proteins, the 670 LP filter was considered as an alternative filter for detection of signals from opt. mBeRFP. However, this filter was clearly not ideal either. These filters were placed in the FL1-FL3 positions in a suitable manner since the fluorescence signals detected originated from proteins excited by the blue laser in the flow cytometer. Signals from the last protein, the red fluorescent protein smURFP, were expected to be detected using the 675/25 nm filter in the FL4 position since the protein was excited by the red laser and the resulting fluorescence signals could only be detected in the FL4 position.

2. Material and methods

2.1 Strains, cultivation media and storage

Ten strains of *Saccharomyces cerevisiae* with at least one chromosomally integrated fluorescent protein gene were used and/or constructed in this project (Table 2). All strains were based on the CEN.PK113-7D strain of *S. cerevisiae.* Liquid cultivations of the yeast strains were performed in yeast peptone dextrose (YPD) medium containing peptone from casein 20 g/L, yeast extract 10 g/L, glucose 20 g/L. Plate cultivations were also performed on YPD, solidified by addition of 15 g/L agar-agar. Geneticin (200 mg/L) was added in all yeast cultivations to maintain selection pressure for the *Cas9-kanMX* plasmid, except for when cultivating for growth characterization and flow cytometry measurements. On selection plates after yeast transformations with the CRISPR/Cas9 system, nourseothricin (100 mg/L) was used to select for the gRNA-*natMX* plasmid, in addition to geneticin (200 mg/L) to select for the *Cas9-kanMX* plasmid. After yeast transformation with a multicopy plasmid, Aureobasidin A (0.3 mg/L) was used to select for the multicopy plasmid, in addition to geneticin (200 mg/) to select for the *Cas9-kanM X* plasmid. Pre-cultures (also referred to as overnight cultures) of the yeast strains were always performed in 50 mL falcon tubes, inoculating 5 mL of YPD with cells from single colonies on plates. All liquid yeast cultivations were carried out at 30°C and 180 rpm in a New Brunswick™ Innova® 43 incubator (Eppendorf Nordic A/S, Hørsholm, Denmark).

Table 2. List of yeast strains used in the project.

Name	Relevant genotype	Fluorescent protein	Reference
CEN.PK113-7D + Cas9	CEN.PK113-7D; *pTEF1p-Cas9-CYC1t_kanMX* (Background strain)	-	Raquel Perruca Foncillas (unpublished)
CEN.PK+pRFP2	CEN.PK113-7D + Cas9; XI-3:: *TEF1p-mEGFP-ADH1t*	mEGFP	Raquel Perruca Foncillas (unpublished)
TMB RP 001	CEN.PK113-7D + Cas9; XI-3:: *TEF1p-yEGFP3-ADH1t*	yEGFP	Raquel Perruca Foncillas (unpublished)
TMB RP 002	CEN.PK113-7D + Cas9; XI-3:: *TEF1p-CyOFP1-ADH1t*	CyOPF1	This project
TMB RP 003	CEN.PK113-7D + Cas9; XI-3:: *TEF1p-smURFP-ADH1t*	smURFP	This project
TMB RP 004	CEN.PK113-7D + Cas9; XI-3:: *TEF1p-opt.CyOFP1-ADH1t*	opt. CyOFP1	This project
TMB RP 005	CEN.PK113-7D + Cas9; XI-3:: *TEF1p-opt.mBeRFP-ADH1t*	opt. mBeRFP	This project
TMB RP 008	TMB RP 004; XI-2:: *TEF1p-mEGFP-ADH1t*	opt. CyOFP1 + mEGFP	This project
TMB RP 009	TMB RP 005; XI-2:: *TEF1p-mEGFP-ADH1t*	opt. mBeRFP + mEGFP	This project
TMB RP 010	TMB RP 003; *pTDH3p-HO-1-CYC1t_AUR1-C*	smURFP	This project

Escherichia coli NEB 5-alpha (New England Biolabs, Ipswich, MA, USA) cells were used for subcloning experiments. Liquid cultivations of *E. coli* were performed in lysogeny broth (LB) medium containing tryptone 10 g/L, yeast extract 5 g/L, NaCl 5 g/L, pH 7.0. Plate cultivations of *E. coli* were performed on LB, solidified by addition of 15 g/L agar-agar. Ampicillin (50 mg/L) was used in plates and in all *E. coli* cultivations to select for successful transformants (plates) and maintain selection pressure for transformants (liquid cultivations). Pre-cultures of *E. coli* were also

always performed in 50 mL falcon tubes, inoculating 5 mL of LB with cells from single colonies on plates. All liquid *E. coli* cultivations were carried out at 37°C on a tilting table.

To store generated yeast and *E. coli* strains for future use, cells were grown overnight as described above and frozen in glycerol (25% v/v final concentration) at -80°C. PCR amplifications were performed using Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Waltham, MA, USA), except for colony PCRs which were performed with DreamTaq DNA Polymerase (Thermo Scientific). All restriction enzymes used were of the FastDigest product line (Thermo Scientific).

2.2 Construction of plasmids and DNA fragment

The plasmids used and constructed in this project are listed in Table 3. Five plasmids were constructed by the author, four of which were made to carry the genes of the fluorescent proteins of interest. These four plasmids were used for yeast transformations using the CRISPR/Cas9 system developed by Jessop-Fabre et al. (2016). The last plasmid was made to carry the gene encoding an enzyme which theoretically catalyzes a reaction producing the cofactor biliverdin, required by one of the proteins for fluorescence. In addition to the plasmids, one DNA fragment was constructed to carry the gene conferring resistance to Aureobasidin A. The plasmids and the DNA fragment were constructed using common molecular biology methods described in the sections below.

Table 3. List of plasmids used and constructed in this project.

Name	Relevant genotype	Reference
pCfB2903	XI-2 MarkerFree backbone	(Jessop-Fabre et al., 2016)
pCfB2904	XI-3 MarkerFree backbone	(Jessop-Fabre et al., 2016)
pCfB3044	gRNA_XI-2; *natMX*	(Jessop-Fabre et al., 2016)
pCfB3045	gRNA_XI-3; *natMX*	(Jessop-Fabre et al., 2016)
pKK-TEV-CyOFP1	*CyOFP1*	(Szczesny et al., 2018)
pBAD-smURFP-RBS-HO-1	*smURFP; HO-1*	(Rodriguez et al., 2016)
pUC57+opt. CyOFP1	*opt. CyOFP1*	Ordered from GenScript Europe
pUC57+opt. mBeRFP	*opt. mBeRFP*	Ordered from GenScript Europe
pRP005	pCfB2904; *TEF1p-yEGFP3-ADH1t*	Raquel Perruca Foncillas (unpublished)
pRP006	pCfB2904; *TEF1p-CyOFP1-ADH1t*	This study
pRP007	pCfB2904; *TEF1p-smURFP-ADH1t*	This study
pRP008	pCfB2904; *TEF1p-opt.CyOFP1-ADH1t*	This study
pRP009	pCfB2904; *TEF1p-opt.mBeRFP-ADH1t*	This study
pRPF2	pCfB2903; *TEF1p-mEGFP-ADH1t*	Raquel Perruca Foncillas (unpublished)
pRS42N	*natNT2*	(Taxis and Knop, 2006)
p426-GPD	*TDH3p-CYC1t; URA3*	(Mumberg, Müller and Funk, 1995)
pUG62AUR	*AUR1p-AUR1-C-AUR1t*	(Bergdahl et al., 2013)
pRP012	p426-GPD; *TDH3p-HO-1-CYC1t*	This study

2.2.4 Construction of pRP006, pRP007, pRP008 and pRP009

The four plasmids pRP006 (CyOFP1), pRP007 (smURFP), pRP008 (opt. CyOFP1) and pRP009 (opt. mBeRFP) were constructed to carry the genes for one fluorescent protein each. In short, this was done by ligating the genes encoding CyOPF1, smURFP, yeast codon optimized CyOFP1 (opt. CyOFP1) and yeast codon optimized mBeRFP (opt. mBeRFP), respectively, into the pRP005 backbone, between the *TEF1p* promoter and *ADH1t* terminator.

The genes encoding CyOFP1 and smURFP were amplified from the plasmids pKK-TEV-CyOFP1 and pBAD-smURFP-RBS-HO-1 (both obtained from Addgene Europe, Teddington, UK) respectively by performing a PCR. During the PCR, restriction sites for the restriction enzymes *Xho*I and *Sfa*AI were added to the beginnings and ends of the genes. Primers 31/32 were used for amplification of CyOFP1 and primers 27/28 were used for amplification of smURFP (Table A1, Appendix A). The products were run through a 0.8 % agarose gel electrophoresis to verify successful amplifications. The PCR products were purified using GeneJET PCR Purification Kit (Thermo Scientific) and digested with *Xho*I and *Sfa*AI yielding ligatable fragments. The genes for opt. CyOFP1 and opt. mBeRFP were extracted from the pUC57+opt. CyOFP1 and pUC57+opt. mBeRFP plasmids (both obtained from GenScript Europe, Leiden, the Netherlands) by digesting the plasmids with *Xho*I and *Sfa*AI, running the digestion mixtures through 0.8 % agarose gel electrophoresis for separation, followed by purification of the correct bands from the gel using a GeneJET Gel Extraction Kit (Thermo Scientific).

The pRP005 plasmid was digested with *Xho*I and *Sfa*AI. The digestion mixture was run through a 0.8 % agarose gel electrophoresis for separation and the pRP005 backbone was purified from the gel. The digested fluorescent protein genes were ligated with the digested pRP005 backbone using T4 DNA ligase (Invitrogen), generating the pRP006, pRP007, pRP008 and pRP009 plasmids carrying the CyOFP1, smURFP, opt. CyOFP1 and opt. mBeRFP encoding genes, respectively. *E. coli* NEB 5-alpha were transformed with the plasmids in accordance with the NEB 5-alpha High Efficiency Transformation Protocol C2987H (New England Biolabs), except for step 3 where the mixture was placed on ice for 45 minutes instead of 30 minutes. Colony PCR was performed to verify the transformants. Primers 31/32 were used for pRP006 (CyOFP1) transformants, primers 27/28 for pRP007 (smURFP) transformants, primers 33/34 for pRP008 (opt. CyOFP1) transformants and primers 35/36 for pRP009 (opt. mBeRPF) transformants (Table A1, App. A). Overnight cultures of two positive transformants per plasmid were used for saving the constructed plasmids in glycerol stock and for plasmid extraction which was performed with a GeneJET Plasmid Miniprep Kit (Thermo Scientific). Extracted plasmids were sent for sequencing. Seq1 was used as forward sequencing primer and seqADH1t was used as reverse sequencing primer (Table A1, App. A).

2.2.5 Construction of pRP012

The pRP012 multicopy plasmid which carries the HO-1 gene encoding heme oxygenase-1 from the cyanobacterium *Synechocystis* and the *URA3* selection marker was constructed by ligating *HO-1* into the backbone of the p426-GPD vector, between the *TDH3p* promoter and *CYC1t* terminator.

HO-1 was amplified from the pBAD-smURFP-RBS-HO-1 plasmid by performing a PCR, adding restriction sites for *Bam*HI to the beginning and *Eco*RI to the end of the gene. Primers 41/42 were used for amplification (Table A1, App. A). The product was purified and digested with *Bam*HI and *Eco*RI. At this stage the product was run through agarose gel electrophoresis to verify successful

amplification. The p426-GPD plasmid was digested using *Bam*HI and *Eco*RI and the digestion mixture was run through agarose gel electrophoresis to verify successful digestion. The digested HO-1 gene was ligated into the digested p426-GPD backbone, generating the pRP012 plasmid.

E. coli transformation, verification and saving of transformants, plasmid extraction and preparation for sequencing was performed in the same way as described for the pRP006-pRP009 plasmids in section 2.2.1, but with primers 41/42 for verification of the *E. coli* transformants and primers 117/118 as forward and reverse sequencing primers (Table A1, App. A).

2.2.6 Construction of DNA fragment for yeast transformation with pRP012

A DNA fragment was constructed to replace the *URA3* marker with the *AUR1-C* gene conferring resistance to Aureobasidin A (along with the *AUR1p* promoter and the *AUR1t* terminator) in pRP012 using *in vivo* plasmid recombination (Oldenburg et al., 1997). This was done by amplifying the *AUR1p-AUR1-C-AUR1t* cassette from the pUG62AUR plasmid by performing a PCR, adding up- and downstream homology to regions flanking the URA3 gene on the p426-GPD plasmid. Primers 43/44 were used for amplification (Table A1, App. A). The PCR product was run through agarose gel electrophoresis to verify successful amplification and then purified.

2.3 Generation of single and double fluorescent protein strains

Seven strains of *S. cerevisiae* were generated in this project. Yeast strains TMB RP 002, TMB RP 003, TMB RP 004 and TMB RP 005 were generated based on the CEN.PK113.7D + Cas9 strain using a CRISPR/Cas9 system, integrating the genes for CyOFP1, smURFP, opt. CyOFP1 and opt. mBeRFP in the XI-3 genomic locus in the respective strains. The strains TMB RP 008 and TMB RP 009 were generated using the same CRISPR/Cas9 system but based on the TMB RP 004 (opt. CyOFP1) and TMB RP 005 (opt. mBeRFP) strains respectively, integrating the gene for mEGFP in the XI-2 genomic locus in both strains. Strains TMB RP 002-005 are referred to as single fluorescent protein strains and TMB RP 008-009 are referred to as double fluorescent protein strains.

2.3.1 Generation of single fluorescent protein strains

The CRISPR/Cas9 system developed by Jessop-Fabre et al. (2016) was utilized to transform the CEN.PK113-7D strain of *S. cerevisiae* according to the protocol available in Appendix B, which was based on the LiAc/SS Carrier DNA/PEG method (Gietz and Schiestl, 2007). Donor DNA was constructed by linearizing the plasmids pRP006 (CyOFP1), pRP007 (smURFP), pRP008 (opt. CyOFP1) and pRP009 (opt. mBeRFP) using *Not*I. Cas9 was recruited to the XI-3 genomic locus by the guide RNA from pCfB3045 to cleave the DNA, making a double stranded break and allowing the endogenous homology directed repair machinery to integrate the donor DNA into the XI-3 sites. This resulted in four strains, each with one chromosomally integrated copy of a fluorescent protein.

Verification of transformants was done by colony PCR in two steps. First, primers annealing directly on the fluorescent protein genes were used to verify that the genes were present in the transformants. In this step, primers 31/32 were used for CyOFP1 transformants, primers 27/28 for smURFP transformants, primers 33/34 for opt. CyOFP1 transformants and primers 35/36 for opt. mBeRPF transformants (Table A1, App. A). The second time, primers annealing in the XI-3 regions up- and downstream of the integration site were used to verify that the genes were integrated in the correct site. In this step, primers 23/24 were used for all transformants (Table A1, App. A).

Two positive transformants per generated strain were saved in glycerol stock for future characterization.

2.3.2 Generation of double fluorescent protein strains

Further modifications were made to two of the single fluorescence protein strains, TMB RP 004 (opt. CyOFP1) and TMB RP 005 (opt. mBeRFP). The gRNA plasmid pCfB3045 was removed from both strains by performing replica plating. Once again, the CRISPR/Cas9 system developed by Jessop-Fabre et al. (2016) was utilized for transformation, with the *Not*I linearized pRPF2 plasmid (mEGFP) as donor DNA and the pCfB3044 plasmid expressing guide RNA directing Cas9 to the XI-2 genomic locus. The donor DNA could again be integrated into the genome using the endogenous homology directed repair system. This resulted in two strains, each with one copy of *mEGFP* in the XI-2 site in addition to the previously integrated fluorescent proteins in the XI-3 sites in the respective strains.

Verification of transformants was done by colony PCR in one step this time, using a primer which annealed inside the mEGFP gene and another primer which annealed in the XI-2 region downstream of the integrated *mEGFP* and flanking *ADH1t* terminator. Primers 13/324 were used for this (Table A1, App. A). The method allowed verification of both integration of the correct gene and correct site for the integration in one step. Two positive transformants per generated strain were saved in glycerol stock for future characterization.

2.4 Generation of yeast strain expressing smURFP and HO-1

The last generated yeast strain was based on the TMB RP 003 (smURFP) strain but not transformed using a CRISPR/Cas9 system. Instead this strain was transformed with two linear DNA fragments, the pRP012 plasmid digested with *Stu*I in the *URA3* sequence and the DNA fragment with the *AUR1p-AUR1-C-AUR1t* cassette flanked by *URA3* up- and downstream homology regions. This method aimed to make use of the plasmid recombination mechanism to rebuild a multicopy plasmid *in vivo,* with *URA3* exchanged for the *AUR1p-AUR1-C-AUR1t* cassette. Aureobasidin A was used for selection.

Colony PCR was performed to verify the transformants. Verification of the presence of *HO-1* was performed with primers 41/42 annealing to the beginning and end of the gene (Table A1, App. A). Overnight cultures were inoculated and the following day the overnight cultures were used to save the strain in glycerol stock.

2.5 Aerobic cultivations in shake flasks

The process of growth characterization of different strains and clones was done in parallel with flow cytometry measurements. Samples taken to determine the optical density at 620 nm (OD_{620}) were also used as samples in the flow cytometry machine.

Shake flask cultivations in YPD were performed in 250 mL baffled shake flasks using a liquid volume of 25 mL. Pre-cultures of strains to be analyzed were obtained from isolated colonies on plates and used to inoculate the YPD to an initial OD_{620} of 0.5. Antibiotics were in general not added to these cultivations, since losing the Cas9 or gRNA plasmids was assumed not to have any notable effect on the growth or fluorescence of the strains. One exception was made when cultivating the TMB RP 010 (smURFP+biliverdin) strain, since selection pressure was required for maintaining the multicopy plasmid with the HO-1 gene in the strain. For these cultivations Aureobasidin A was added to a final concentration of 0.3 mg/L.

After inoculation, a sample was taken to measure the fluorescence at the start of cultivation. The shake flasks were then placed in a shaking incubator at 30°C, 180 rpm. Following samples were taken at 2, 3, 4, 5, 6, 7 and 24 hours of cultivation time. Optical density of the samples was measured at 620 nm using an Ultrospec 2100 Pro spectrophotometer (Amersham Biosciences Corp., USA).

Flow cytometry was carried out on a BD Accuri C6 flow cytometer, equipped with an automatic BD CSampler (BD Biosciences, San Jose, CA, USA). Prior to conducting measurements on samples, a quality control to ensure proper functionality of the flow cytometer was performed in accordance with the standard procedure of the lab. Also, some of the standard emission filters were exchanged for others, depending on the goals of each run. The emission filters used throughout all runs were 510/15 nm, 585/40 nm, 610/20 nm, 670 LP and 675/25 nm (all purchased from BD Biosciences). The 488 nm and 640 nm excitation lasers were used for excitation of the samples.

Samples were diluted to $OD_{620} < 1.0$ as necessary in phosphate-buffered saline (PBS), vortexed and loaded in the autosampler. Samples were run with the following settings:

- Run limits:
    - 10 000 events
    - 150 µL
- Speed: Medium (35 µL/min)
- FSC-H: Less than 80 000
- Wash settings: 1 cycle

Data was overviewed in the BD CSampler Analysis Software (BD Biosciences). Further data analysis was performed in FlowJo v10.4 (Becton, Dickinson and Company) and in Microsoft Excel.

3. Results

3.1 Generation and evaluation of single fluorescent protein strains

3.1.1 Generation of single fluorescent protein strains

The fluorescent proteins investigated in this project were chosen based on their attributes documented in FPbase (Lambert, 2019); these were CyOFP1, smURFP a yeast codon optimized version of CyOFP1 and a yeast codon optimized version of mBeRFP. The main attribute of interest was the brightness since higher brightness generally means stronger fluorescent signals. Therefore, quite naturally, the brighter the protein, the better for use as a biosensor. For initial investigation of the fluorescent activity of the proteins, four strains of *S. cerevisiae* were generated. These were TMB RP 002 (CyOFP1), TMB RP 003 (smURFP), TMB RP 004 (opt. CyOFP1) and TMB RP 005 (opt. mBeRFP) each with one gene encoding a fluorescent protein behind the strong constitutive promoter *TEF1,* integrated into the XI-3 genomic locus. To confirm that the correct genes had been integrated into the correct site, five transformants of each strain were picked and yeast colony PCR was performed in two steps, each time followed by agarose gel electrophoresis.

First, primers annealing to the beginning and end of the FP genes were used to confirm that the genes were present in the obtained transformants. The original plasmids from which the genes had been cloned were used as positive controls in these cases. As seen in Figure 4, all screened transformants from all four strains had integrated the respective genes, despite issues with the lane for the positive control for TMB RP 003 (smURFP) (Figure 4B), the positive control lanes for TMB RP 004 (opt. CyOFP1) and the closest ladders for TMB RP 005 (opt. mBeRFP) (Figure 4C).

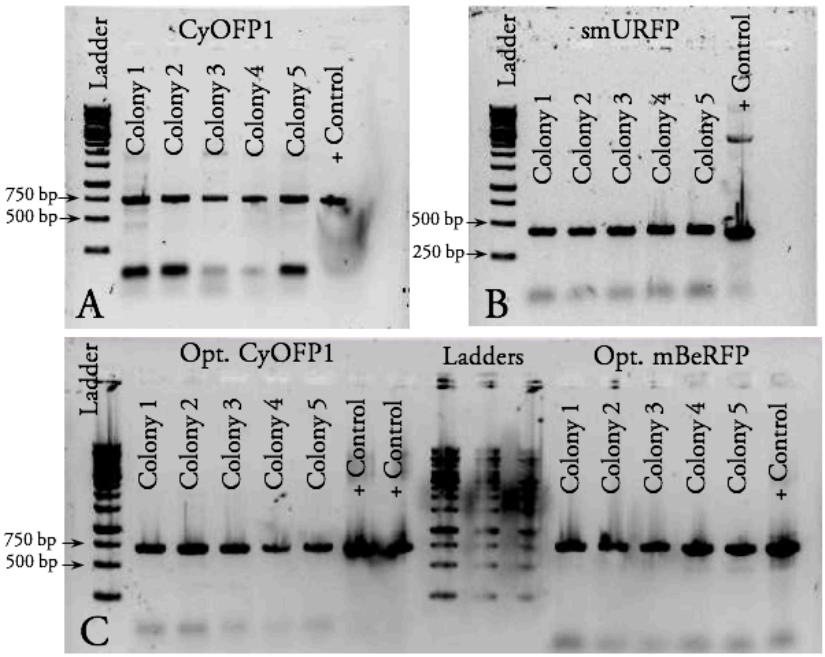

Figure 4. Agarose gels loaded with GeneRuler™ 1kb DNA Ladder, yeast colony PCR products from TMB RP 002 (CyOFP1) (A), TMB RP 003 (smURFP) (B), TMB RP 004 (opt. CyOFP1) (C left) and TMB RP 005 (opt. mBeRFP) (C right) and positive controls. Primers 31/32 were used for TMB RP 002, primers 27/28 for TMB RP 003, primers 33/34 for TMB RP 004 and primers 35/36 for TMB RP 005. The expected sizes of the PCR products were 743 bp for CyOFP1, 428 bp for smURFP, 705 bp for opt. CyOFP1 and 708 bp for opt. mBeRFP.

Transformants 3 and 4 of TMB RP 002 (CyOFP1), along with transformants 1 and 2 of TMB RP 003 (smURFP), TMB RP 004 (opt. CyOFP1) and TMB RP 005 (opt. mBeRFP) were selected for further characterization.

To confirm that the FP genes had been integrated into the correct genomic locus, primers annealing in the XI-3 region, up- and downstream of the integrated genes, were used to verify that the genes indeed had been integrated in the XI-3 site. The background strain, with no genetic modifications in the XI-3 site was used as a negative control in these cases. Based on the gels shown in Figure 5, including a re-run of TMB 002 (CyOFP1) transformant 3 (Figure 5B) due to a probable loading issue in the first run (Figure 5A), it was concluded that all transformants had the respective genes integrated into the correct site.

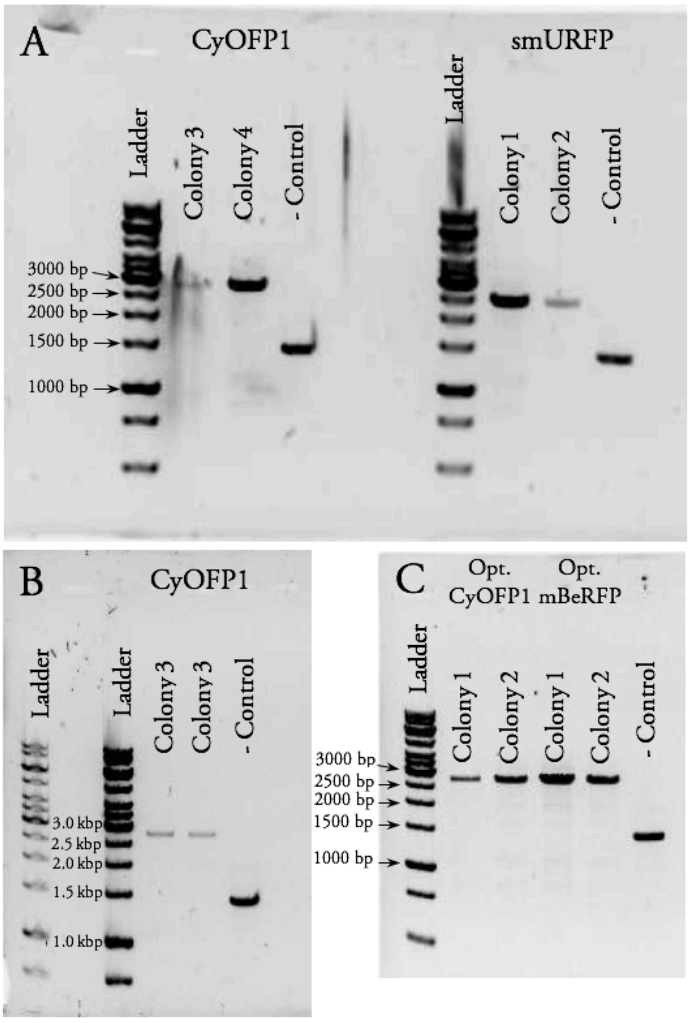

Figure 5. Agarose gels loaded with GeneRuler™ 1kb DNA Ladder, yeast colony PCR products from TMB RP 002 (CyOFP1) (A, left and B), TMB RP 003 (smURFP) (A, right), TMB RP 004 (opt. CyOFP1) (C, left) and TMB RP 005 (opt. mBeRFP) (C, right) and negative controls (PCR products from the background strain in all cases). Primers 23/24 were used for all strains. The expected sizes of the PCR products were 2746 bp for CyOFP1 (both versions), 2440 bp for smURFP, 2749 bp for opt. mBeRFP and 1440 bp for the negative control.

Transformants 3 and 4 of TMB RP 002 (CyOFP1), along with transformants 1 and 2 of TMB RP 003 (smURFP), TMB RP 004 (opt. CyOFP1) and TMB RP 005 (opt. mBeRFP) were saved in glycerol stock as described in section 2.1.

3.1.2 Growth characterization of single fluorescent protein strains in shake flasks

To see if the genetic modifications made in the single fluorescent protein strains had any effect on their capability to grow, growth characterization was performed in rich medium (YPD containing 20 g/L glucose) by measuring OD_{620} over 24 hours of cultivation time, with frequent sampling during the log phase and a 24-hour sample to see how the strains grew over a longer period of time.

The previously constructed strains TMB RP 001 (yEGFP) and CEN.PK+pRPF2 (mEGFP) were also included. The results from one clone of each strain showed that integration of the proteins into the XI-3 genomic locus had very little or no effect on the growth characteristics of the strains (Figure 6). It should be noted that TMB RP 001 (yEGFP) #1, TMB RP 002 (CyOFP1) #3 and TMB RP 003 (smURFP) #1 only were characterized once, while the other clones were characterized at least three times. This was a consequence of the TMB RP 001 (yEGFP) #1, TMB RP 002 (CyOFP1) #3 and TMB RP 003 (smURFP) #1 strains not showing desirable flow cytometry results during the first run. Therefore, these strains were excluded from further testing, whereas the other strains were more promising and hence, were subject to additional growth characterizations and flow cytometry measurements (expanded on in following sections).

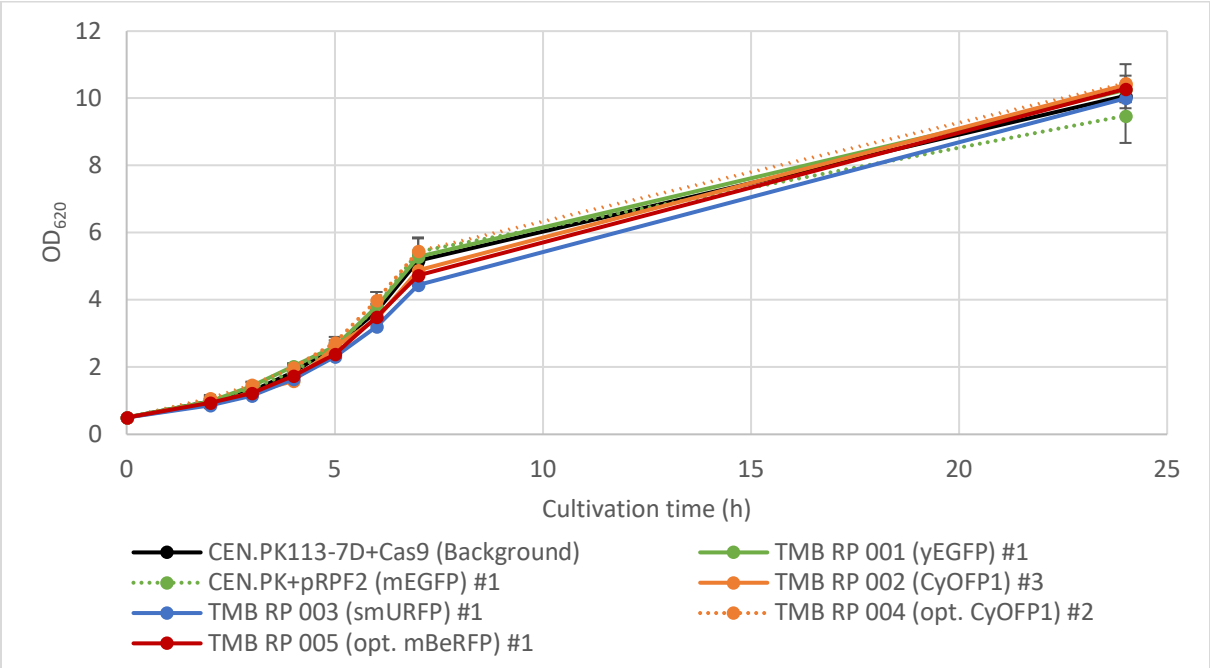

Figure 6. Growth curves for the strains with genes encoding a fluorescent protein integrated in the XI-3 genomic locus. Samples were taken at 0, 2, 3, 4, 5, 6, 7 and 24 hours of cultivation time. Except for TMB RP 001 (yEGFP) #1, TMB RP 002 (CyOFP1) #3 and TMB RP 003 (smURFP) #1, the curves represent averages of several runs, with error bars showing the standard deviations.

3.1.3 Fluorescent activities of fluorescent proteins expressed in *S. cerevisiae*

The fluorescent activities of the examined proteins were measured in parallel with growth characterization. The same samples from 0-24 hours of cultivation time were run in a flow cytometer to assess the mean fluorescence over time. Although most of the proteins had filters for signal detection which were well aligned with their emission peaks, data was collected in all filters for all proteins to scan for fluorescence spillover. It was observed that fluorescence activity peaks occurred in general after three hours of cultivation time (as illustrated for green fluorescent protein in Figure 7). With this in mind, bar graphs were also generated using the peak fluorescence data from the three-hour samples.

Green fluorescent proteins (GFPs)

The fluorescent activities of the two GFPs (multimeric yEGFP and monomeric mEGFP) used in this project should theoretically be detected using the 510/15 nm emission filter. Figure 7 shows the mean fluorescence over time for TMB RP 001 (yEGFP), CEN.PK-pRPF2 (mEGFP) and the background strain, detected using the intended filter.

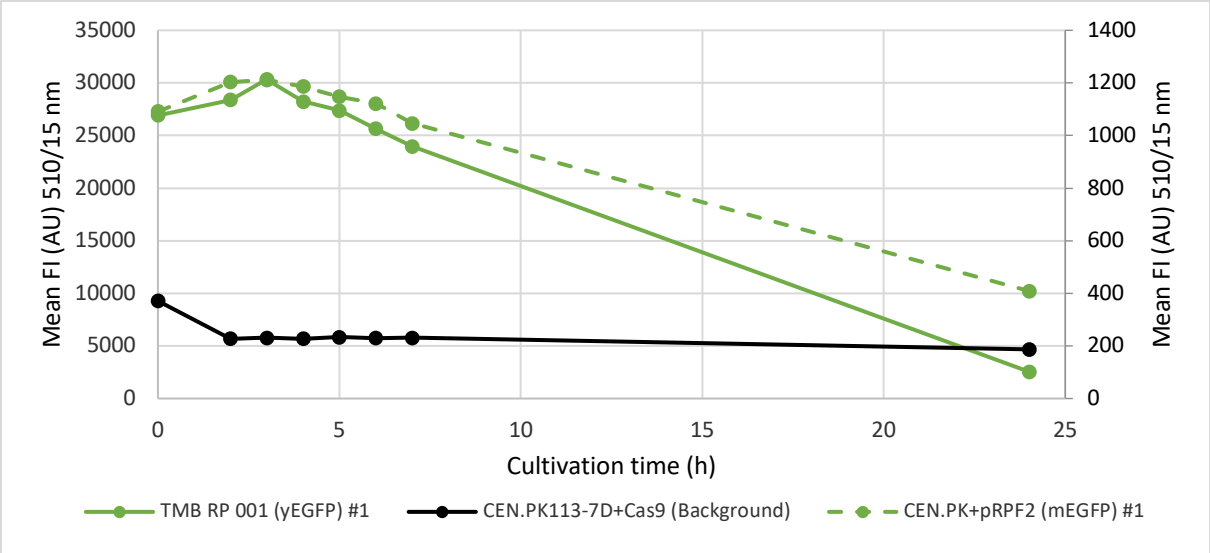

Figure 7. Mean fluorescence intensities (FI) over time for the background strain, TMB RP 001 (yEGFP) #1 and CEN.PK+pRPF2 (mEGFP) #1. Data for TMB RP 001 (yEGFP) #1 is associated with the primary Y-axis and for the background strain and CEN.PK+pRPF2 (mEGFP) #1 is associated with the secondary Y-axis. Data was collected using the 510/15 nm emission filter in the FL1 position.

It was clear that the signal from the TMB RP 001 (yEGFP) strain was much stronger (131-fold) than the signal from the background strain. The signal from CEN.PK-pRPF2 (mEGFP) was also stronger (over 5-fold) than the signal from the background strain. When generating the bar graph at three hours of cultivation time (Figure 8), it was also seen that strong signals from TMB RP 001 (yEGFP) were also detected in the positions where other filters were placed, especially with the 510/15 nm and 610/20 nm filters. This extent of spillover fluorescence was unexpected. Fluorescence from CEN.PK-pRPF2 (mEGPF) however did not spill over as much and showed levels quite similar to the background strain in the other filters.

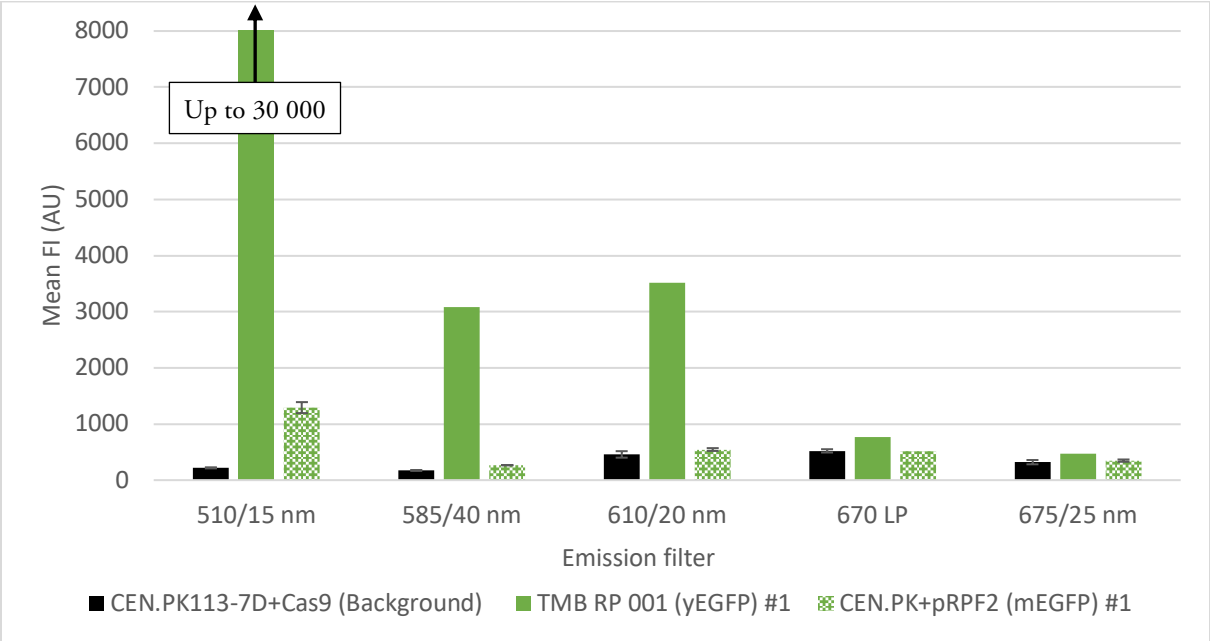

Figure 8. Mean fluorescence intensities (FI) at three hours of cultivation time for the background strain, TMB RP 001 (yEGFP) #1 and CEN.PK+pRPF2 (mEGFP) #1. Data was collected with the filters in the positions which were deemed optimal. The background and CEN.PK+pRPF2 (mEGFP) #1 data points are averages of measurements from several runs, with error bars showing the standard deviations. ***Notes:*** *The data point for TMB RP 001 (yEGFP) #1 in the 510/15 nm filter reaches over 30 000 AU which is not shown. The data point for TMB RP 001 (yEGFP) #1 in the 585/40 nm filter is from TMB RP 001 (yEGFP) #3 – a clone with theoretically the identical genotype.*

Orange fluorescent proteins (OFPs)

The fluorescent activities of the two OFPs used in this project should theoretically be detected using the 585/40 nm emission filter, or possibly using the 610/20 nm emission filter since it nearly aligned with the emission peak of the OFPs. Figure 9 shows the fluorescence over time for TMB RP 002 (CyOFP1), TMB RP 004 (opt. CyOFP1) and the background strain, detected using these two filters. Both strains with OFPs had stronger signals than the background strain. The yeast codon optimized CyOFP1 strain (TMB RP 004) outperformed the original CyOFP1 strain (TMB RP 002) by a large margin with signals more than threefold higher than those from the original CyOFP1 strain (TMB RP 002). It was also clear that the signals detected using the 610/20 nm emission filter were much stronger than the signals detected using the 585/40 nm emission filter.

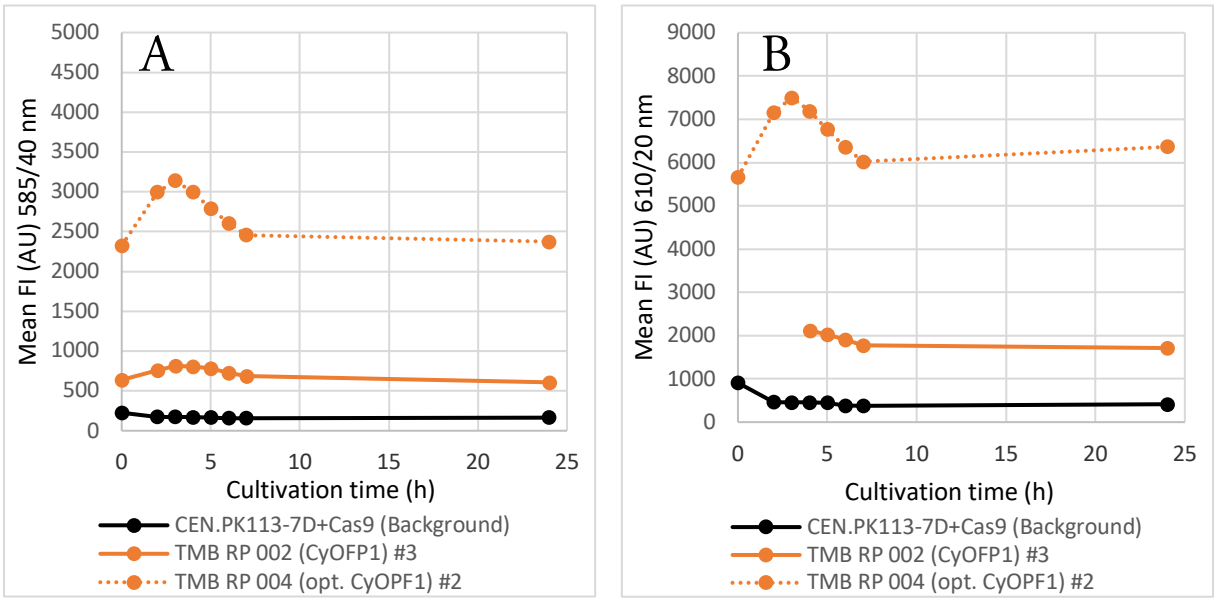

Figure 9. Mean fluorescence intensities (FI) over time for the background strain, TMB RP 002 (CyOFP1) #3 and TMB RP 004 (opt. CyOFP1) #2. Data in A was collected using the 585/40 nm emission filter in the FL2 position. Data in B was collected using the 610/20 nm emission filter in the FL3 position. ***Note:*** *In B, no data was collected for TMB RP 002 (CyOFP1) #3 during the first three hours of cultivation time using this filter setup.*

When scanning over all filters (Figure 10), it was confirmed that TMB RP 004 (opt. CyOFP1) had stronger mean fluorescence intensity than TMB RP 002 (CyOFP1), with 4.1- and 3.6-fold stronger signals in the 585/40 nm and 610/20 nm filters, but also 3.3-fold stronger signal in the 670 LP filter. The signals in the 675/25 nm filter were slightly stronger for TMB RP 004 (opt. CyOFP1) and the signals in the 510/15 nm filter for both OFP strains were as weak as the signals from the background strain.

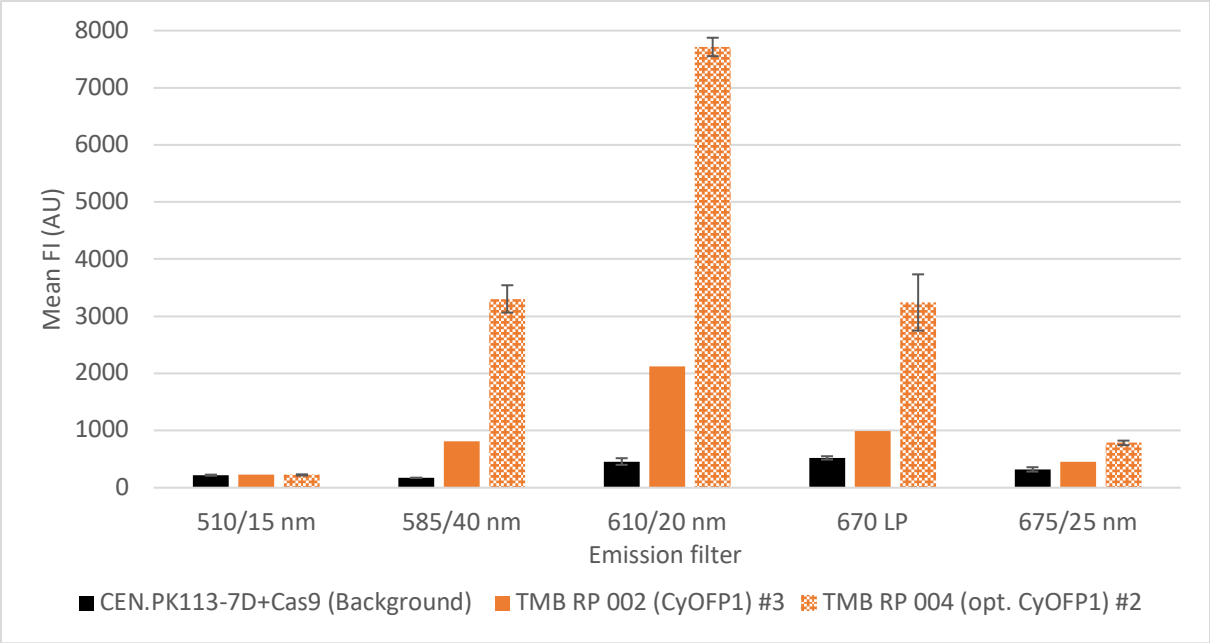

Figure 10. Mean fluorescence intensities (FI) at three hours of cultivation time for the background strain, TMB RP 002 (CyOFP1) #3 and TMB RP 004 (opt. CyOFP1) #2. Data was collected with the filters in the positions which were deemed optimal. The background and TMB RP 004 (opt. CyOFP1) #2 data points are averages of measurements from several runs, with error bars showing the standard deviations. ***Notes:*** *The data point for TMB RP 002 (CyOFP1) #3 associated with the 610/20 nm filter is from the 4-hour sample. The value is therefore slightly lower than it should have been, assuming that the fluorescence peak was at three hours of cultivation time.*

Red fluorescent proteins (RFPs)

In theory, the fluorescent activity of TMB RP 005 (opt. mBeRFP) should be detected using the 610/20 nm or 670 LP emission filters in the FL3 position. Also, the fluorescent activity of TMB RP 003 (smURFP) should be detected using the 675/25 nm emission filter in the FL4 position because the protein is excited by the 640 nm laser and the flow cytometer only detects the fluorescence of proteins excited by this laser in the FL4 position.

The fluorescence intensities over time for TMB RP 005 (opt. mBeRFP), TMB RP 003 (smURFP) and the background strain, detected using the 610/20 nm, 670 LP and 675/25 nm filters are shown in Figure 11. It should however be noted that no comparable data was collected from TMB RP 003 (smURFP) using the 610/20 filter in the FL3 position.

TMB RP 005 (opt. mBeRFP) had strong signals in both the 610/20 nm and 670 LP filters, clearly distinguishable from the signal coming from the background strain. TMB RP 005 (opt. mBeRFP) also displayed a signal in the 675/25 nm filter, which was clearly distinguishable from the other two strains but not as strong as its signals using the other filters.

As suspected, no strong signal from TMB RP 003 (smURFP) was detected in the 670 LP filter where it had the same signal as the background strain, because the filter was placed in the FL3 position. It was however surprising that the smURFP strain displayed essentially no more activity than the background strain in the 675/25 nm filter, when this filter in the FL4 position was the theoretically ideal setup.

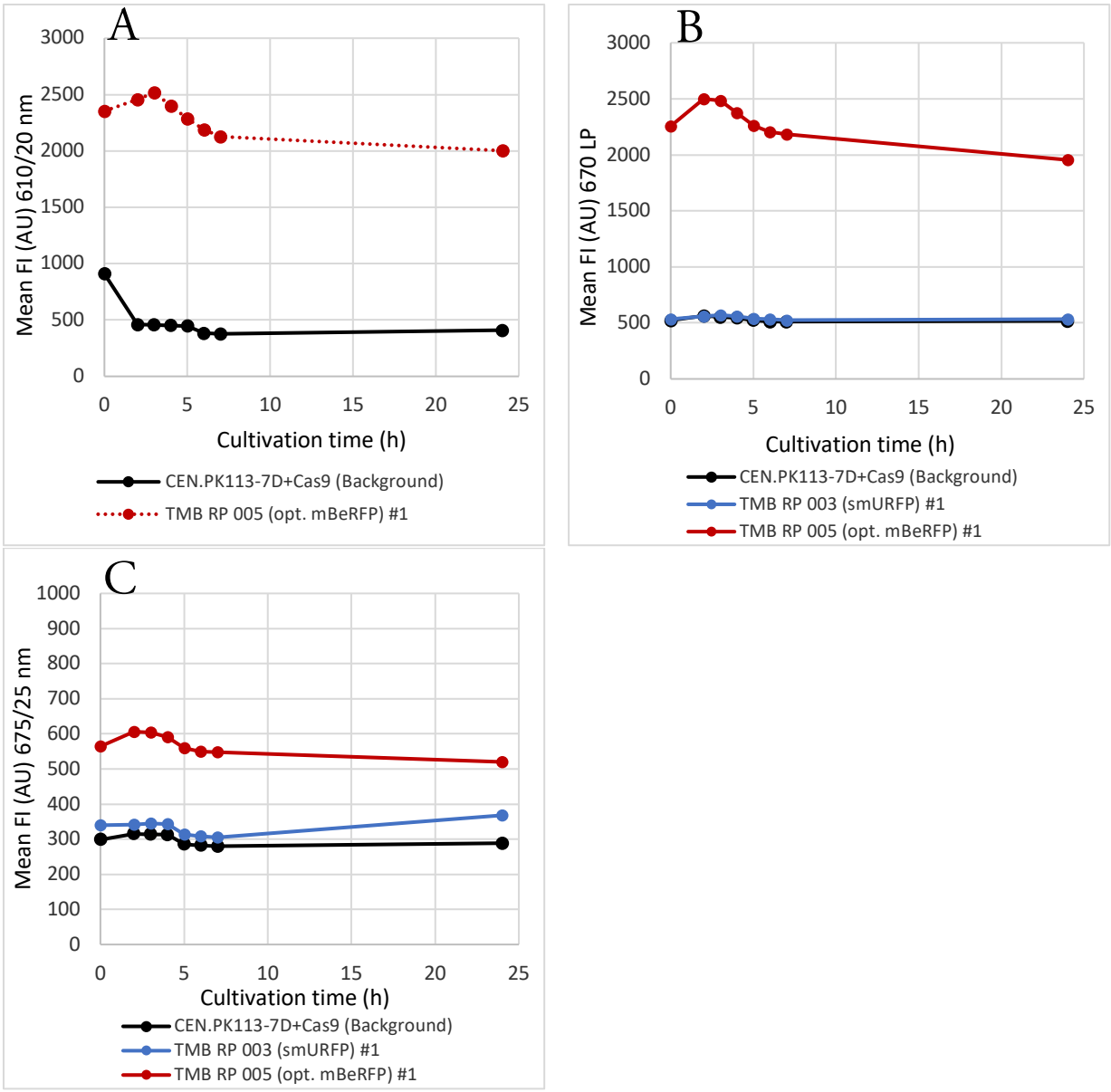

Figure 11. Mean fluorescence intensities (FI) over time for the background strain, TMB RP 003 (smURFP) #1 (not in A) and TMB RP 005 (opt. mBeRFP) #1. In A, data was collected using the 610/20 nm emission filter in the FL3 position, but no comparable data was collected for TMB RP 003 (smURFP) using this setup. In B, data was collected using the 670 LP nm emission filter in the FL3 position. In C, Data was collected using the 675/25 nm emission filter in the FL4 position.

In addition to what could be seen in Figure 11, it was observed that TMB RP 005 (opt. mBeRFP) also had a stronger signal than the background strain using the 585/40 nm filter, but not nearly as strong as its signals in the 610/20 nm and 670 LP filters (Figure 12). Its signal in the 510/15 nm filter was essentially the same as the signal from the background strain.

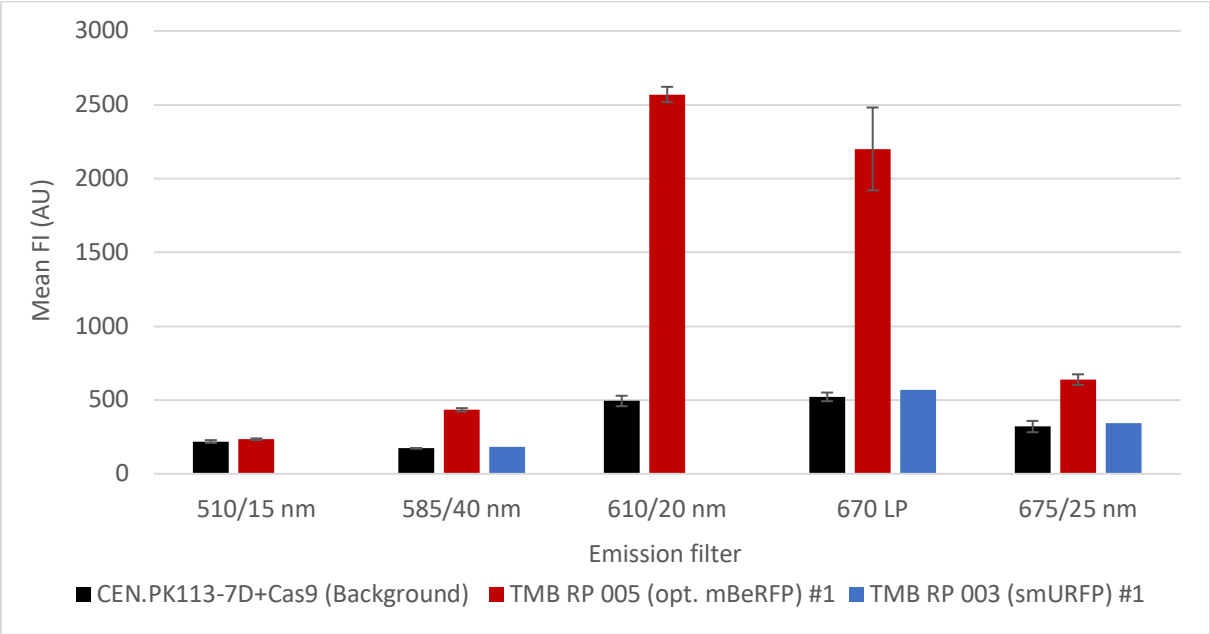

Figure 12. Mean fluorescence intensities (FI) at three hours of cultivation time for the background strain, TMB RP 003 (smURFP) #1 and TMB RP 005 (opt. mBeRFP) #1. Data was collected with the filters in the positions which were deemed optimal. The background and TMB RP 005 (opt. mBeRFP) #1 data points are averages of measurements from several runs, with error bars showing the standard deviations. ***Notes:*** *No comparable data was collected for TMB RP 003 (smURFP) #1 using the 610/20 nm filter and no data was collected at all for the same strain using the 510/20 nm filter since no signal was expected.*

Summary of the single fluorescent protein strains

To summarize, it was discovered that all examined proteins except for smURFP had sufficiently high fluorescence activities to easily distinguish them from the background strain (Figure 13). The signals were detected using the filters which were theoretically optimal; however, there was some unwanted and unexpected spillover in a few cases. The strongest signals from the green fluorescent proteins (GFPs) were detected in the 510/15 nm filter, the strongest signals from the orange fluorescent proteins (OFPs) were detected in the 585/40 nm and 610/20 nm filters and the strongest signals from the red fluorescent protein (RFP) with notable fluorescent activity, opt. mBeRPF, were detected in the 610/20 nm and 670 LP filters. As expected, the signals from the GPFs, OFPs and mBeRFP in the 675/25 nm filter were very low since theoretically none of these proteins were excited by the red laser (640 nm) of the flow cytometer. It was unexpected that smURFP did not show any notable signal using the 675/25 filter and it was concluded that smURFP did not work when integrated into the XI-3 site of *S. cerevisiae.*

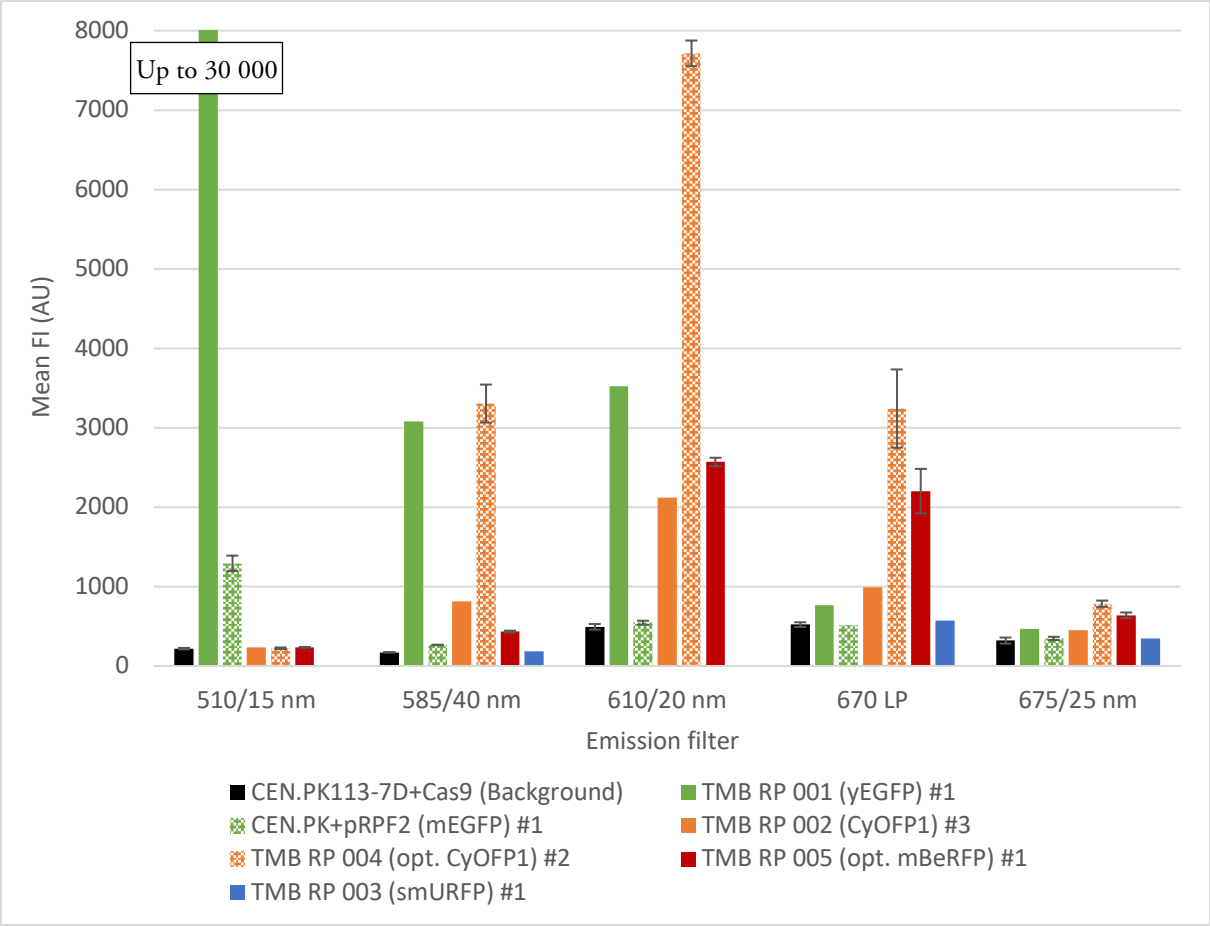

Figure 13. Summary of the mean fluorescence intensities (FI) at three hours of cultivation time for the strains with genomic integrations of the examined proteins. Data was collected with the filters in the positions which were deemed optimal. The data points for the background, CEN.PK+pRPF2 (mEGFP) #1, TMB RP 004 (opt. CyOPF1) #2 and TMB RP 005 (opt. mBeRFP) #1 strains are averages of measurements from several runs, with error bars showing the standard deviations. The remaining strains were only measured once, hence the absence of error bars. ***Notes:*** *The data point for TMB RP 001 (yEGFP) #1 in the 510/15 nm filter reaches over 30000 AU which is not shown. The data point for TMB RP 001 (yEGFP) #1 in the 585/40 nm filter is from TMB RP 001 (yEGFP) #3 – a clone with theoretically the identical genotype. The data point for TMB RP 002 (CyOFP1) #3 associated with the 610/20 nm filter is from the 4-hour sample. The value is therefore slightly lower than it should have been, assuming that the fluorescence peak was at three hours of cultivation time. No comparable data was collected for TMB RP 003 (smURFP) #1 using the 610/20 nm filter and no data was collected at all for the same strain using the 510/20 nm filter since no signal was expected.*

3.2 Generation and evaluation of strains carrying two fluorescent proteins

3.2.1 Strain construction

Having two or more fluorescent proteins integrated into the same strain was sought because it would be a step closer to the end-goal of having multiple biosensors in one strain in order to measure several properties simultaneously. However, the mean fluorescence intensities of the single fluorescent protein strains showed that not all fluorescent proteins were suitable for combinations due to the fluorescence spillover. The use of yEGFP for combinations was not considered since it had heavy, unwanted spillover into several filters. Neither was CyOFP1 considered for combinations due to showing weaker signals overall compared to the yeast codon optimized version. Since opt. CyOFP1 had spillover fluorescence which gave stronger signals in the 610/20 nm and 670 LP filters than opt. mBeRFP in these filters it was deemed unsuitable to combine these two proteins, since ideally one protein with a dominating signal in a given filter was sought after. The options were narrowed down and combining the remaining green fluorescent protein (mEGFP) with the remaining orange fluorescent protein (opt. CyOFP1) and the same green fluorescent

protein with the only possible red fluorescent protein (opt. mBeRFP) in separate strains were thought to be reasonable options. The idea behind this was that fluorescence from mEGFP did not spill over much into the filters where fluorescence from opt. CyOFP1 and opt. mBeRFP were detected and neither did fluorescence from opt. CyOFP1 nor opt. mBeRPF spill over into the filter where fluorescence from mEGFP was mainly detected. Therefore, two strains with two fluorescent proteins in each were generated by integrating mEGFP into the XI-2 genomic locus of the TMB RP 004 (opt. CyOFP1) #2 and TMB RP 005 (opt. mBeRFP) #1 strains, which already had one copy of their respective fluorescent proteins integrated into the XI-3 sites.

As with the single fluorescent protein strains, yeast colony PCR followed by agarose gel electrophoresis with the PCR products was performed to verify that the correct mEGFP gene was integrated in the correct site. A primer pair where one primer annealed inside the gene encoding mEGFP and one primer annealed in the XI-2 region downstream of the integrated mEGFP and flanking *ADH1t* terminator was used. This way, it could both be confirmed that the correct genes were integrated and that they were integrated in the correct site.

For both strains, the PCR products from all colonies were of the expected sizes, 1334 bp, confirming that mEGFP had been integrated in the XI-2 site in both strains (Figure 14). Transformants 1 and 2 of TMB RP 008 (opt. CyOFP1 + mEGFP) and TMB RP 009 (opt. mBeRFP + mEGFP) were saved in glycerol stock as described in section 2.1.

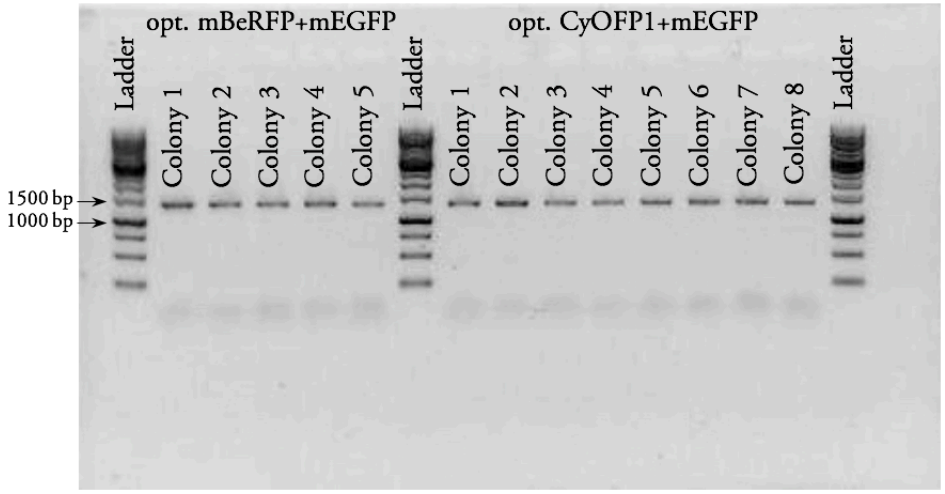

Figure 14. Agarose gel loaded with GeneRuler™ 1kb DNA Ladder and yeast colony PCR products from TMB RP 009 (opt. mBeRFP + mEGFP) and TMB RP 008 (opt. CyOFP1 + mEGFP). Primers 13/324 were used for both strains.

3.2.2 Growth characterization of the strains with double FPs shake flasks

Growth characterization in rich medium (YPD containing 20 g/L glucose) showed again that the introduction of the mEGFP gene did not seem to have an adverse effect on growth (Figure 15).

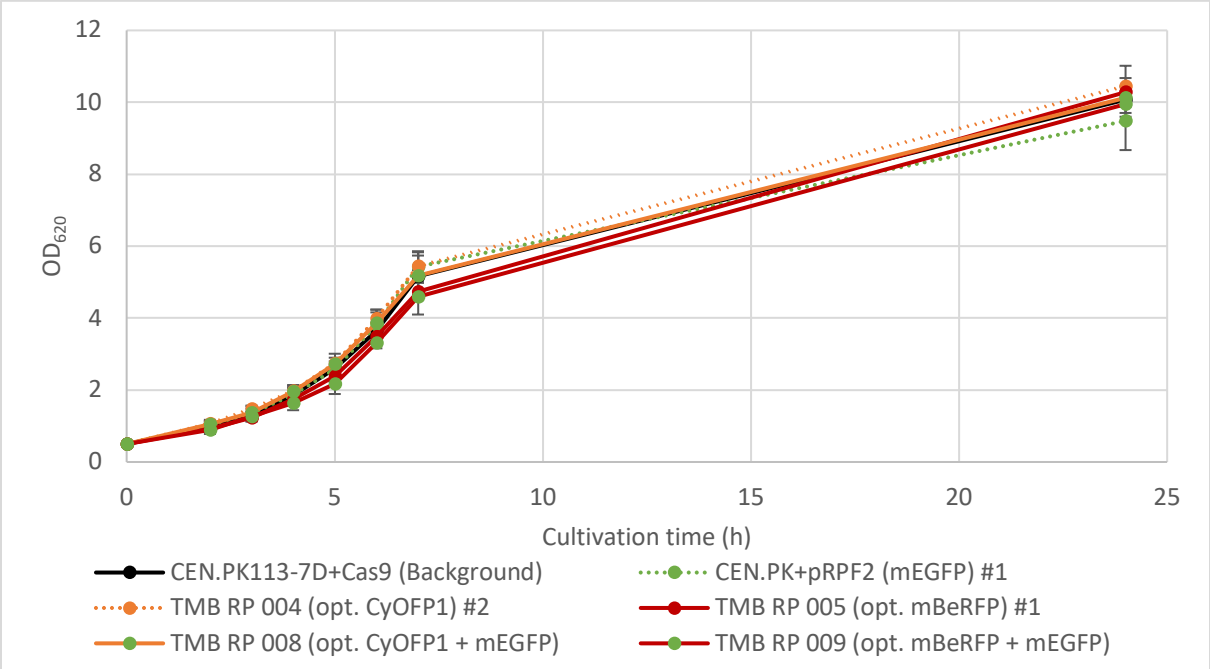

Figure 15. Growth curves for the double fluorescent protein strains with mEGFP integrated into the XI-2 genomic locus, TMB RP 008 (opt. CyOPF1 + mEGFP) and TMB RP 009 (opt. mBeRFP + mEGFP), along with the background strain, parental strains TMB RP 004 (opt. CyOFP1) #2 and TMB RP 005 (opt. mBeRFP) #1 and CEN.PK+pRPF2 (mEGFP) #1. Samples were taken at 0, 2, 3, 4, 5, 6, 7 and 24 hours of cultivation time. With the exceptions of the double FP strains, the curves represent averages of several runs, with error bars showing the standard deviations. The curves for the double FP strains represent the averages of two clones, with error bars showing the standard deviations.

3.2.3 Fluorescent activities of co-expressed fluorescent proteins in *S. cerevisiae*

The mean fluorescence intensities at three hours of cultivation time showed that the signals from the double fluorescent protein strains in the 510/15 nm filter were comparable to the signal from the mEGFP strain (CEN.PK+pRPF2) (Figure 16). Similarly, in the 585/40 nm and 610/20 nm filters, the signals from the optimized CyOFP1 + mEGFP strain (TMB RP 008) were comparable to the signals from the optimized CyOFP1 strain (TMB RP 004). In the same filters, the signals from the optimized mBeRFP + mEGFP (TMB RP 009) strain were roughly the same as the signals from the optimized mBeRFP strain (TMB RP 005) but seemed to be slightly higher. This showed that the signals originating from each protein remained of essentially the same strengths when the proteins were integrated into the same strain and it was this result that was hoped for, but also expected based on the theory and previous results. This particular experiment was considered very successful and showed promise for the continuation of the project with the end-goal of having three or more biosensors with different fluorescent protein reporters in one yeast strain.

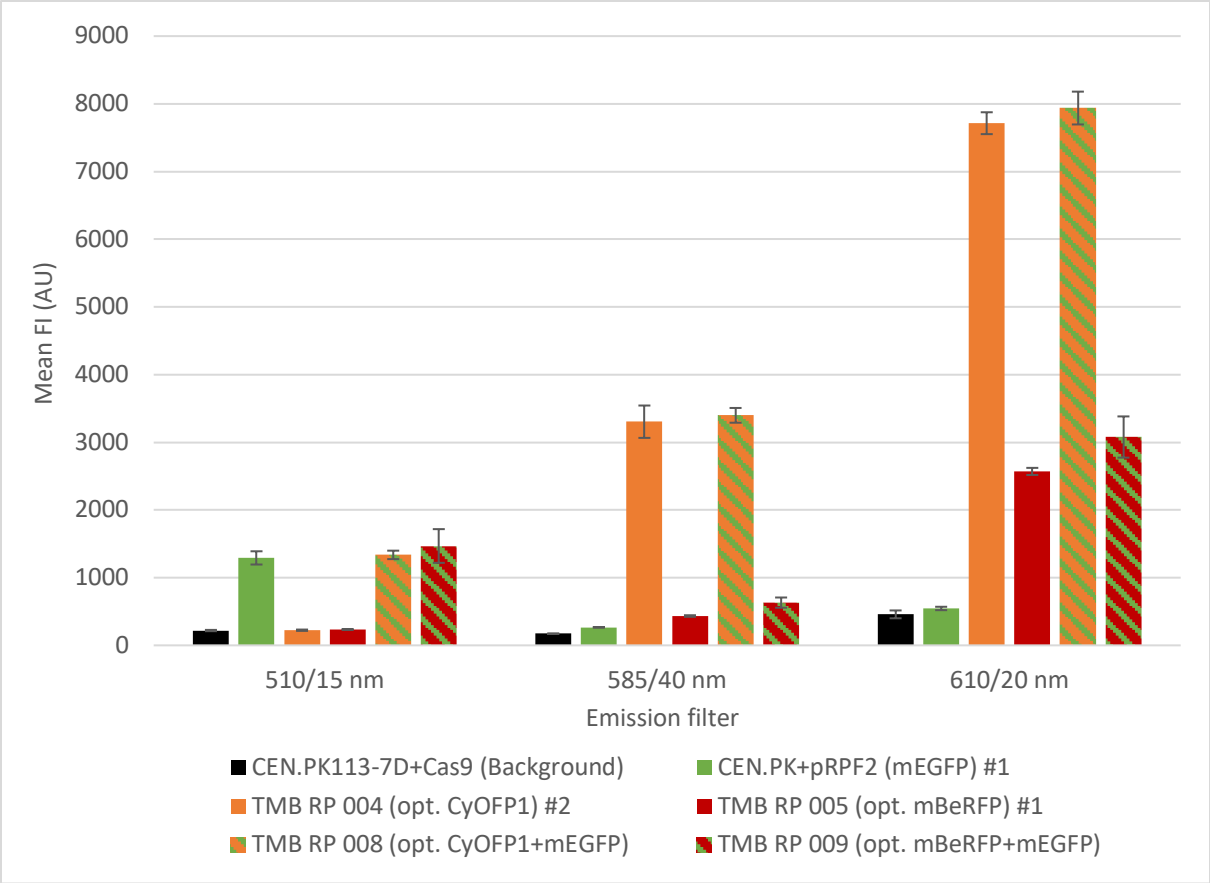

Figure 16. Mean fluorescence intensities (FI) at three hours of cultivation time for the double fluorescent protein strains with mEGFP integrated into the XI-2 genomic locus, TMB RP 008 (opt. CyOPF1 + mEGFP) and TMB RP 009 (opt. mBeRFP + mEGFP), along with the background strain, parental strains TMB RP 004 (opt. CyOFP1) #2 and TMB RP 005 (opt. mBeRFP) #1 and CEN.PK+pRPF2 (mEGFP) #1. Data was collected with the filters in the positions which were deemed optimal. The data points for the background, CEN.PK+pRPF2 (mEGFP) #1, TMB RP 004 (opt. CyOPF1) #2 and TMB RP 005 (opt. mBeRFP) #1 strains are averages of measurements from several runs, with error bars showing the standard deviations. The data points for the double FP strains represent the averages of two clones, with error bars showing the standard deviations.

3.3 Generation and evaluation of modified smURFP strain

3.3.1 Generation of modified smURFP strain

The mean fluorescence intensity data for the RFPs suggested that TMB RP 003 (smURFP) either had no fluorescent activity at all, or at best, just barely noticeable activity. A possible reason for this may have been a lack of biliverdin, the cofactor required by smURFP for fluorescence (Rodriguez et al., 2016). *S. cerevisiae* has an endogenous gene, *HMX1,* encoding a heme oxygenase which catalyzes a reaction producing biliverdin (Saccharomyces Genome Database, 2020). Doubting if this production was happening at all, or in sufficient levels required by smURPF, it was attempted to supply the TMB RP 003 (smURFP) strain with (more of) the cofactor to see if it would make a difference in the mean fluorescence intensity. The HO-1 gene encoding heme oxygenase-1 from *Synechocystis,* optimized for bacterial expression, was available on the pBAD-smURFP-RBS-HO-1 plasmid and due to time limitations, this gene was used in the attempt to supply TMB RP 003 (smURFP) with biliverdin instead of the endogenous *HMX1* from *S. cerevisiae.*

Cloning experiments resulted in the pRP012 multicopy plasmid, harboring *HO-1* and the *URA3* selection marker. The simple approach of transforming TMB RP 003 (smURFP) with the pRP012 multicopy plasmid and utilizing the *URA3* marker for selection could not be used in this project because the strain was not auxotrophic. Instead, the last strain was generated by transforming TMB

RP 003 (smURFP) with the open pRP012 multicopy plasmid, linearized by digestion with *Stu*I in the *URA3* marker sequence, along with the DNA fragment with the *AUR1p-AUR1-C-AUR1t* cassette flanked by *URA3* up- and downstream homology regions. *In vivo* plasmid recombination resulted in an episomal multicopy plasmid (not recovered and therefore not named) with *HO-1* and the plasmid-borne *URA3* gene exchanged for the *AUR1p-AUR1-C-AUR1t* cassette. In this way, the *AUR1-C* marker conferring resistance to the antibiotic Aureobasidin A could be utilized for selection and the generated strain was named TMB RP 010 (smURFP+biliverdin). To confirm successful transformation and plasmid recombination, colony PCR was performed followed by agarose gel electrophoresis. Primers annealing to the beginning and end of the heme oxygenase-1 gene were used to confirm that the gene (and therefore successfully repaired plasmid) was present in the screened colonies (Figure 17). The original plasmid from which the gene had been cloned was used as positive control. Transformants 2 and 4 of TMB RP 003 (smURFP+biliverdin) were saved in glycerol stock as described in section 2.1

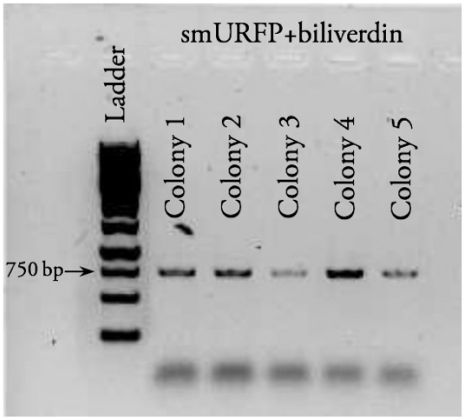

Figure 17. Agarose gel loaded with GeneRuler™ 1kb DNA Ladder, yeast colony PCR products from TMB RP 003 (smURFP)) transformed with pRP012 (StuI) and a DNA fragment and positive control (pBAD-smURFP-RBS-HO-1 plasmid). The expected sizes of the PCR products were 747 bp.

3.3.2 Growth characterization of modified smURFP strain

Clones were grown in rich medium (YPD containing 20 g/L glucose), supplemented with Aureobasidin A to a final concentration of 0.3 mg/L (Figure 18). Addition of the multicopy plasmid to the original smURFP strain (TMB RP 003) gave an adverse effect on growth, although the clones eventually reached OD_{620} similar to both the background and TMB RP 003 (smURFP) #1 strains (Figure 18). However, it was not possible to know if the effect was connected to the presence of the antibiotic or to the expression of the heme oxygenase encoding gene.

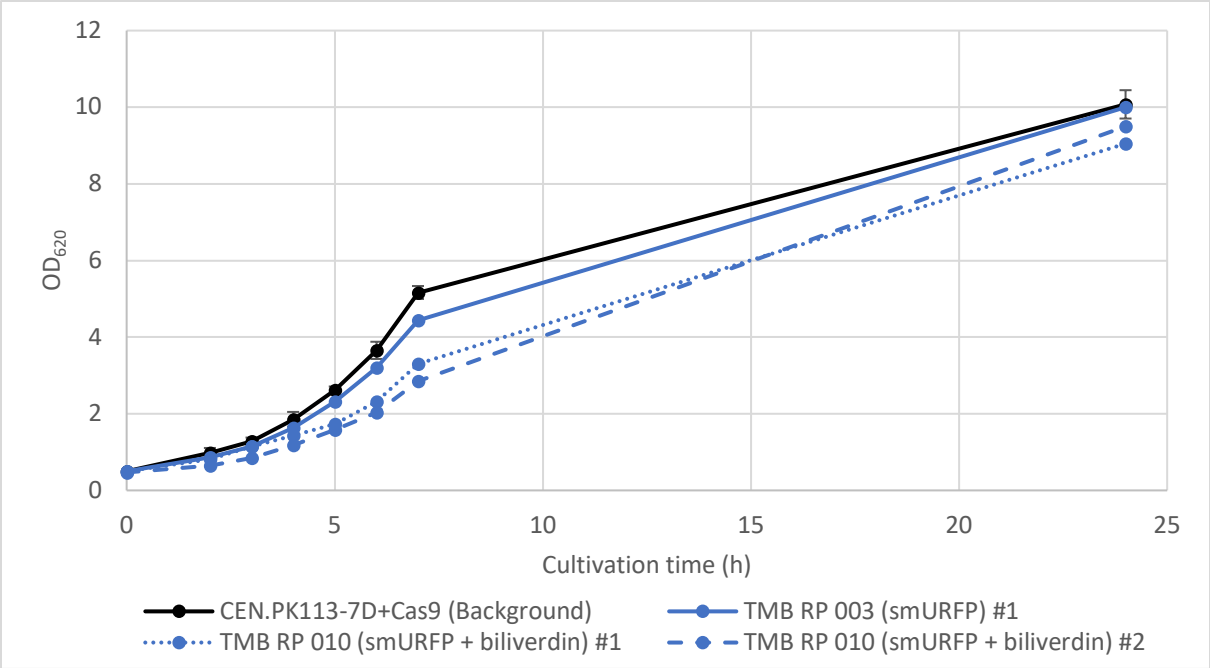

Figure 18. Growth curves for the background strain, TMB RP 003 (smURFP) #1, TMB RP 010 (smURFP+biliverdin) #1 and #2. Samples were taken at 0, 2, 3, 4, 5, 6, 7 and 24 hours of cultivation time. The curve for the background strain represents and averages of several runs, with error bars showing the standard deviations. The other curves represent single runs. ***Note:*** *TMB RP 010 (smURFP+biliverdin) #2 was started at OD = 0.47, instead of the standard OD = 0.5 due to poor growth in the overnight cultures.*

3.3.3 Fluorescent activity of modified smURFP strain

The fluorescent activity of the generated strain TMB RP 010 (smURFP+biliverdin) was detected using the 675/25 nm filter in the FL4 position following the reasoning for the TMB RP 003 (smURFP) strain. Figure 19 shows the mean fluorescence over time for the two TMB RP 010 (smURFP+biliverdin) clones and the background strain, detected using the 675/25 nm filter in the FL4 position. Again, the signal was at best barely higher than the signal from the background strain. The same trend was also seen regardless of which filter was used (data not shown).

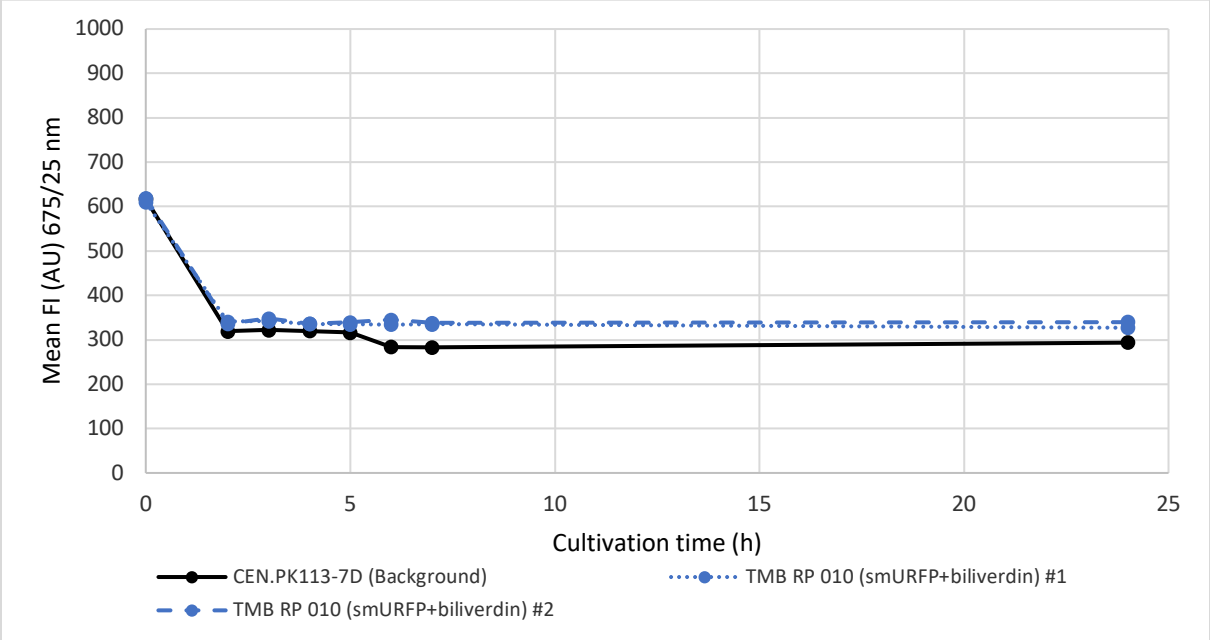

Figure 19. Mean fluorescence intensities (FI) over time for the background strain, TMB RP 010 (smURFP+biliverdin) #1 and #2. Data was collected using the 675/25 nm emission filter in the FL4 position.

3.4 Difference in signal strength between different filter positions

During the fourth run, all samples were measured twice in the flow cytometer, but in between each measurement, the positions of the 610/20 nm and 670 LP filters were swapped. This revealed that placing the same emission filter in the different positions in the flow cytometer yielded differences in signal strength using the same samples. This was the case for all examined proteins. To exemplify, it was observed that the signal for TMB RP 004 (opt. CyOFP1) peaked at roughly 8000 AU with the 610/20 nm filter in the FL3 position, compared to peaking at barely 1000 AU when the filter was in the FL2 position (Figures 20A and 20B). This corresponded to an 8-fold increase in signal strength when the filter was placed in the FL3 position compared to the FL2 position. The same trend was seen for the 670 LP filter, where the signal again was much higher when the filter was placed in the FL3 position compared to when it was placed in the FL2 position (Figures 20C and 20D).

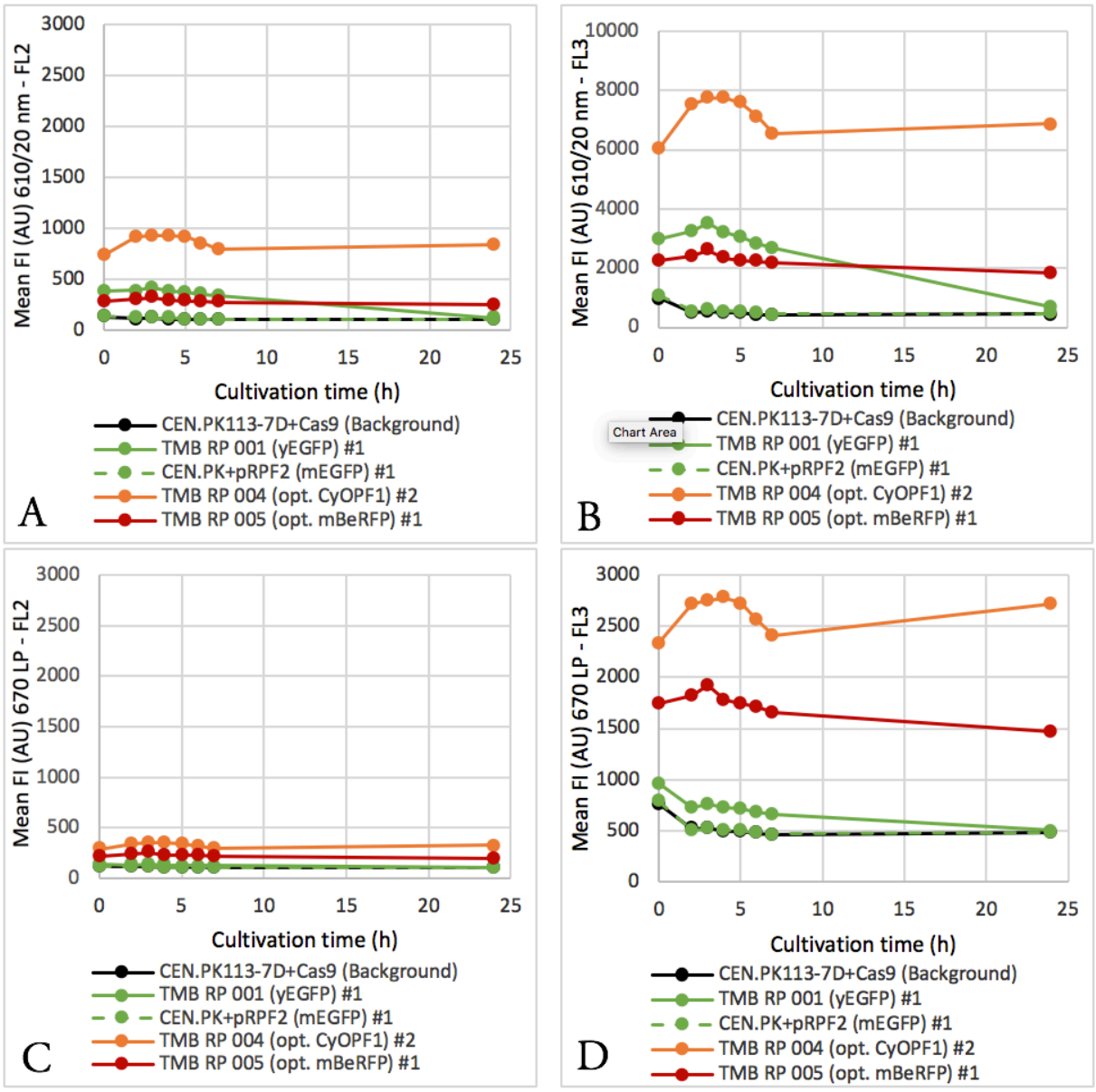

Figure 20. Measurements of mean fluorescence intensities (MFI) are shown over a 24-hour cultivation period. All graphs were generated using data from the fourth run. Graphs A and B show the mean fluorescence intensities collected in the FL2 and FL3 positions respectively using the 610/20 nm filter. Graphs C and D show the mean fluorescence intensities collected in the FL2 and FL3 positions respectively using the 670 LP filter.

This discovery prompted an investigation into the relationships between the different filter positions in regard to signal strengths. The 24-hour samples from the fourth run were measured several times over in the flow cytometer, swapping the 510/15 nm, 610/20 nm and 670 LP filters between the FL1, FL2 and FL3 positions. This experiment made it very clear that the signals were strongest in the FL3 position followed by the FL1 position and last the FL2 position, regardless of which filter was placed in each position (Figure 21). Data from the 670 LP filter is not shown but highlighted the same trend as the data in Figure 21.

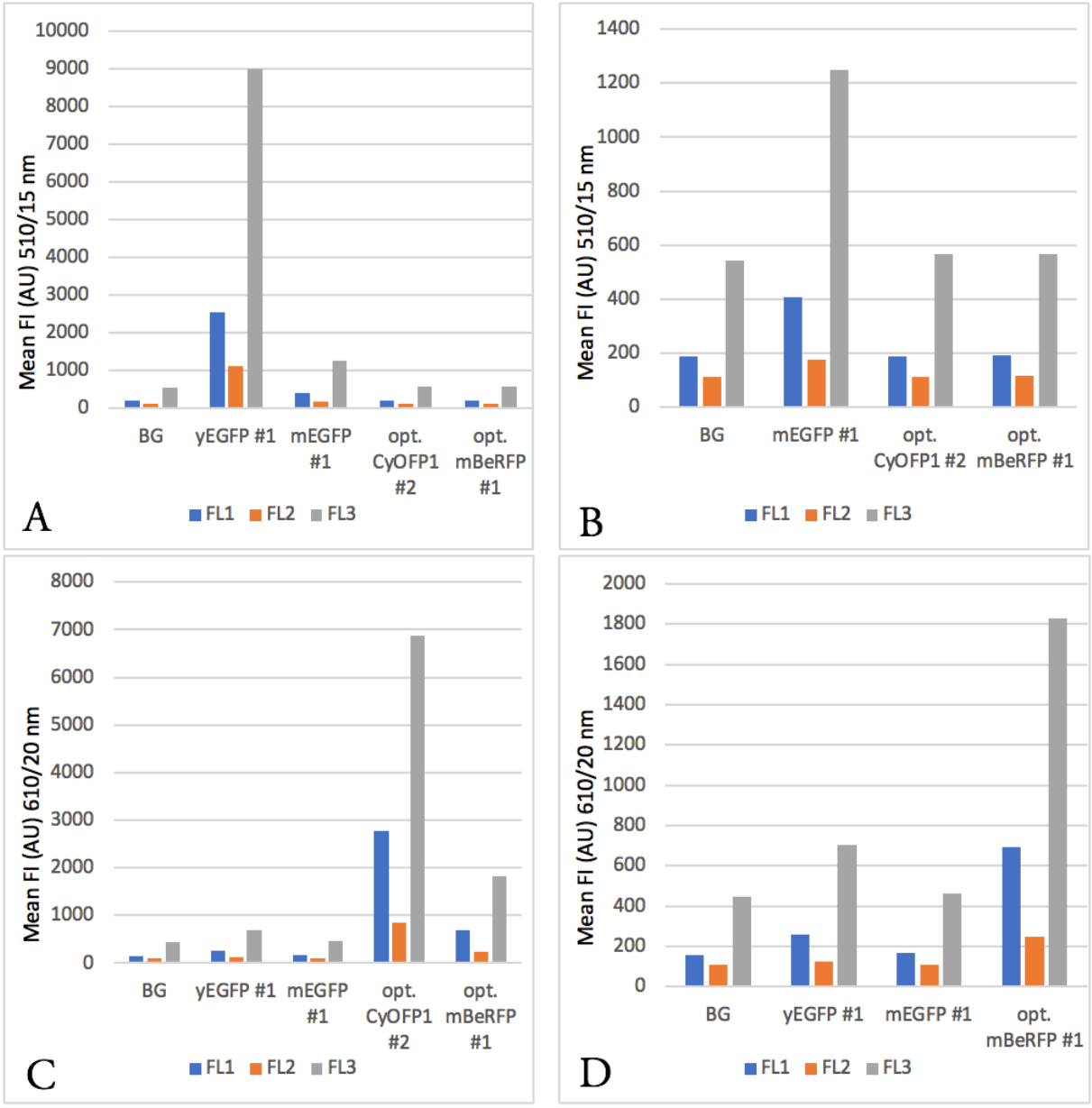

Figure 21. The mean fluorescence intensities (MFI) of the proteins examined in the fourth run measured in the 24-hour samples. Graphs A and B show the signals captured using the 510/15 nm filter in the different filter positions. In B, the signals from yEGFP clone #1 have been removed to more clearly show the trend. Graphs C and D show the signals captured using the 610/20 nm filter in the different filter positions. In D, the signals from opt. CyOFP1 clone #2 have been removed to more clearly show the trend. BG: background strain.

3.5 Mixing samples of different strain populations

Also during the fourth run, samples from three hours of cultivation time from the individual fluorescent protein strains were diluted to the same OD and mixed. This was done as an experiment to see if it was possible to visualize the different populations (strains) using dot plots, where each dot represents the data collected from one cell. The fluorescence intensities of the mixed samples were measured in the flow cytometer.

Figures 22A and 22B show green fluorescence intensity on the X-axis, detected using the 510/15 nm filter in the FL1 position and red fluorescence intensity on the Y-axis, detected using the 610/20 nm filter in the FL2 position. Unfortunately, the FL3 position was not used for the 610/20 nm filter because it was not known at the time that this position would have given stronger signals, as concluded in section 3.4. Figure 22A shows the single strain samples of the background strain, TMB RP 001 (yEGFP) #1, CEN.PK+pRPF2 (mEGFP) #1, TMB RP 004 (opt. CyOFP1) #2 and TMB RP 005 (opt. mBeRFP) #1 plotted in one figure. This figure was essentially a control, showing how a sample with an equal mixture of these five strains should look. The black shapes are so-called gates, which were applied to roughly include the cells of each strain. Gates were drawn in plots with only one strain, before combining all strains with gates in the same plot (Figure 22A). In practice however, no mixture was done with these five strains.

The resulting dot plot of a mixture done with the CEN.PK+pRPF2 (mEGFP) #1 and TMB RP 004 (opt. CyOFP1) #2 strains is shown in Figure 22B. In this figure, the same gates from the "control" plot were applied to be able to compare and quantify the percentage of cells of each strain in the samples. Although the gates were manually drawn and cannot be configured to accurately capture precisely all cells of one particular strain in a given sample, it was very clear that populations from each sample used in the mixtures were easy to distinguish using dot plots and gates. The numbers below the population labels show the frequencies of each population and these numbers were not too far off from 50% of each, which was the theoretical distribution when mixing two strains with equal OD. If the gates were slightly more generously drawn, the numbers would likely have been even closer to 50% for each population.

Figure 22C shows green fluorescence intensity on the X-axis, detected using the 510/15 nm filter in the FL1 position and orange fluorescence intensity on the Y-axis, detected using the 585/40 nm filter in the FL2 position. In this figure, the three single strain samples of CEN.PK+pRPF2 (mEGFP) #1, TMB RP 004 (opt. CyOFP1) #2 and TMB RP 008 (opt. CyOFP1 + mEGFP) were plotted in one figure. It was clear that the strain with both opt. CyOFP1 and mEGFP had the same level of green fluorescence as the strain with only the green fluorescent protein (mEGFP) and the same level of orange fluorescence as the strain with only the orange fluorescent protein (opt. CyOFP1).

Additionally, fluorescence spillover was seen using these plots. The clearest example is the spillover from yEGFP in Figure 22A where the entire population has high green fluorescence, but also high red fluorescence as indicated by a large portion of the population having similar intensities as the orange (opt. CyOFP1) and red (opt. mBeRFP) fluorescent proteins.

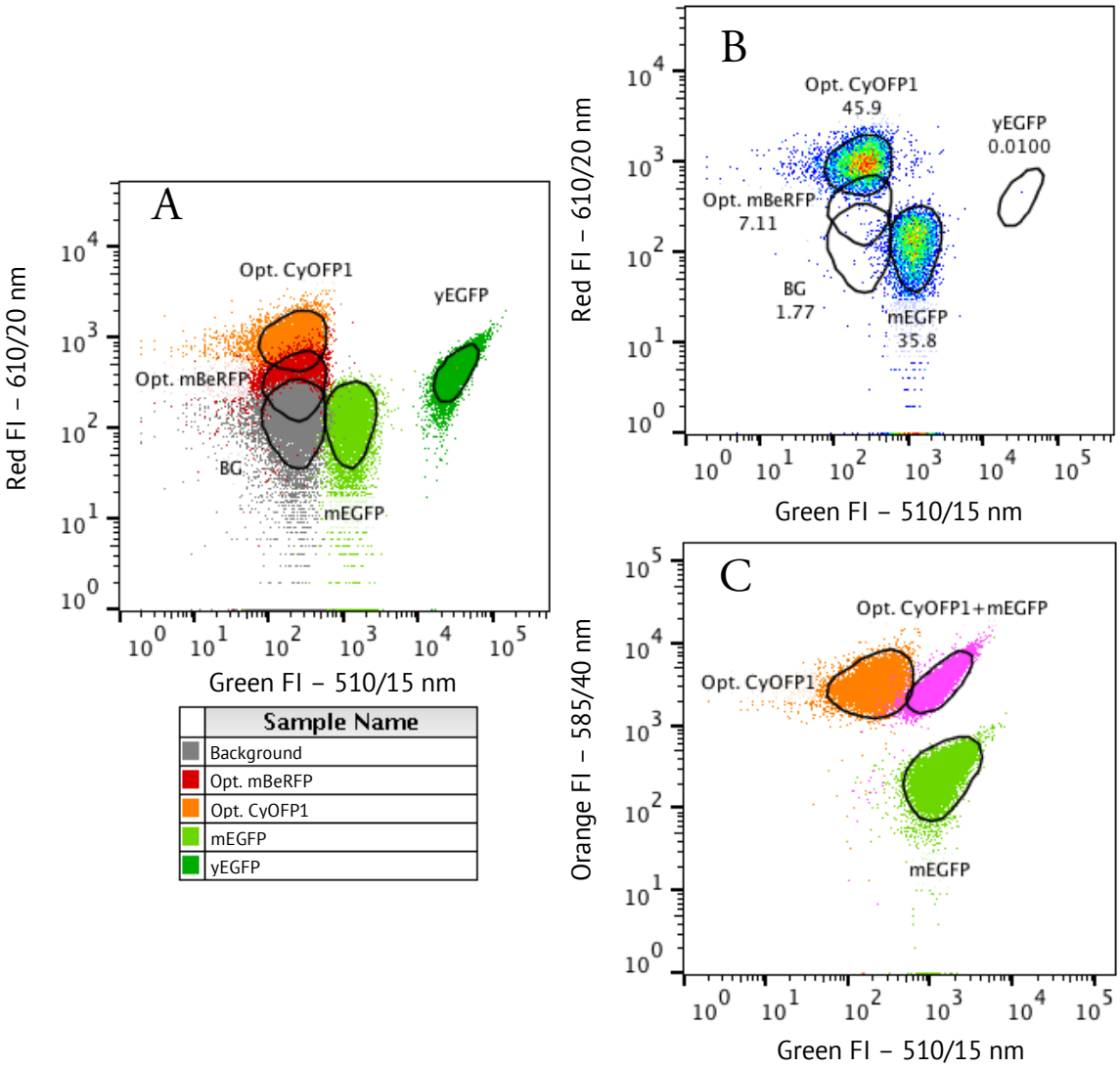

Figure 22. A and B: Dot plots showing green fluorescence intensity on the X-axes and red fluorescence intensity on the Y-axes. In A, single strain samples are plotted in one figure. In B, one mixed sample of CEN.PK+pRPF2 (mEGFP) #1 and TMB RP 004 (opt. CyOFP1) #2 is plotted. Numbers below population labels represent the frequencies of each population in %. C: Dot plot showing single strain samples of CEN.PK+pRPF2 (mEGFP) #1, TMB RP 004 (opt. CyOFP1) #2 and TMB RP 008 (opt. CyOFP1 + mEGFP), with green fluorescence intensity on the X-axis and orange fluorescence intensity on the Y-axis. In all plots, gates are represented by the manually drawn shapes with black lines. BG: background strain.

4. Discussion

Using the CRISPR/Cas9 system developed by Jessop-Fabre et al. (2016), genes encoding the fluorescent proteins CyOFP1, smURFP and mBeRFP were individually integrated into the XI-3 genomic locus of *S. cerevisiae* CEN.PK113-7D. Successful integration of the genes marked the achievement of the first goal of the project, resulting in the four strains TMB RP 002 (CyOFP1), TMB RP 003 (smURFP), TMB RP 004 (opt. CyOFP1) and TMB RP 005 (opt. mBeRFP). Growth characterization of the generated strains showed positive results, with no noticeable adverse effect on the growth of the strains. This was not too surprising as the XI-3 site previously had been determined suitable for gene integrations (Mikkelsen et al., 2012). However, it was crucial to test and establish since it is highly important for the future applications of the proteins that the assay is non-invasive, implying among other things that growth is affected as little as possible, ideally not at all, by integration of the protein encoding genes into the genome. Regardless of the *in vivo* fluorescence of a protein, it could not be used in a biosensor if its integration would adversely affect the growth of the host strain.

The next phase of the project was to evaluate the mean fluorescence intensities of the proteins *in vivo* in *S. cerevisiae* using flow cytometry. Three out of the four generated single fluorescent protein strains had fluorescence signals which were determined to be strong enough to have potential for future use in biosensors. This conclusion was drawn when seeing that their fluorescence intensities were very clearly distinguishable from the background and reproducible with small standard deviations. Especially with the small sample size, this was seen as very promising. Quantification of the fluorescence signals in relation to the background signals revealed that the fluorescence from the three candidates with potential for future use (CyOFP1, opt. CyOFP1 and opt. mBeRFP) were at best 19.1-, 4.7- and 5.2-fold stronger than the signals from the background strain respectively. These numbers were based on data from the 585/40 nm filter for the CyOFP1 proteins and the 610/20 nm filter for mBeRPF. It should also be kept in mind that these signals were detected with the fluorescent protein genes behind the strong constitutive promoter *TEF1* and using other promoters would likely result in stronger or weaker signals depending on the utilized promoter.

Unfortunately, integration of the smURFP gene into the XI-3 site did not result in fluorescence signals stronger than those from the background strain. Since the gene was integrated in the same site and between the same promoter and terminator as in the other strains and those proteins were expressed and worked well, lack of gene expression for smURFP was likely not the reason why the smURFP strain did not show strong signals as hoped. The most probable reason was thought to be the lack of biliverdin required for smURFP fluorescence (Rodriguez et al., 2016). The endogenous heme oxygenase enzyme (encoded by the *HMX1* gene) in *S. cerevisiae* was likely not supplying smURFP with as much biliverdin as hoped. In hindsight this was maybe not so surprising since the *HMX1* gene seems to be regulated by oxidative stress and/or iron starvation, conditions which the strains in this project were not subject to (Saccharomyces Genome Database, 2020).

An attempt to supply the TMB RP 003 (smURFP) strain with biliverdin to increase the fluorescence signals was performed by transforming the strain with a multicopy plasmid carrying the *HO-1* gene from *Synechocystis*, encoding heme oxygenase-1, but optimized for bacterial expression. The *HO-1* gene was used due to time limitations, instead of using *HMX1,* encoding endogenous heme oxygenase to *S. cerevisiae*, which would have been a more logical approach. Unfortunately, the flow cytometry results for the resulting strain, TMB RP 010 (smURFP+biliverdin), were no better than the results for TMB RP 003 (smURFP). Why this method was not successful was not established. Since the amount of biliverdin produced was not

measured in any way, it could not be determined to what levels the cofactor gene was expressed. Neither could it be said with certainty that the cofactor was functional. The fact that the HO-1 gene was not optimized for yeast could have been the problem and this could be tested by ordering a codon optimized version of the gene and performing the same transformation again using the codon optimized gene. However, a previous study on smURFP had not utilized codon optimization of *HO-1* and still detected fluorescence from smURFP in *E. coli* and in mammalian cells (Rodriguez et al., 2016). In the same study, smURFP fluorescence was even detected in mammalian cells without exogenous biliverdin. In the mammalian cells, a codon optimized version of smURFP was used and this may indicate that codon optimization of smURFP may also be required for functionality of the protein in *S. cerevisiae*. Further investigation into getting smURFP to work in *S. cerevisiae* would require some time but may very well be worthwhile because if a solution is found it would result in a fluorescent protein with a signal in the 675/25 nm filter where none of the other fluorescent proteins tested so far have strong signals, thus making it suitable for combinations with the other proteins. This was why the attempt to supply TMB RP 003 (smURFP) was performed. It is suggested that the best approach for further investigation is to test a codon optimized version of smURFP, since that worked for mammalian cells in the study by Rodriquez et al. (2016). Continuing work with the non-codon optimized smURFP used in this project may be in vain if the problem stems from expression or folding issues of smURFP due to codon usage discrepancies.

The previously constructed strains with green fluorescent proteins integrated in the XI-3 site showed large differences in fluorescence signals *in vivo*. The signals from yEGFP were very strong, over 23 times stronger than the signals from mEGFP in the 510/15 nm filter. At first, the strong signal from yEGFP was seen as something positive, however its fluorescence was also detected in unexpected, high levels in the 585/40 nm and 610/20 nm filters. The signals from mEGFP instead were quite moderate overall in comparison to yEGFP. Since they were more than 5-fold stronger than the signals from the background strain in the 510/15 nm filter, which was intended for detection of green fluorescence, mEGFP was still considered to be sufficiently bright for potential use in a biosensor. Being clearly distinguishable from the background and also having very little fluorescence spillover (especially in comparison to yEGFP), it was a fairly obvious choice to use mEGFP in combination with other proteins when generating the double fluorescent protein strains.

It was interesting to compare the mean fluorescence intensities of CyOFP1 and opt. CyOFP1, to see the effect of codon optimization. A previous study had shown that codon optimizing fluorescent proteins for yeast could both increase the practical brightness of the fluorescent proteins as well as decrease it, attributing these altered characteristics to changes in the folding process, function of the proteins and how they are degraded (Botman et al., 2019). The authors also found that all codon optimized fluorescent proteins in their study had increased expression levels. In this project, opt. CyOFP1 clearly outperformed its non-optimized counterpart in all filters where fluorescence signals were expected, likely at least thanks to higher expression and possibly also thanks to improved folding and functioning as suggested by Botman et al. (2019), although none of these explanations were verified in this project. In general, the experimental evidence from this project and from the study of Botman et al. seems to indicate that codon optimization results in brighter fluorescent proteins *in vivo* and if seeking the strongest possible fluorescence signals perhaps codon optimizing the proteins is the way to go. However, this must still be experimentally verified on a case by case basis and researchers must also be aware that stronger fluorescence also increases the risk of fluorescence spillover.

After assessing the mean fluorescence intensities of the proteins individually, the next step was to find suitable combinations of proteins to integrate into the same strain. At the start of the project, the proteins were chosen partly because of their emission spectra and it was thought that it would be possible to combine a green, orange and red protein. The main concern beforehand was if they would have strong enough fluorescence intensities, since too low intensities had been the problem with previously trialed proteins (unpublished data). Because of this, the main reason for choosing these proteins was their brightness's as documented in the fluorescent protein database FPbase (Lambert, 2019). After performing some of the measurement runs, it turned out that the main problem in the project actually was to find suitable combinations of proteins because of the signals being too strong and resulting in spillover.

Growth characterization of the generated double fluorescent protein strains showed promising results as their growth was not adversely affected by integration of mEGFP into the XI-2 site. The results for the mean fluorescence intensities lived up almost precisely to the expectations. The mean fluorescence intensities of TMB RP 008 (opt. CyOFP1 + mEGFP) and TMB RP 009 (opt. mBeRFP + mEGFP) detected in the 510/15 nm filter were very similar to the mean fluorescence intensity of CEN.PK+pRPF2 (mEGFP). This suggested that practically the entire signals from the double fluorescent protein strains in this filter originated from the mEGFP. Following the same reasoning, it was suggested that the signals from TMB RP 008 (opt. CyOFP1 + mEGFP) in the 585/40 nm and 610/20 nm filters were originating from the opt. CyOFP1. A similar pattern was seen for the TMB RP 009 (opt. mBeRFP + mEGFP) strain, however the signals in primarily the 585/40 nm and 610/20 nm filter were slightly higher than the signals from TMB RP 005 (opt. mBeRFP). It is possible that this is an artefact of poor statistical certainty due to just having two biological replicates but may also indicate that one or several mutations have occurred in one of, or both trialed TMB RP 009 (opt. mBeRFP + mEGFP) clones resulting in stronger signals from opt. mBeRFP. It is not likely that increased expression of mEGFP accounted for the increase in signal when expressed in the XI-2 site (in the double fluorescent protein strains) compared to the XI-3 site (in the CEN.PK+pRPF2 strain). If this was the case, there would probably have been a noticeable difference in the 510/15 nm filter, but no significant difference was recorded.

Gaining the knowledge that the same filter gave different mean fluorescence intensity results when placed in the different positions of the flow cytometer had a fairly large impact on the remaining part of the project. Before it was noticed (the first three runs out of six), it had been assumed that filters would give the same results regardless of if they were placed in the FL1, FL2 or FL3 positions. It was known from the start that the FL4 position was exclusively for detecting signals from proteins excited with the red laser and placing a filter there would definitely yield a different result. Gaining the knowledge made data handling, planning and execution of future flow cytometry measurements a bit more complex. Handling of data from these first four runs was made more tedious since all data collected using a given filter could not be compared if runs performed with the given filter were done with that filter in different positions. With this knowledge in mind, the last two runs were performed with the filters in the same positions which were deemed optimal based on an assessment of the signal strengths in the previous four runs. The differences between the filter positions are likely caused by variations in voltage settings for the photomultiplier tubes in the flow cytometer. In future experiments, it is necessary to consider the differences to make sure that data from different fluorescent proteins is compared in a correct manner.

Dot plots can be valuable tools to visualize the differences in fluorescence between different strains and to visualize the heterogenous fluorescence of the strain populations that can be easy to forget

about when using the mean fluorescence intensities of the populations for analysis. These plots can also provide a way to visualize fluorescence spillover and they may be a better tool for analysis overall since the method of color compensation is based on modifications to the dot plots.

To summarize, in this project fluorescent protein candidates were successfully integrated into the genome of *S. cerevisiae* CEN.PK113-7D, and three of these showed potential for use in biosensors in the future. One of the tested fluorescent proteins did not work at all, even when efforts were made to supply it with its required cofactor. Strains with combinations of two fluorescent proteins were successfully generated and behaved essentially as expected, which was promising for the continuation of the project. Reaching the goal of having three compatible fluorescent proteins chromosomally integrated in one strain is technically not too far away but may prove to be a complicated process as more fluorescent proteins need to be tested to find combinations of proteins with as little fluorescence spillover as possible. Lastly, it seems far-fetched that a combination of proteins without any fluorescence spillover at all can be found and therefore it is recommended to further investigate the method of color compensation to be able to take care of unavoidable fluorescence spillover.

Acknowledgements

I would like to thank my supervisors Marie Gorwa-Grauslund and Raquel Perruca Foncillas for valuable input throughout the project and for thorough feedback regarding the report. I especially want to thank Raquel for teaching me proper lab techniques and always being available to answer questions about the lab work.

I would also like to thank all the staff and fellow students at the department for a great Spring. I have really enjoyed carrying out my project alongside you all!

References

Abcam plc and Davies, D. (2020). *Fluorescence compensation in flow cytometry*. Available at: https://www.abcam.com/protocols/fluorescence-compensation-in-flow-cytometry [Accessed 27 May 2020].

BD Biosciences. (2014). *BD Accuri ™ C6 Flow Cytometer Optical Filter Guide*. [Fact sheet]. Available at: https://www.bdbiosciences.com/documents/BD_Accuri_Optical_Filter_Guide.pdf [Accessed 1 June 2020].

BD Biosciences. (2015). *BD Accuri ™ C6 Flow Cytometer, Technical Specifications*. [Fact sheet]. Available at: https://www.bdbiosciences.com/documents/Accuri_C6_TechSpecs.pdf [Accessed 1 June 2020].

Bergdahl, B., Sandström, A. G., Borgström, C., Boonyawan, T., van Niel, E. W. J. and Gorwa-Grauslund, M. F. (2013). Engineering Yeast Hexokinase 2 for Improved Tolerance Toward Xylose-Induced Inactivation. *PLoS ONE*, 8 (9), pp.1–10. doi:10.1371/journal.pone.0075055.

Borodina, I. and Nielsen, J. (2014). Advances in metabolic engineering of yeast Saccharomyces cerevisiae for production of chemicals. *Biotechnology Journal*, 9 (5), pp.609–620. doi:10.1002/biot.201300445.

Botman, D., de Groot, D. H., Schmidt, P., Goedhart, J. and Teusink, B. (2019). In vivo characterisation of fluorescent proteins in budding yeast. *Scientific Reports*, 9 (1), pp.1–14. doi:10.1038/s41598-019-38913-z.

Chu, J., Oh, Y., Sens, A., Ataie, N., Dana, H., Macklin, J. J., Laviv, T., Welf, E. S., Dean, K. M., Zhang, F., Kim, B. B., Tang, C. T., Hu, M., Baird, M. A., Davidson, M. W., Kay, M. A., Fiolka, R., Yasuda, R., Kim, D. S., Ng, H-L. and Lin, M. Z. (2016). A bright cyan-excitable orange fluorescent protein facilitates dual-emission microscopy and enhances bioluminescence imaging in vivo. *Nature Biotechnology*, 34 (7), pp.760–767. doi:10.1038/nbt.3550.

Cormack, B. P., Bertram, G., Egerton, M., Gow, N. A. R., Falkow, S. and Brown, A. J. P. (1997). Yeast-enhanced green fluorescent protein (yEGFP): A reporter of gene expression in Candida albicans. *Microbiology*, 143 (2), pp.303–311. doi:10.1099/00221287-143-2-303.

Cormack, B. P., Valdivia, R. H. and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, 173 (1), pp.33–38. doi:10.1016/0378-1119(95)00685-0.

Cranfill, P. J., Sell, B. R., Baird, M. A., Allen, J. R., Lavagnino, Z., De Gruiter, H. M., Kremers, G. J., Davidson, M. W., Ustione, A. and Piston, D. W. (2016). Quantitative assessment of fluorescent proteins. *Nature Methods*, 13 (7), pp.557–562. doi:10.1038/nmeth.3891.

D'Ambrosio, V. and Jensen, M. K. (2017). Lighting up yeast cell factories by transcription factor-based biosensors. *FEMS Yeast Research*, 17 (7), pp.1–12. doi:10.1093/femsyr/fox076.

Gietz, R. D. and Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, 2 (1), pp.31–34. doi:10.1038/nprot.2007.13.

Givan, A. L. (2011). Flow Cytometry: An Introduction. In: Hawley, T. S. and Hawley, R. G. (Eds). *Flow Cytometry Protocols*. 3rd Ed. New York: Humana Press. pp.1–30. doi:https://doi.org/10.1007/978-1-61737-950-5_1.

Jessop-Fabre, M. M., Jakočiūnas, T., Stovicek, V., Dai, Z., Jensen, M. K., Keasling, J. D. and Borodina, I. (2016). EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes into Saccharomyces cerevisiae via CRISPR-Cas9. *Biotechnology Journal*, 11 (8), pp.1110–1117. doi:10.1002/biot.201600147.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012). *A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity*. 337 (6096), pp.816–821. doi:10.1126/science.1225829.

Lambert, T. J. (2019). FPbase: a community-editable fluorescent protein database. *Nature Methods*, 16 (4), pp.277–278. doi:10.1038/s41592-019-0352-8.

Mikkelsen, M. D., Buron, L. D., Salomonsen, B., Olsen, C. E., Hansen, B. G., Mortensen, U. H. and Halkier, B. A. (2012). Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metabolic Engineering*, 14 (2), pp.104–111. doi:10.1016/j.ymben.2012.01.006.

Mumberg, D., Müller, R. and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene*, 156 (1), pp.119–122. doi:https://doi.org/10.1016/0378-1119(95)00037-7.

Oldenburg, K. R., Vo, K. T., Michaelis, S. and Paddon, C. (1997). Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Research*, 25 (2), pp.451–452. doi:10.1093/nar/25.2.451.

Ran, F. A., Hsu, P. D., Lin, C.-Y., Gootenberg, J. S., Konermann, S., Trevino, A., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y. and Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154 (6), pp.1380–1389. doi:10.1016/j.cell.2013.08.021.

Rodriguez, E. A., Tran, G. N., Gross, L. A., Crisp, J. L., Shu, X., Lin, J. Y. and Tsien, R. Y. (2016). A far-red fluorescent protein evolved from a cyanobacterial phycobiliprotein. *Nature Methods*, 13 (9), pp.763–769. doi:10.1038/nmeth.3935.

Saccharomyces Genome Database. (2020). *HMX1/YLR205C Overview*. Available at: https://www.yeastgenome.org/locus/S000004195 [Accessed 27 May 2020].

Szczesny, R. J., Kowalska, K., Klosowska-Kosicka, K., Chlebowski, A., Owczarek, E. P., Warkocki, Z., Kulinski, T. M., Adamska, D., Affek, K., Jedroszkowiak, A., Kotrys, A. V., Tomecki, R., Krawczyk, P. S., Borowski, L. S. and Dziembowski, A. (2018). Versatile approach for functional analysis of human proteins and efficient stable cell line generation using FLP-mediated recombination system. *PLoS ONE*, 13 (3), pp.1–29. doi:10.1371/journal.pone.0194887.

Taxis, C. and Knop, M. (2006). System of centromeric, episomal, and integrative vectors based on drug resistance markers for Saccharomyces cerevisiae. *BioTechniques*, 40 (1), pp.73–78. doi:10.2144/000112040.

Wikimedia Commons user Mariuswalter. (2020). *DNA Repair-colourfriendly.png*. https://commons.wikimedia.org/wiki/File:DNA_Repair-colourfriendly.png. [Accessed 2 June 2020].

Yang, J., Wang, L., Yang, F., Luo, H., Xu, L., Lu, J., Zeng, S. and Zhang, Z. (2013). mBeRFP, an Improved Large Stokes Shift Red Fluorescent Protein. *PLoS ONE*, 8 (6), pp.6–11. doi:10.1371/journal.pone.0064849.

Zacharias, D. A., Violin, J. D., Newton, A. C. and Tsien, R. Y. (2002). Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science*, 296 (5569), pp.913–916. doi:10.1126/science.1068539.

Appendices

A. Primers used in the project

Table A1. List of primers used in the project. In sequences with both upper- and lower-case letters, lower case indicates the sequences annealing to a gene.

Number/name	Sequence (5' → 3')	Used for
23	GGCCGTTATTTGTGCTTGAT	Colony PCR verifications.
24	CGGTTGTGATATTGTTCCTGC	Colony PCR verifications.
27	TACAAACTCGAGTATatgaaaacttctgaacaacg	Amplification of *smURFP*, adding *Xho*I restriction site. Colony PCR verifications.
28	ATGCGTGCGATCGCTTAggacatagccttgatgat	Amplification of *smURFP*, adding *Sfa*AI restriction site. Colony PCR verifications.
31	TACAAACTCGAGTATatggtgagcaagggc	Amplification of *CyOFP1*, adding *Xho*I restriction site. Colony PCR verifications.
32	ATGCGTGCGATCGCttacttgtacagctcgtccat	Amplification of *CyOFP1*, adding *Sfa*AI restriction site. Colony PCR verifications.
33	ATGGTGTCTAAGGGTGAAGAG	Colony PCR verifications.
34	TCACTTGTACAACTCATCCATACC	Colony PCR verifications.
35	ATGGTTTCTAAAGGTGAAGAAT	Colony PCR verifications.
36	TTATTTATGACCCAATTTAGATGGC	Colony PCR verifications.
41	AAAGGATCCTATatgagtgtcaacttagcttc	Amplification of *HO-1*, adding *Bam*HI restriction site. Colony PCR verifications.
42	TTAGAATTCGTCctagccttcggaggtg	Amplification of *HO-1*, adding *Eco*RI restriction site. Colony PCR verifications.
43	CGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGaaagtgcccatcagtgt	Amplification of *AUR1p*, *AUR1-C* and *AUR1t*, adding homology upstream of *URA3* in pUG62-AUR
44	CAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGcagaggaaagaataacgcaa	Amplification of *AUR1p*, *AUR1-C* and *AUR1t*, adding homology downstream of *URA3* in pUG62-AUR
117	CTTAAACTTCTTAAATTCTAC	Forward primer for sequencing of *HO-1* inserted between *TDH3p* and *CYC1t*
118	GAATGTAAGCGTGACATAAC	Reverse primer for sequencing of *HO-1* inserted between *TDH3p* and *CYC1t*
13	GAACTTCAGGGTCAGCTTGC	Colony PCR verifications.
324	ACTGGGAACAGAAATCGACC	Colony PCR verifications.
seq1	ACTAGTCCTGCAGATAGCTTC	Forward primer for sequencing of genes inserted between *TEF1p* and *ADH1t*
seqADH1t	TCGCTTATTTAGAAGTGTCAAC	Reverse primer for sequencing of genes inserted between *TEF1p* and *ADH1t*

B. Yeast transformation protocol used

Materials:

- YPD medium
- Sterile MQ water
- LiOAc solution: 100 mM Lithium acetate, 10 mM TRIS-HCL pH 7.5, 1 mM EDTA
- PEG 3350 (50%)
- Lithium acetate (1.0 M)
- TRIS-HCL (0.5 M, pH 7.5)
- EDTA (100 mM)
- ssDNA (2 mg/mL)
- DMSO (100%)

Method:

1. Inoculate 5 mL YPD medium supplemented with 5 µL G418 and incubate overnight in a shaking incubator at 30°C
2. Prepare baffled shake flasks with 25 mL YPD medium and put them in the 30°C incubator to pre-warm the medium, saving 1-1.5 cultivation time the next day.
3. Use the pre-culture to inoculate the pre-warmed 25 mL YPD medium to and OD_{620} of 0.4. Add 25 µL G418. Incubate in a shaking incubator at 30°C until the OD has reached about 1.5. This will take approximately 4 hours. Make sure the cells have divided at least 1.5 times.
4. Boil a 1.0 mL sample of ssDNA for 5 min and quickly chill on ice water. (It is not necessary or desirable to boil the carrier DNA every time. Keep a small aliquot in your own freezer box and boil after 3-4 freeze thaws. But keep on ice when out.)
5. Harvest the culture in a sterile 50 mL Falcon tube at 3000 rpm for 5 minutes.
6. Decant the supernatant. Resuspend the cells in 25 mL sterile MQ water and repeat centrifugation.
7. Decant the water and resuspend the cells in 1 mL LiOAc solution.
8. Pellet the cells at top speed for 15 seconds and remove the LiOAc solution with a micropipette.
9. Resuspend the cells with LiOAc solution to a final volume of 500 µL ± 400 µL. Adjust the volume to the OD_{620} of the culture you harvested, resuspension volume = OD x shake flask volume x 0.01.
10. Vortex the cell suspension and pipette 50 µL samples into labelled tubes.
11. Carefully add the following ingredients in the order listed:
    a. 240 µL PEG (50% w/v)
    b. 36 µL Lithium acetate (1.0 M)
    c. 8 µL TRIS-HCL (0.5 M, pH 7.5)
    d. 4 µL EDTA (100 mM)
    e. 25 µL ssDNA (2.0 mg/mL)
    f. X µL of plasmid DNA (volume necessary to reach 0.1-10 µg of plasmid)
    g. 50 - X µL water
12. Vortex each tube until the cell pellet has been completely mixed. Usually takes about 30 seconds.
13. Incubate the transformation mixture at 30°C for 30 minutes, standing still.

14. Add 40 µL DMSO (100%) to the transformation mixture and vortex quickly to mix. The DMSO enhances the permeability of the cells but be careful since it might be toxic to your cells.
15. Heat-shock the cells at 42°C for 20 minutes in water bath.
16. Centrifuge at 8000 rpm for 15 seconds and remove the transformation mixture with a micropipette.
17. In case of making gene disruptions with antibiotic marker, resuspend the cells in 500 µL YPD medium and incubate for two hours in the shaking incubator at 30°C before plating.