Construction & Evaluation of a Reporter Gene Displaying Aldehydes on the Cell Surface

By

Christine Wong

A thesis submitted for the requirements for the Master of Science in Chemistry

Department of Chemistry and Biomolecular Sciences

University of Ottawa

Ottawa, Ontario, Canada

June 2020

Abstract

Reporter genes are often used to observe expression of promoters, which may change from its natural behaviour as a result of stress or disease states. Reporter genes are useful because they are easily detectable by a variety of imaging methods, including fluorescence microscopy techniques, magnetic resonance imaging, and positron emission tomography. Previously, methyl 5-MeO-*N*-aminoanthranilate (MMNA) had been synthesized as an aldehyde-conditional fluorophore and was tested in physiological conditions to identify the Aldehydic Load of cells. Thus, it was hypothesized that a reporter protein displaying an aldehyde on the cell surface can be identified by MMNA. This reporter protein would contain a substrate recognition site for formylglycine generating enzyme (FGE) that converts a specific cysteine residue into a formylglycine residue. This will result in production of an aldehyde at the N-terminal of the transmembrane domain of platelet derived growth factor receptor β. In this way, the protein product, Aldehyde-presenting FGE-dependent Readout (Alfred), would display an aldehyde on the extracellular surface of the cell. Alfred was expressed in A549 human lung cancer cells using the Tet-On® Inducible System, which allows expression of a gene of interest by use of doxycyclin (dox) as a chemical trigger. Microscopy of Alfred-transfected cells, induced by dox and probed with MMNA, showed no difference in fluorescence between non-transfected and Alfred-transfected cells. The overexpression of FGE to increase thiol-to-aldehyde conversion, and the imaging of cells at longer timepoints (48 and 72 hours) to allow localization of the protein to the cell surface, were attempted. In addition, Alfred was constitutively expressed in another transfection experiment in efforts to increase gene expression. However, these efforts to evaluate Alfred did not improve the microscopy results. Western

blotting confirmed FGE overexpression in transgenic cells. Blotting against the Myc-tag in Alfred showed no detected proteins in Alfred-transfected cells. In conjunction with the microscopy images, these results suggest that Alfred is not expressed and cannot be detected as a reporter gene. Comparison to previous works allows the identification of potential approaches to improve Alfred functionality, including the absence of the hemagglutinin epitope, the choice of aldehyde probe used, the choice of cell line used, and the method of analyzing microscopy. Future directives are postulated to identify sources that hinder Alfred expression, and to improve visualization of Alfred over homeostatic aldehydes.

Dedication

To my wonderful parents and sister who never cease to encourage and inspire me. Thank you each for supporting me in your own ways. 多謝你哋咁多年嚟嘅支持。

Acknowledgements

I would like to thank Dr. Adam Shuhendler for providing the opportunity to work on this project, and for his guidance and support throughout the process. In addition, I thank the Molecular Medicine Lab, in particular Alexia Kirby, Yen Truong, Nick Calvert, and Anthony Rossl, for their help in eukaryotic cell culture technique and providing invaluable advice.

I would also like to thank the other labs in Marion 2, notably the Host-Microbe Interactions Laboratory, the Boddy lab, and the daCosta]:[lab for their help and expertise, and generally making this experience fun and exciting.

Statement of Contribution

Dr. Adam Shuhendler and Melissa McNulty collaborated in creating the Alfred gene sequence.

Dr. Adam Shuhendler and Anthony Rossl both created gene maps shown in this report.

Dr. Derrick Gibbings gifted the Tet-On® Inducible Systems plasmids to the Molecular Medicine Lab.

Anthony Rossl and Jordan Brazeau-Henrie both converted the Sanger sequencing results and compared the amino acid sequence to the reference sequence.

XL1-blue cells were provided by the Boddy Lab.

A549 cells were donated by Dr. Christina Addison.

MDA-MB-231 and MDA-MB-231-BR cells were donated by Dr. Paula Foster.

Kyle Tomaro provided the positive control containing a Myc-tagged protein for use in verifying antibody fidelity in Western blot.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Amp (Amp'; AmpR)	ampicillin (Amp resistance)
BME	β-mercaptoethanol
BSA	bovine serum albumin
CEST-MRI	chemical exchange saturation transfer magnetic resonance imaging
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
dox	doxycyclin
ECMV	Encephalomyocarditis virus
FBS	fetal bovine serum
FGE	formylglycine generating enzyme
FRET	fluorescent resonance energy transfer; förster resonance energy transfer
G418	geneticin
GFP	green fluorescent protein
GOI	gene of interest
HA	hemagglutinin
IgG	immunoglobin
IRES	internal ribosome entry site
l-Gln	l-glutamine
MCS	multiple cloning site
MMNA	methyl 5-MeO-*N*-aminoanthranilate
Neo (Neo'; NeoR)	neomycin (Neo resistance)

NSCLC	non-small-cell lung cancer
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDGFR	platelet derived growth factor receptor
penstrep	penicillin-streptomycin
PET	positron emission tomography
SUMF1	sulfatase modifying factor 1
TM	transmembrane
uTM	unaltered transmembrane protein

1. Introduction

1.1. Reporter genes

In the quest to observe natural and disease cellular behaviour, identifying gene expression has become an important technique in research. Some applications include analyzing expression of promoters, describing protein interactions, or simply verifying successful transfection of vectors into cells. The most convenient method is to use reporter genes, which are easy to detect and quantify [1,2,3]. Reporter genes allow visualization of transcription through induction of the promoter, and thus the reporter gene is co-expressed with the gene of interest (GOI) [1]. The reporter gene is ligated before or after the GOI, and always after the promoter [1]. By co-expressing the reporter, the activity of the promoter can be assessed easily instead of quantifying the translated protein, since the protein may be degraded or post-translationally modified before evaluation [1]. Second, protein interaction can be analyzed by attaching different reporter genes to the GOI, and observing the localization of the reporter proteins, provided that the reporter protein does not affect normal activity and localization of the protein of interest [4]. Third, transfection success can be verified by inducing the promoter and seeing reporter gene expression before continuing with the rest of the experiment [5]. It is important that the reporter gene is not endogenous to the cells being observed, or else visualization of the reporter gene would not only show the presence of the GOI.

Common reporter gene products are fluorescent proteins. These are available in a variety of colours with maximum excitation and emission at different wavelengths, including the well-known green fluorescent protein (GFP) originally derived from *Aequorea victoria* [2],

which emits in the green range of the visible light spectrum. Modifications to this protein have been made, which include increasing or decreasing the brightness, optimizing the excitation and emission curves, and improving pH resistance [2]. These developments allow the use of fluorescent proteins in new technologies including fluorescent resonance energy transfer (FRET) for identifying protein interactions [2], fluorescent timers which change colour at a constant rate over time [6], and MS2-GFP which requires a bacteriophage RNA-binding protein to be fused with GFP to bind specifically to certain viral RNA hairpin loops [3]. While many reporter genes already exist, new methods of visualizing promoter expression will always be relevant in order to tackle the various experimental conditions of new and upcoming research.

1.2. Formylglycine generating enzyme and the aldehyde tag

Formylglycine generating enzyme (FGE), also known as sulfatase modifying factor 1 (SUMF1), is endogenously expressed in the endoplasmic reticulum of eukaryotic cells, and modifies sulfatases to be used in lysozymes or to be secreted [7,8,9]. In order to target and modify the sulfatases, FGE recognizes a specific 13-mer of amino acids, L**C**TPSRAALLTGR, in the unfolded polypeptide [7]. This 13-mer can be minimized to a 6-mer, where the first 6 amino acids of the 13-mer sequence will suffice for enzyme recognition [7]. In this sequence, the enzyme converts the cysteine residue to a formylglycine residue, where a thiol is oxidized to become an aldehyde (Figure 1.2.1) [7,8]. This specific aldehyde will be referred to as the "aldehyde tag".

Previously, Wu et al. have created various fusion proteins that expressed the aldehyde tag [7]. They incorporated the 13-mer or 6-mer recognition sequence into monoclonal antibody F_c fragments, full-length monoclonal antibodies, cell surface proteins, and cytosolic proteins [7]. In

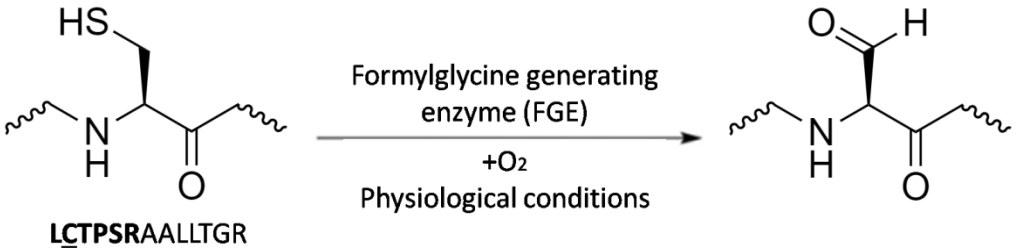

Figure 1.2.1. Formylglycine generating enzyme converts the cysteine residue in a specified 13-mer of amino acids to a formylglycine residue. FGE recognizes a 13-mer or 6-mer (in bold) peptide substrate depicted below the cysteine substrate, converting the underlined cysteine residue to a formylglycine residue by oxidizing the thiol into an aldehyde.

each experiment, they positioned the 13- or 6-mer recognition sequence at the N-terminal [7]. They overexpressed FGE in their cells, and after expressing the protein, they probed with an aldehyde probe (aminooxy-FLAG or biotin hydrazide) [7]. They also created mutant constructs, where the critical cysteine residue was replaced by an alanine residue [7]. They discovered that the proteins without the 13-mer sequence (the wild type and the mutant 13-mer construct) were not detectable by the probe, whereas the proteins with the wild type 13-mer sequence were detected [7]. In the antibodies, they were also able to calculate percent conversion of cysteine to formylglycine, showing that tagging at the N-terminal with the 13-mer resulted in 67% conversion, 91% conversion with transient overexpression of FGE, and 77% conversion in cells stably transfected with FGE [7]. Studying the cell surface proteins, they discovered that both transmembrane (TM) and membrane-associated proteins were able to be modified by FGE [7]. In the cytosolic proteins, they noticed that there was a low level of detection by the probe [7]. They determined that the modification was likely due to contamination during the cell lysis process, and also hypothesized that FGE activity may be present in the cytosol, though it has previously been uncharacterized [7]. Overall, the study showed that FGE conversion of the cysteine to

formylglycine is specific to the recognition sequence, and the aldehyde tag can be introduced in monoclonal antibodies, cell membrane-associated proteins, and cytosolic proteins [7].

1.3. Probes for the detection of aldehyde tags

Since none of the 20 essential amino acids contain an aldehyde functional group, this allows an opportunity to label or tag proteins at a specific location by introducing an aldehyde. One method is to create a protein containing the FGE recognition sequence, allowing the creation of this specific aldehyde via FGE. Another method of introducing an aldehyde is by creating an unnatural amino acid with an aldehyde group [10]. The latter method requires the nonsense-codon suppressor tRNA strategy, or expressed protein ligation technology [11]. These two methods, while more versatile in the functional group they can add, are more complex. Using FGE allows creation of the aldehyde through endogenous cellular mechanisms, since the protein will be traversing the endoplasmic reticulum *en route* to the cell surface [12]. This is a much simpler method than introducing an unnatural amino acid, and may facilitate broader adoption in a range of cell types.

In order to tag these aldehydes, an aldehyde probe can be used. An example of an aldehyde probe utilizes a hydroxylamine or aminooxy functional group (Figure 1.3.1.) [13]. The nitrogen in the aminooxy functional group acts as a nucleophile to attack the carbonyl carbon in the aldehyde, pushing the electrons in the pi bond to the oxygen [10]. Since the oxygen is now negatively charged, one pair of its electrons attacks one of the hydrogen atoms bound to the nitrogen [13]. Then, the condensation reaction occurs; simultaneously the lone pair of the nitrogen reacts with the carbon of the aldehyde group (now a hydroxide group) [13]. This results in a water molecule and the oxime product [13].

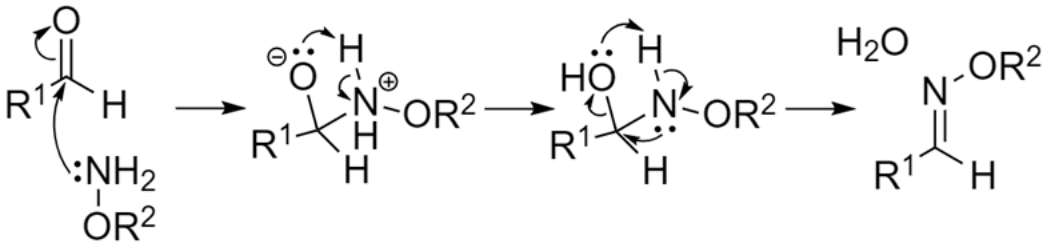

Figure 1.3.1. Mechanism of an aminooxy functional group and an aldehyde reaction to create an oxime. This reaction can also proceed with hydroxylamine, in which the R^2 group represents a hydrogen atom, also leading to an oxime product.

For the purpose of imaging, the R^2 group in Figure 1.3.1., which is attached to the nitrogen in the hydroxylamine, can be a dye or a fluorescent molecule. In a study by Tuley et al., an unnatural amino acid, *meta*-formyl-phenylalanine was reacted with coumarin dye which contained the aminooxy group [10]. Site specific incorporation of *meta*-formyl-phenylalanine into sfGFP allowed the aminooxy-aldehyde reaction to be monitored via FRET [10]. In the study by Wu et al. discussed earlier, the aminooxy functional group was bound to Alexa Fluor 647, a red fluorescent dye [7]. This was used to identify the aldehydes on the aldehyde-tagged TM protein that they investigated [7].

Another example of an aldehyde reactive group is a hydrazine (Figure 1.3.2.). Similarly to the reaction with aminooxy functional group, the carbon in the carbonyl group undergoes a nucleophilic attack from the terminal nitrogen in the hydrazine group. This is followed by isomerization, moving the hydrogen from the nitrogen to the oxygen. The product is formed by a condensation reaction and loss of the oxygen in the aldehyde group, resulting in the nitrogen-carbon pi bond to form the hydrazone. This can be a very fast reaction, quicker than strain-promoted click chemistry [14].

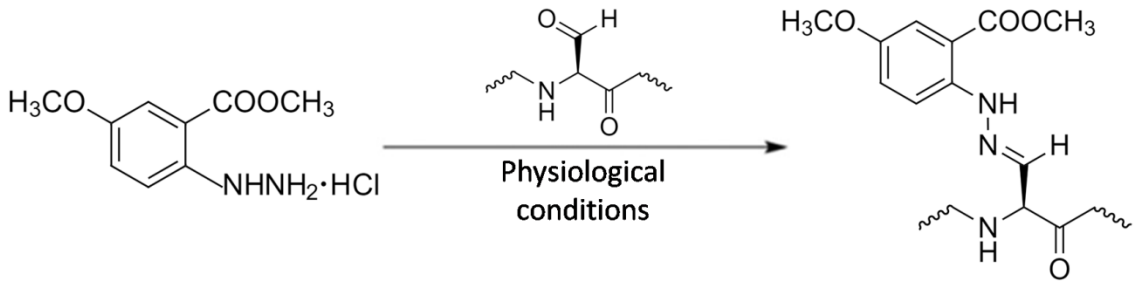

Figure 1.3.2. Reaction of a hydrazine and an aldehyde to create a hydrazone. Methyl 5-MeO-*N*-aminoanthranilate is reacted with a formylglycine residue within a protein to site-specifically tag the protein.

Methyl 5-MeO-*N*-aminoanthranilate (MMNA), previously described by the Molecular Medicine Lab, is an example of a hydrazine, and is fairly easy to synthesize [14]. 5-MeO-*N*-aminoanthranilate was originally designed as a chemical exchange saturation transfer magnetic resonance imaging (CEST MRI) agent, and esterification of this molecule resulted in MMNA [14]. The hydrazone product of the reaction of MMNA with an aldehyde was observed to be fluorescent in the blue-green region of the visible light spectrum [14]. One advantage of MMNA is that its fluorescence is conditional upon aldehyde binding. Since the product of the MMNA-aldehyde reaction is the fluorescent compound, the probe requires the reaction to produce a signal, and therefore does not require washing away the excess probe [14]. Another advantage is that MMNA has a broad reactivity for aldehydes, and is expected to react with formylglycine residues produced by FGE [14]. The excitation and emission matrix of each hydrazone product can be characterized, allowing fluorescent fingerprinting to identify carbonyl compounds within a mixed sample [15]. Thus, MMNA allows determination of the total aldehyde content in the sample by live cell microscopy, an emerging biomarker called the Aldehydic Load [14].

1.4. The Tet-On® Inducible System

The Tet-On® Inducible System created by Takara Bio USA, Inc (Clontech® Laboratories, Inc) is a two-component vector system to induce gene expression with a doxycycline (dox) trigger [5,16]. In particular, pTRE3G-ZsGreen1 and pCMV-TET3G are 2 vectors that can be used together (Figure 1.4.1.) [5]. pCMV-TET3G contains the $p_{CMV\ IE}$ promoter to express the Tet-On® 3G gene [5]. This vector is selectable with neomycin resistance (Neo') and ampicillin resistance (Amp') (Figure 1.4.1.) [5]. pTRE3G-ZsGreen1 contains the p_{TRE3G} promoter expressing a reporter gene called ZsGreen (its excitation and emission curves are very similar to GFP), followed by an internal ribosome entry site (IRES) and a multiple cloning site (MCS) to express and insert the gene of interest (GOI) [5]. This vector is selectable with ampicillin resistance (Amp') and ZsGreen fluorescence with a fluorescence-based cell sorter (Figure 1.4.1.) [5].

Once the strong promoter $p_{CMV\ IE}$ (hereafter referred to as p_{CMV}) is constitutively induced by the host cell, the Tet-On® 3G gene is expressed to create a transactivator protein (hereafter referred to as TET3G) [5,16]. When bound to dox, TET3G undergoes a conformational change that allows it to selectively bind to the *tet* sequence within p_{TRE3G} in pTRE3G-ZsGreen1, inducing transcription of ZsGreen and the GOI (Figure 1.3.2.) [5,16]. TET3G does not bind to endogenous promoters within the cell, and without dox binding, TET3G does not promote transcription of p_{TRE3G} [5,16]. In addition, there is no transcription induced by p_{TRE3G} in the absence of the active TET3G because p_{TRE3G} does not have the normal enhancer elements associated with strong promoters [16]. Thus, expression of the GOI can be turned on using dox as a chemical trigger, and turned off in the absence of dox.

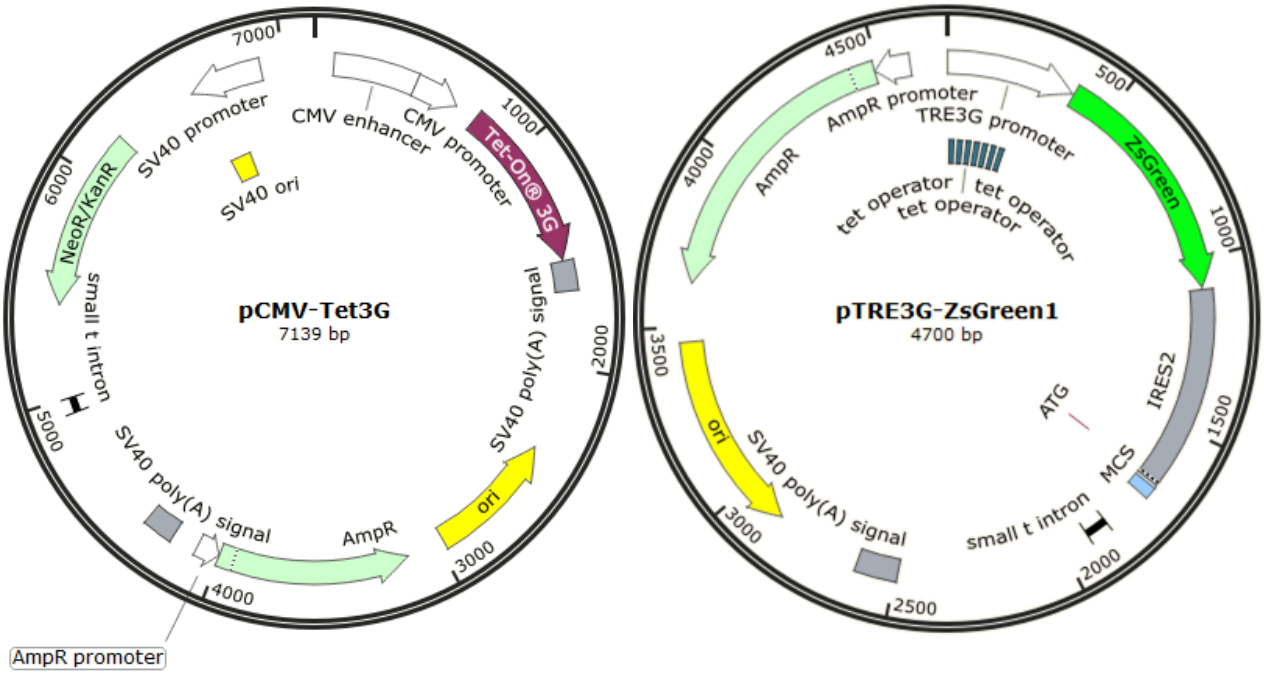

Figure 1.4.1.Plasmid map of pCMV-TET3G and pTRE3G-ZsGreen1 in the Tet-On® Inducible System. pCMV-TET3G constitutively expresses Tet-On® 3G, a gene that encodes for the TET3G transactivator protein. TET3G, when bound to doxycycline (dox), binds the p_{TRE3G} promoter in pTRE3G-ZsGreen1 to express ZsGreen. The gene of interest can be ligated into the multiple cloning site (MCS) within pTRE3G-ZsGreen1, to be expressed in the presence of dox.

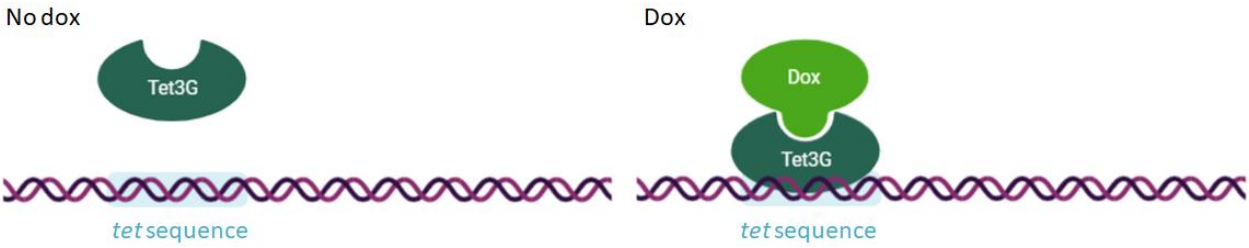

Figure 1.4.2. Binding of the TET3G transactivator protein to the *tet* recognition sequence within p_{TRE3G}. TET3G does not bind to the *tet* sequence within p_{TRE3G} in the absence of dox. Upon dox binding, transcription is activated.

The IRES located after ZsGreen should not affect transcription, but rather it is related to translation. Originally discovered in viral protein synthesis, it is located in the 5'UTR of the mRNA [17]. It facilitates translation in a cap-independent fashion by using complex RNA secondary structures to allow initiation factor and IRES trans-acting factor docking, eventually enabling ribosome docking (hence, internal) [17], whereas normally, ribosomes dock via the 5' guanine cap on mRNA [18]. The secondary structure allows the classification of the IRES into 4 types (I to IV), even though more than 50% of the sequence of the IRES is not conserved [17]. The IRES included in pTRE3G-ZsGreen1 is from Encephalomyocarditis virus (ECMV), which classifies it as a Type II IRES [19].

1.5. Alfred

This project was conceptualized from the findings of Wu et al. [7]. A new reporter gene is created in this project, named Alfred to represent Aldehyde-presenting FGE-dependant readout. The Alfred gene contained MluI and BamHI restriction enzyme sites, start and stop codons, the leader sequence of an immunoglobin (IgG) κ light chain, the 13-mer amino acid sequence recognized by FGE, a human c-Myc proto-oncogene epitope tag, and the TM domain of platelet derived growth factor receptor β (PDGFRβ); the gene length totalled to 297 bp (Figure 1.5.1.). Unlike Wu et al.'s construct, Alfred does not contain the hemagglutinin A epitope nor the multiple cloning site residues (MCS) (Figure 1.5.2.). This was because these residues were hypothesized to not play a functional role in creation of the aldehyde or localization of this aldehyde on the cell surface.

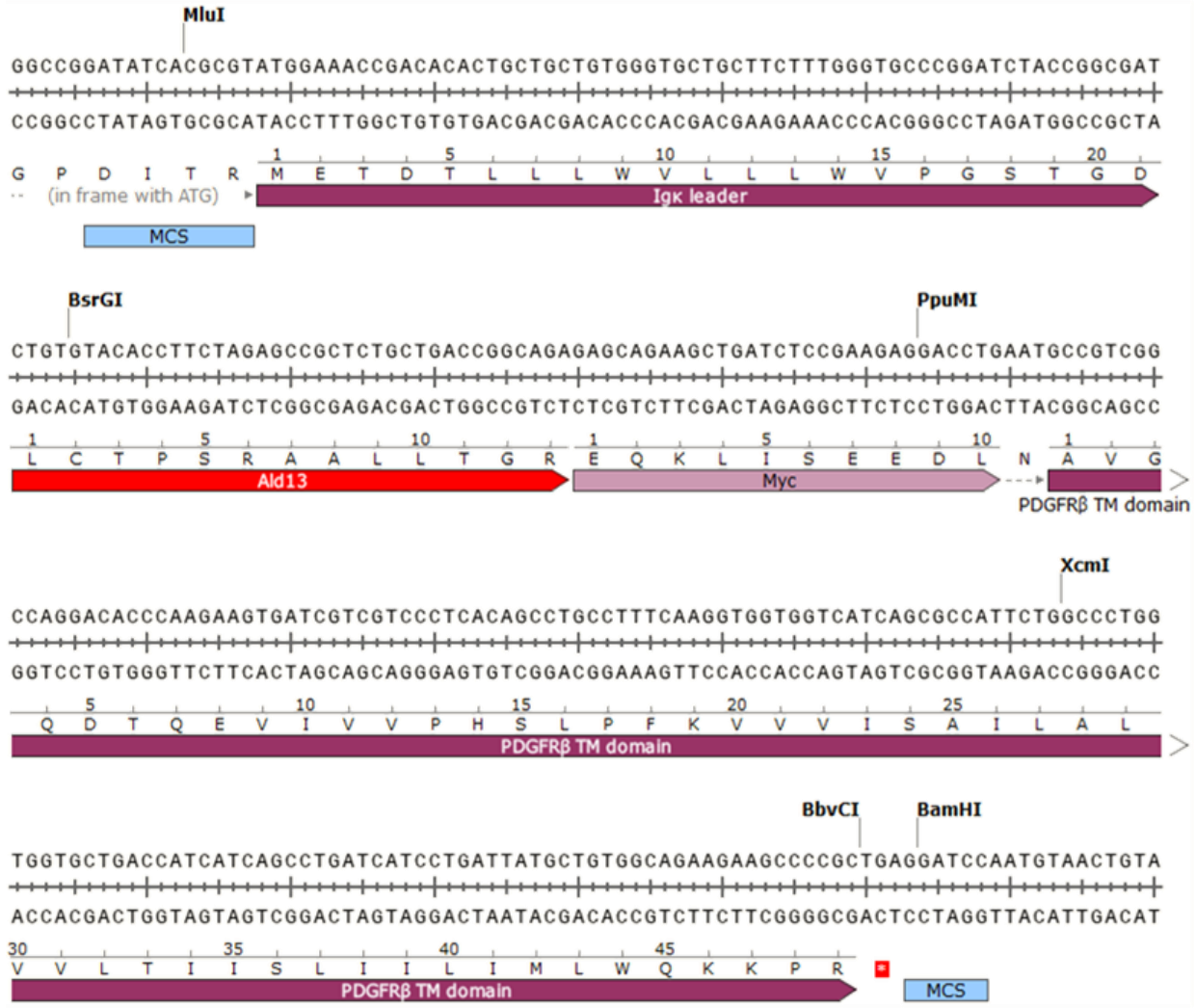

Figure 1.5.1. Gene sequence of the Alfred gene. The gene map was created in SnapGene. The gene is composed of the IgG κ leader sequence (Igκ leader), the FGE substrate site (Ald13), the Myc epitope tag (Myc), and the platelet derived growth factor receptor β transmembrane domain (PDGFRβ TM domain). This was ligated into pTRE3G-ZsGreen1 *via* the MluI and BamHI restriction sites.

The Tet-On® Inducible System was used in order to control Alfred expression. The Alfred gene is the GOI, and was ligated using Gibson assembly into the MCS of pTRE3G-ZsGreen1 (Figure 1.4.1.) to create pTRE3G-ZsGreen1-Alfred (sequence in Appendix A). In Gibson assembly, the overlapping sequence is usually created with PCR primers; however

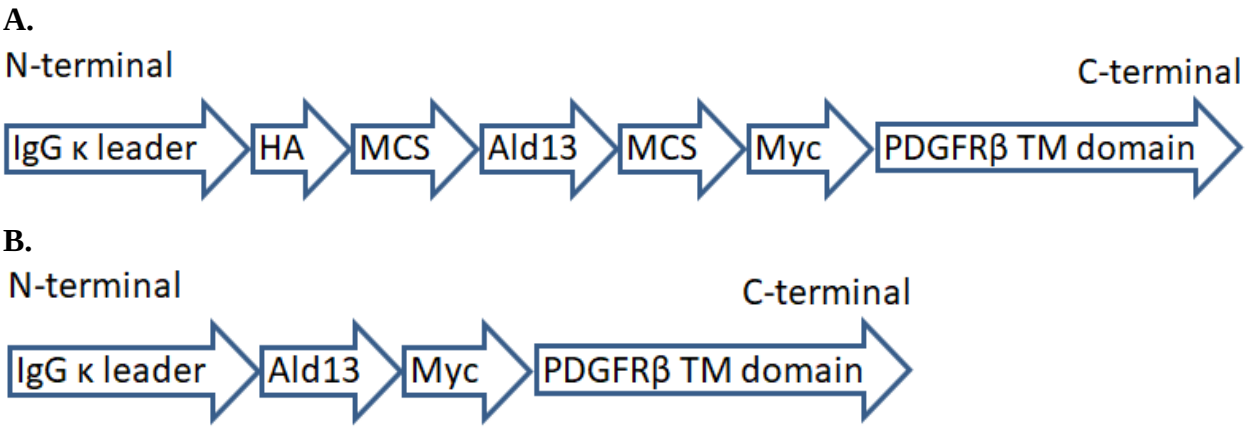

Figure 1.5.2. Comparison of Alfred design to that of a previously published reporter gene. The immunoglobin κ light chain leader sequence (IgG κ leader), FGE 13-mer recognition site (Ald13), Myc epitope tag (Myc), and the platelet derived growth factor receptor β transmembrane domain (PDGFRβ TM domain) are present in **A.** Wu et al.'s construct [7] as well as **B.** Alfred. The hemagglutinin A epitope (HA) and multiple cloning site (MCS) residues are not present in Alfred.

in this project, Gibson assembly was possible because of the overlapping restriction enzyme sites. Both the vector and the gene contained MluI and BamHI, coordinated in the correct order for the gene to be ligated in the appropriate orientation to be transcribed with the p_{TRE3G} promoter.

Transcription of the Alfred gene should occur with the addition of dox. During IRES-dependent translation, the signal polypeptide sequence contained in the IgG κ leader sequence should target the polypeptide for translocation to the endoplasmic reticulum, and the critical cysteine residue should be modified by FGE before the eventual translocation of Alfred to the cell membrane [12]. The aldehyde tag located at the N-terminal end of the PDGFRβ can then be presented on the cell surface. Thus, Alfred makes use of FGE and displays the converted aldehyde tag on the cell surface to be detected by aldehyde probes such as MMNA. Similarly to Wu et al.'s work, Alfred should be seen as punctate fluorescent foci on the cell's surface [7]. It is important to note that in this system, ZsGreen is expressed alongside Alfred, which may allow

FRET to occur between ZsGreen and MMNA. That is, when ZsGreen is excited, its emission wavelength may be released at the optimal wavelength to excite MMNA, resulting in an increased signal from MMNA. Therefore, MMNA should first be tested with ZsGreen, or a similar fluorescent protein, to ensure that FRET does not interfere with MMNA fluorescence.

One of the main features of this reporter gene is that it is invisible without an aldehyde probe. Therefore, Alfred would not interfere with the imaging of other reporters expressed simultaneously once the aldehyde probe is washed away. In addition, the reporter gene can be used as a multi-modality reporter gene system. Since the gene produces an aldehyde tag, the aldehyde can be reacted with a variety of probes, and thus be used for many different applications beyond fluorescence [14,15], such as CEST-MRI [20], and PET-based imaging [21]. Therefore, the signal from this reporter gene originates from the probe used to detect it and not from the protein itself, allowing development and optimization of new aldehyde probes.

2. Goals and Objectives

For this study, A549 cells were chosen for their ability to develop into tumours within living systems. Methyl 5-MeO-*N*-aminoanthanilate (MMNA) was chosen as the aldehyde probe for its easy detection in fluorescence microscopy and its similarity to previously designed magnetic resonance imaging (MRI) and positron emission tomography (PET) probes. These two factors would allow future studies to involve imaging of Alfred-expressing tumours in MRI and PET imaging.

Furthermore, the Tet-On® Inducible System was chosen to selectively turn on expression of Alfred using doxycyclin (dox) as a chemical trigger. Presence of dox alongside the constitutively expressed TET3G transactivator protein from the Tet-On® Inducible System will allow expression of the p_{TRE3G} promoter, resulting in cap-dependent translation of ZsGreen and internal ribosome entry site (IRES)-dependent translation of Alfred.

The Alfred gene is designed to compose of 4 main parts to display an aldehyde on the cell surface. First, the IgG κ leader sequence targets the protein for localization in the endoplasmic reticulum membrane as part of the secretory pathway. Second, the formylglycine generating enzyme (FGE) substrate sequence for cysteine-to-formylglycine conversion by FGE in the endoplasmic reticulum allows the production of an aldehyde at the N-terminal of Alfred. Third, the Myc tag can be used for detection of Alfred with anti-Myc antibodies. Finally, the transmembrane domain of the platelet derived growth factor receptor β (PDGFRβ) integrates into the cell membrane, simultaneously allowing the N-terminal aldehyde produced by FGE to be displayed on the cell surface.

The goal of this study is to display Alfred's aldehyde on the cells surface and detect this aldehyde using MMNA as a proof of concept. To accomplish this, Alfred will be inserted into pTRE3G-ZsGreen1 within the multiple cloning site (MCS) after the IRES. This system will be transfected into A549 cells, and expression of Alfred will be induced by the addition of dox. Fluorescence microscopy will be used to image ZsGreen and MMNA. The microscope settings for these two fluorescent molecules and the fidelity of the Tet-On® Inducible System will first be determined by transfecting the cells with varying combinations of the vectors as outlined in Section 3.1.2. Another test using GFP-stable cells alongside tagging with MMNA will ensure that fluorescent resonance energy transfer (FRET) does not occur between ZsGreen and MMNA. Finally, expression of Alfred as detected by MMNA in fluorescence microscopy should show punctate fluorescent foci on the cell surface. The presence of these foci should appear only in Alfred-transfected cells, but not in non-transfected cells.

3. Materials and Methods

3.1. Materials

3.1.1. Plasmids

The lyophilized Alfred gene block was obtained from ThermoFisher Scientific (Ottawa, ON, Canada). The Tet-On® Inducible Systems plasmids (pCMV-TET3G, pTRE3G-ZsGreen1) were generously donated by Dr. Derrick Gibbings at the University of Ottawa (Ottawa, ON, Canada). BamHI-HF, MluI-HF, DpnI, NdeI, and ScaI and CutSmart buffer were obtained from New England BioLabs (Ipswich, MA, USA). Primers were obtained from ThermoFisher Scientific (Ottawa, ON, Canada). Q5 polymerase, 5× Q1 reaction buffer, NEBuilder® HiFi DNA Assembly Master Mix, and 1kb Plus DNA Ladder were obtained from New England BioLabs (Ipswich, MA, USA). dNTP kit, DreamTaq Green PCR Master Mix (2×), GeneRuler Express DNA Ladder, and 1Kb Plus DNA Ladder were obtained from ThermoFisher Scientific (Ottawa, ON, Canada). PerfectTaq DNA Polymerase was obtained from 5Prime. E.Z.N.A.® Plasmid Mini Kit I and E.Z.N.A.® Gel Extraction Kit were obtained from Omega Bio-Tek, Inc. (Norcross, GA, USA). pcDNA3.1 Neo(+) hSUMF1 (pCMV-FGE) and mCardinal-N1 were obtained from Addgene (Watertown, MA, USA). Agarose was obtained from Sigma-Aldrich (Oakville, ON, Canada). 10× TBE buffer and electrophoresis plastics were obtained from Bio-Rad Laboratories Inc (Mississauga, ON, Canada). SYBR® Safe DNA gel stain was obtained from ThermoFisher Scientific (Ottawa, ON, Canada).

3.1.2. Bacterial culture

Bacterial culture plastics were obtained from ThermoFisher Scientific (Ottawa, ON, Canada). Agar and LB Broth, Miller were obtained from ThermoFisher Scientific (Ottawa, ON, Canada). Ampicillin sodium salt and kanamycin sulfate were obtained from Sigma-Aldrich (Oakville, ON, Canada). Antibiotics are prepared in sterile-filtered 1000× aliquots and stored at -20°C. DH5α competent cells were obtained from ThermoFisher Scientific (Ottawa, ON, Canada) and XL1-blue competent cells were generously donated by The Boddy Lab at the University of Ottawa (Ottawa, ON, Canada).

3.1.3. Mammalian cell culture

Cell culture plastics were obtained from ThermoFisher Scientific (Ottawa, ON, Canada) or VWR International (Mississauga, ON, Canada). A549 cells were generously donated by Dr. Christina Addison at the Ottawa Hospital Research Institute (Ottawa, ON, Canada) and MDA-MB-231 cells and MDA-MB-231-BR cells were donated by Dr. Paula Foster at the Robarts Research Institute at Western University (London, ON, Canada). Phosphate-buffered saline (PBS) and Dulbecco's Modified Eagle Media (DMEM) were obtained from Sigma-Aldrich (Oakville, ON, Canada). DMEM, fetal bovine serum (FBS), L-glutamine (L-Gln), penicillin-streptomycin (penstrep), trypsin-EDTA, and DMSO were obtained from ThermoFisher Scientific (Ottawa, ON, Canada). $CaCl_2$•$7H_2O$, HEPES, NaCl, dextrose, Na_2PO_4•$7H_2O$, KCl, NaOH, and EDTA were obtained from Sigma-Aldrich (Oakville, ON, Canada). All chemicals are used as-is from the supplier. cOmplete™ Ultra Tablets, Mini, *EASYpack* Protease Inhibitor Cocktail was obtained from Roche (Basel, Switzerland). RIPA lysis buffer, 10×, β-mercaptoethanol (BME), and Ponceau S solution were obtained from Sigma-Aldrich (Oakville,

ON, Canada). Pierce™ BCA Protein Assay Kit and Tween®20 were obtained from ThermoFisher Scientific (Ottawa, ON, Canada). Mini-PROTEAN® TGX™ Precast Gels, Trans-Blot® Turbo™ Mini PVDF Transfer Packs, XT Sample Buffer, 4×, 2× Laemmli Sample Buffer, 10× Tris/Glycine/SDS Buffer, Goat Anti-Rabbit IgG (H+L) HRP Conjugate, Goat Anti-Mouse IgG (H+L) HRP Conjugate, ProteinPlus All Blue ladder, and Clarity™ Western ECL Substrate were obtained from Bio-Rad Laboratories Inc (Mississauga, ON, Canada). Primary anti-FGE antibody, anti-Myc antibody, and anti-tubulin antibody were obtained from Abcam (Cambridge, UK). Skim milk powder pasteurized was obtained from Loblaws (Toronto, ON, Canada). Geneticin (G418), hygromycin, puromycin, doxycyclin (dox), were obtained from ThermoFisher Scientific (Ottawa, ON, Canada). Whatman™ qualitative filter paper, Grade 1 was obtained from Sigma-Aldrich (Oakville, ON, Canada). Methyl 5-MeO-*N*-aminoanthranilate (MMNA) was synthesized by Dr. Mojmir Suchy according to our previous publication [14].

3.1.4. Common machinery and services

DNA concentration was determined using a BioSpectrometer® obtained from Eppendorf (Mississauga, ON, Canada). Agarose gel electrophoresis was imaged using an AlphaImager® EC obtained from Cell Biosciences (San Leandro, CA, USA). Plasmids were sequenced at the StemCore Laboratories at the Ottawa Hospital Research Institute (Ottawa, ON, Canada) or at Génome Québec (Montreal, QC, Canada). Protein quantification was conducted on a Synergy™ H4 Hybrid Multi-Mode Microplate Reader obtained from BioTek (Winooski, VT, USA). Western blot was conducted on a ChemiDoc XRS+ system obtained from Bio-Rad Laboratories Inc (Mississauga, ON, Canada). All live cell microscopy was conducted on a Zeiss Observer Z1,

fitted with LSM-TPMT, LSM 800, Heating Unit XL S, TempModule S, and CO_2 Module S. ZEN was the program used for imaging, and Image J was used for image processing.

3.2 Common methods

3.2.1. Bacterial culture materials

To make agar plates, 15 g/L agar and 25 g/L LB broth, Miller were fully dissolved in ddH_2O, then autoclaved. When the liquid had cooled, antibiotic was added. Near flame, the liquid was poured into petri dishes, and allowed to solidify. The plates were sealed with parafilm and stored upside-down at 4 °C. To make LB media, 25 g/L LB broth, Miller was dissolved in ddH_2O, then autoclaved. The broth was stored at room temperature.

3.2.2. Bacterial transformation and plasmid harvest

All bacterial culture protocols were performed near flame when the bacteria were exposed to air. Competent bacterial cells were thawed on ice, and then 50-100 ng of DNA was added directly into the tube and mixed gently. The cells were incubated on ice for 30 minutes, heat shocked at 42 °C for 45 seconds, and then immediately put on ice for 3 minutes. LB broth was added to reach 1 mL, and then gently mixed. The cells were incubated in a shaker at 200 rpm at 37 °C for 1 hour and then pelleted by centrifuging at 5000 ×g for 1 min. The media was removed until 100 µL of media was remaining, and the pellet was resuspended. This was spread on an agar plate with the appropriate antibiotic selection marker. After letting the plate dry for 3 minutes, the plate was placed in an incubator upside-down at 37 °C for 16-20 hours. The plate was then sealed with parafilm and stored upside-down at 4 °C for up to 2 months.

The appropriate antibiotic was used to grow selected colonies in LB media, and then the DNA was extracted with an E.Z.N.A.® Plasmid Mini Kit I (as per the standard protocol) and the concentration was verified with a BioSpectrometer®.

3.2.3. Creation of pTRE3G-ZsGreen1-Alfred

The pTRE3G-ZsGreen1 vector was cleaved with MluI-HF and BamHI-HF using CutSmart Buffer as per the protocol, but modified to use 2.5 µg of plasmid. The linearized vector was purified with an E.Z.N.A.® Gel Extraction Kit (as per the protocol). The sequence of Alfred was codon optimized for human cells by ThermoFisher Scientific, as requested. The lyophilized Alfred gene block was resuspended in the appropriate volume of ddH_2O to a stock concentration of 20 ng/µL, and vortexed thoroughly. The linearized vector and resuspended gene block were assembled as per the HiFi DNA Assembly protocol. The Gibson assembly has low transformation efficiency, and therefore 5 µL of the assembly was used in transformation. The DNA was extracted with an E.Z.N.A.® Plasmid Mini Kit I (as per the standard protocol) and the concentration was determined with a BioSpectrometer®.

The ligation was verified with agarose gel electrophoresis. The extracted DNA was cleaved with MluI-HF and BamHI-HF as per the standard protocol. A 1% agarose gel was made by mixing agarose in fresh TBE at 1 g/100 mL and melted in a microwave for 45 seconds in 15 second intervals. When cool (but not solidified), 1 µL/10 mL SYBR™ safe or ethidium bromide was added to the liquid. This was poured into a gel mould with a well comb and allowed to solidify. TBE was added as a running buffer. The PCR product was mixed with loading dye and then added into the wells; 5 µL ladder was used. The gel was run at 120-140 V

for ~40 minutes, and then imaged with an AlphaImager. After screening with the gel, the extracted DNA was sent to the Ottawa Hospital Research Institute for sequencing.

3.2.4. Creation of pCMV-Alfred

The pCMV backbone vector was obtained from pCMV-FGE using primers CA_FGE_F and CA_FGE_R with PCR as per the Q5 polymerase protocol. Likewise, the Alfred gene was obtained from pTRE3G-Alfred using primers CA_Alf_F and CA_Alf_R. The >20 bp overlap between the fragments was created with the primer design. The fragments were assembled as per the HiFi DNA Assembly protocol, and 5 µL of the assembly was used for transformation. The DNA was extracted with an E.Z.N.A.® Plasmid Mini Kit I (as per the standard protocol) and the concentration was determined with a BioSpectrometer®. The ligation was verified with agarose gel electrophoresis by amplifying the extracted DNA with PCR using primers CA_SEQ_F1 and CA_SEQ_R1 using PerfectTaq DNA Polymerase or DreamTaq Green PCR Master Mix (2×) as per the standard protocol. The DNA was then sent to Génome Québec for sequencing.

3.2.5. Mammalian cell culture

All mammalian cell culture protocols were performed in a biosafety cabinet (BSC). DMEM+ was prepared by supplementing DMEM with 10% v/v FBS, 1% v/v penstrep, and L-Gln (as needed). Cells were cultivated according to the guidelines provided by ThermoFisher Scientific [22] unless otherwise stated.

A cell culture was started by thawing a cell stock and adding 1×10^6 cells to a T-75 flask. The media was replaced the next day to remove the cryoprotectant in the frozen culture. The cells were allowed to grow to ~80% confluency (generally every 3-4 days) and then were

passaged. To passage the cells, they were lifted with enough trypsin-EDTA to cover the bottom of the flask, and incubated at 37 °C until fully lifted. The trypsin was neutralized with an equal volume of DMEM+. The cells were transferred into a Falcon tube and centrifuged at 218 ×g at 5 °C for 5 minutes to pellet the cells. The trypsin-DMEM+ was aspirated, and then the pellet was resuspended in DMEM+. Cells were plated at the suggested plating confluency (usually a 1/4-1/10 dilution).

3.2.6. Transient transfection of mammalian cells

In a 1.5 mL microcentrifuge tube, 0.061 µL/µL (final volume) $CaCl_2$ solution and 0.05-0.1 µg/µL (final volume) of each DNA plasmid (or at the desired ratio) were mixed. One volume of 2× HBS was added dropwise, and then the mixture was incubated at room temperature for 10 minutes (the mixture became cloudy). The transfection mixture was added to wells of a 6-, 12-, 24-, or 96-well plate containing cells, at 1:10 mixture:media (the transfection mixture was added to the media in the wells). The cells were incubated at 37 °C for 4-6 hours, washed 1-2 times with pre-warmed PBS, and then the appropriate volume (2 mL, 1 mL, 1 mL, or 200 µL, respectively) of pre-warmed media was added.

3.2.7. Creation of stably transfected cell lines

The optimal antibiotic concentration was determined for selection. Wild type cells were plated in a 12-well plate, and then a range of concentrations of antibiotics diluted in DMEM+ was added in triplicate. The cells were monitored daily for cell death, and the media was replaced when cell death was high, as well as prior to imaging. The antibiotic concentration that killed 100% of the wild type cells in 3-5 days was chosen as the concentration used for selection.

Cells were plated to be confluent after 48 hours. Cells were lifted 24 hours after transient transfection (with pCMV-TET3G to create A549-TET3G or pCMV-FGE to create A549-FGE) and then distributed among 3-4 10×10 cm dishes in selection media (DMEM+ with antibiotic). Cells were well dispersed to allow individual colony formation without passaging. Every 4-5 days (or more frequently if the cell death was high), media was changed by aspirating the old media and adding fresh DMEM+ and antibiotic. Colonies were formed after 2-3 weeks, until discernable without a microscope. Small isolated colonies were selected by dipping a small piece of autoclaved Whatman™ filter paper in trypsin-EDTA and placing it on the colony. The petri dish was closed, and after 5 minutes, the paper was lifted and shaken gently in media to dislodge the cells into wells of a 24-well plate containing DMEM+ and antibiotic. Optionally, the paper was left for a few minutes before removing it from the media. As many colonies as possible were selected (at least 24 colonies in total). The cells were allowed to grow into a large patch without cells overlapping, and then moved to 2-3 wells of a 6-well plate (at least 6 colonies). When cells had reached sufficient confluency, the wells were pooled into a T-75 flask, 1 for each colony, and grown in a maintenance antibiotic concentration of 0.2 μg/mL. Cells were then frozen into stocks and tested for expression of the transfected plasmid.

3.2.8. Harvesting of protein from cell lysates

A549-FGE cells were grown in a flask, and A549 cells were plated in a 6-well plate and transfected with pCMV-Alfred. The cells were lifted with trypsin-EDTA (for FGE expression in A549-FGE cells) or EDTA-PBS (for Alfred expression in A549 cells transiently transfected with pCMV-Alfred) and pelleted by centrifuging. The trypsin-DMEM+ was aspirated, and the pellet redissolved in PBS to wash. The cells were pelleted again and the PBS was aspirated. On ice,

1 tablet of cOmplete™ Ultra Tablets, Mini, *EASYpack* Protease Inhibitor Cocktail was dissolved in 1.5 mL autoclave ddH_2O, and then 142 µL of the protease inhibitor was added to 855 µL of cold RIPA buffer (diluted to 1×). The cell pellet was dissolved in 200 µL of the cold RIPA buffer-protease inhibitor solution. The cells were vortexed, incubated on ice for 20 minutes, and then centrifuged at 16000 ×g at 4 °C for 20 minutes. The cell lysate (supernatant) was moved to a new microcentrifuge tube and then stored at -80 °C.

Protein content in the cell lysates was quantified using the Pierce™ BCA Protein Assay Kit. A standard curve was generated by diluting the 2000 µg/mL bovine serum albumin (BSA) stock in RIPA buffer to create a concentration range from 31.25 µg/mL to 2000 µg/mL, increasing by a factor of 2 for each concentration. A 1/20 dilution of the cell lysate was prepared. BCA working reagent was prepared by mixing 50:1 solution A:solution B. The cell lysate or BSA standards were added to each well of a 96-well plate. BCA working reagent was added to each well containing protein at a 1:19 cell lysate:BCA working reagent ratio, and added to an extra well for the 0 µg/mL standard. The plate was incubated for 30 min at 37 °C in the dark, and then the absorbance was read at 37 °C at 562 nm. The concentration of protein in the cell lysate was quantified with the standard curve.

3.2.9. SDS-page and Western blot

Cell lysates were diluted to 3 µg/µL with PBS. Loading buffer was prepared at 2× with XT Sample Buffer, 4×, with 10% BME (for final 5% BME). The loading buffer was then added to the cell lysate dilution, and heated at 95-99 °C for 10 minutes. The liquid was collected by a brief centrifugation. The Mini-PROTEAN® TGX™ Precast Gel (4-20% polyacrylamide in a top-bottom gradient) was prepared as per the standard protocol, and 20 µL of sample was loaded

into each well, and 5 µL of ProteinPlus All Blue ladder was used. The gel was run as per the standard protocol. 1× TBS-T was prepared by mixing 10× TBS and 1% Tween 20 in ddH_2O. Blocking buffer, or 5% milk, was prepared by dissolving 5 g/100 mL skim milk powder pasteurized in TBS-T. When the bands had migrated almost to the bottom of the gel, the precast gel was cracked open and the proteins were transferred from the gel to a Trans-Blot® Turbo™ Mini PVDF Transfer Packs as per the standard protocol. The PVDF membrane was transferred to a Falcon tube (protein side towards the inside). The blocking buffer was added to the PVDF membrane, and incubated at room temperature for 1 hour on a roller. Primary antibody solutions were prepared by diluting in 5% milk + 0.01% azide: anti-FGE antibody was diluted 1:500; anti-Myc antibody was diluted 1:1000; anti-tubulin antibody was diluted 1:5000 in TBS-T. Secondary antibody solution was prepared by diluting in TBS-T and prepared fresh each time: anti-rabbit antibody was diluted 1:10000; anti-mouse antibody was diluted 1:10000. The blocking buffer was drained, and the primary antibody was added. This was left to incubate overnight at 4 °C on a roller. Then, the primary antibody (anti-FGE or anti-Myc) was removed, and the PVDF membrane was washed 3 times with TBS-T for at least 10 minutes each on a roller. The secondary antibody was added and incubated at room temperature for 1 hour on a roller, and then removed. The PVDF membrane was washed again prior to imaging on a ChemiDoc XRS+, using the Clarity™ Western ECL Substrate kit as per the standard protocol. The membrane was then washed with TBS-T, but not stripped. Antibody binding was also performed with anti-tubulin antibody in the same manner, with the following changes: the primary antibody was incubated at room temperature for 1 hour; the secondary antibody was incubated at room temperature for 20 minutes. Optionally, the PVDF membrane was verified for

presence of proteins, by covering the membrane with Ponceau S for 5 minutes. The Ponceau S was washed off with distilled H_2O.

3.2.10. Cell imaging

The Zeiss microscope included an incubation chamber, which was set to 37 °C and held at 5% CO_2. Cells were imaged at 10× in a Nunc™ 24-Well Plate, Round, or at 20× in a Nunc™ MicroWell™ 96-Well Optical-Bottom Plates with Polymer Base. The DMEM+ was aspirated and replaced with DMEM+ (no phenol red). For imaging aldehydes, 10 mM MMNA was prepared in DMSO, and used at a final concentration of 0.01 mM (from first microscopy experiment until Fig 4.2.3) or 0.02 mM (from Figure 4.2.4 onwards) per well. When necessary, the probe was further diluted with DMEM+, immediately before adding to wells. No more than 1 µg/mL DMSO was present in the final volume. The cells were allowed to incubate 20 minutes for maximum fluorescence. All pictures were taken within 1 hour after adding the probe, prior to cyclization of the MMNA.

3.2.11. Prediction of protein secondary structure

The amino acid sequence of Wu et al.'s construct [7] and Alfred were inputted into Chou & Fasman Secondary Structure Prediction Server (CFSSP) and the improved self-optimized prediction method (SOPMA), available online (CFSSP: http://www.biogem.org/tool/chou-fasman/index.php; SOPMA: https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html). Parameters were left as the default.

3.3. Protocols for imaging

Separate channels for each fluorescent source (ZsGreen, MMNA, and mCardinal) were created, and the emission wavelengths were adjusted to eliminate bleedthrough of signal between channels. Images were taken on the required channels and then combined to create an overlay image. Each colour was adjusted linearly; the same colour adjustments were applied across all images in a figure, except for the brightfield images which were adjusted individually for each picture. Colour adjustments varied between figures.

3.3.1. Establishing controls and imaging parameters

Wild type A549 cells were plated in a 24-well plate at 3×10^4 cells/well. After 24 hours, cells were transfected with pCMV-TET3G and pTRE3G-ZsGreen1 (1:4), or only pTRE3G-ZsGreen1. After 24 hours, fresh sterile-filtered dox was prepared, and 1000 ng of dox was added to select wells to induce expression. After 24 hours, MMNA was added and the cells were imaged for ZsGreen (λ_{ex} = 488 nm, emission window set to 499-540 nm) and MMNA (λ_{ex} = 405 nm, emission window set to 410-485 nm).

3.3.2. Verifying absence of FRET

MDA-MB-231 and MDA-MB-231-BR (containing GFP) cells were plated. After 24 hours, MMNA was added and the cells were imaged for GFP (λ_{ex} = 488 nm, emission window set to 499-540 nm) and MMNA (λ_{ex} = 405 nm, emission window set to 410-495 nm).

3.3.3. Alfred expression

Wild type A549 cells were plated in a 24-well plate at 3×10^4 cells/well. After 24 hours, cells were co-transfected with pTRE3G-ZsGreen1-Alfred (1:4). After 24 hours, fresh sterile-filtered dox was prepared, and 1000 ng of dox was added to select wells to induce expression. After 24 hours, MMNA was added and the cells were imaged for ZsGreen (λ_{ex} = 488 nm, emission window set to 499-540 nm) and MMNA (λ_{ex} = 405 nm, emission window set to 410-495 nm).

3.3.4. Alfred expression in the A549-TET3G cell line

A cell line for stable TET3G transactivator expression (A549-TET3G) was created using pCMV-TET3G and selected with G418. Expression was tested with pTRE3G-ZsGreen1 transfection and cells imaged for ZsGreen (λ_{ex} = 488 nm, emission window set to 495-540 nm).

Wild type A549 or A549-TETG cells were plated in a 24-well plate at 3×10^4 cells/well. After 24 hours, cells were transfected with pTRE3G-ZsGreen1-Alfred. After 24 hours, fresh sterile-filtered dox was prepared, and 1000 ng of dox was added to select wells to induce expression. After 24 hours, MMNA was added and the cells were imaged for ZsGreen (λ_{ex} = 488 nm, emission window set to 499-540 nm) and MMNA (λ_{ex} = 405 nm, emission window set to 410-495 nm).

3.3.5. Alfred expression in the A549-TET3G cell line with overexpression of FGE

Wild type A549 or A549-TETG cells were plated in a 24-well plate at 3×10^4 cells/well. After 24 hours, cells were co-transfected with pCMV-FGE and pTRE3G-ZsGreen1-Alfred (1:1). After 24 hours, fresh sterile-filtered dox was prepared, and 1000 ng of dox was added to wells to

induce expression. After 24 hours, MMNA was added and the cells were imaged for ZsGreen (λ_{ex} = 488 nm, emission window set to 499-540 nm) and MMNA (λ_{ex} = 405 nm, emission window set to 410-495 nm).

This procedure was repeated in a time-course experiment. Cells were plated and imaged at the 24, 48, and 72 hour timepoints after transfection.

3.3.6. Constitutive Alfred expression

A cell line for stable FGE expression (A549-FGE) was created with pCMV-FGE and selected with G418. Transfection was verified with Western blot.

Wild type A549 or A549-FGE cells were plated in a 24-well plate at 3×10^4 cells/well. After 24 hours, cells were co-transfected with pCMV-Alfred and mCardinal-N1 (1:1). Cells were lifted with EDTA-PBS from the 24-well plate and moved to a glass-bottom 96-well plate for imaging. MMNA was added and the cells were imaged for mCardinal (λ_{ex} = 633 nm, emission window set to 640-754 nm) and Alfred (λ_{ex} = 405 nm, emission window set to 410-495 nm).

This procedure was repeated in a time-course experiment. Cells were plated and imaged at the 24, 48, and 72 hour timepoints after transfection.

4. Results

4.1. Alfred expression with the Tet-On® Inducible System

This experiment aimed to express Alfred, a transmembrane (TM) protein displaying an aldehyde at the cell surface, in mammalian cells. A549, a lung cancer cell line, was chosen to express this system, as it is a cancer-initiating cell line. This will allow creation of tumours in living systems, which paves the way for future projects. The Alfred gene was ligated into pTRE3G-ZsGreen1 and induced by the transactivator protein TET3G, produced by pCMV-TET3G in the Tet-On® Inducible System, in the presence of doxycyclin (dox). The Alfred-derived aldehyde should react with the aldehyde probe developed previously by the Molecular Medicine Lab, methyl 5-MeO-*N*-anthranilate (MMNA), and present an increased amount of fluorescence on the surface of the cells, or the perimeter of the cell shape when viewing a cross section. In particular, Alfred-induced fluorescence is expected to co-localize with cells that express ZsGreen, the positive marker of Tet-On induction.

4.1.1. Creation of pTRE3G-ZsGreen1-Alfred

The lyophilized Alfred gene was resuspended, and then ligated into pTRE3G-ZsGreen1 after the IRES into the multiple cloning site (MCS) with Gibson assembly. The product was amplified in *E. coli* and screened with restriction enzyme digestion to verify ligation (Figure 4.1.1.).

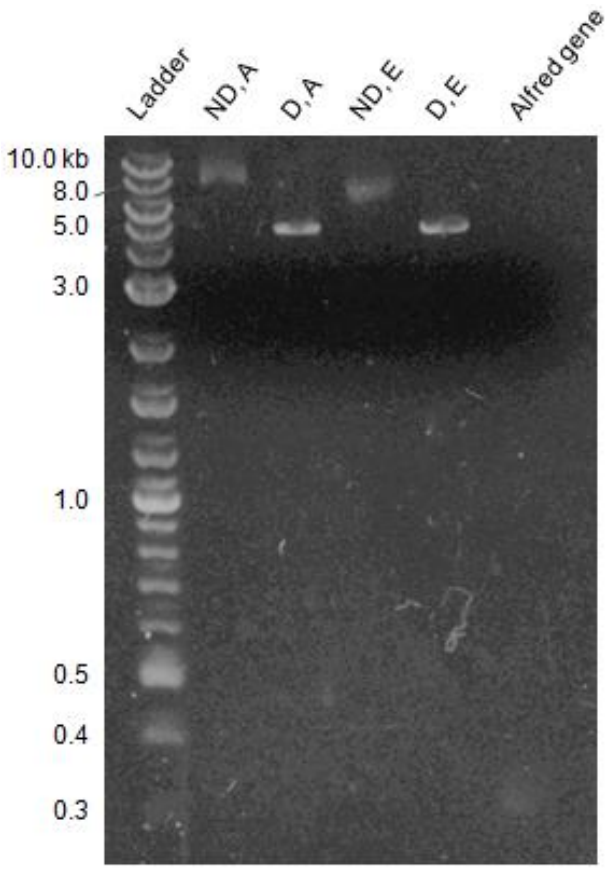

Figure 4.1.1. Restriction enzyme digest screening of pTRE3G-ZsGreen1-Alfred. Plasmid DNA was extracted from transformed competent XL1-blue cells and digested with MluI-HF and BamHI-HF before screening with electrophoresis on a 1% agarose gel. Non-digested pTRE3G-ZsGreen1-Alfred was 4967 bp; non-digested pTRE3G-ZsGreen1 was 4700 bp; the Alfred gene was 267 bp. Ladder: 1 Kb Plus DNA ladder; ND: non-digested; D: restriction enzyme digested with MluI-HF and BamHI-HF; A: pTRE3G-ZsGreen1-Alfred; E: pTRE3G-ZsGreen1.

From the agarose gel electrophoresis (Figure 4.1.1.), it can be seen that the non-digested pTRE3G-ZsGreen1-Alfred and pTRE3G-ZsGreen1, which are circular, had migrated above their linear lengths. The digested pTRE3G-ZsGreen1 was located at ~4700 bp and the Alfred gene block was located at ~300 bp. In the digested pTRE3G-ZsGreen-Alfred sample, the original pTRE3G-ZsGreen1 reappeared at ~4700 bp, but the Alfred gene did not reappear at ~300 bp. This does not confirm a successful ligation, but since the non-digested pTRE3G-ZsGreen1

migrated further than the pTRE3G-ZsGreen1-Alfred, the vector was sent for DNA sequencing, which confirmed successful ligation (Appendix A).

4.1.2. Establishing controls and imaging parameters for the Tet-On® Inducible System in A549 cells

A549 cells were transiently transfected with various combinations of the Tet-On® Inducible Systems plasmids to establish the positive and negative controls (Table 4.1.1.). These controls were used to test the fidelity of the Tet-On® Inducible System, to establish controls for subsequent experiments, and to determine the emission wavelengths to be detected for each fluorescence source.

Table 4.1.1. Positive and negative controls established for the Tet-On® Inducible System.

Control name	**Purpose**
+M +E +D	Positive control to show that TET3G (from pCMV-TET3G) induces expression of the ZsGreen (from pTRE3G-ZsGreen1) in the presence of dox.
+M +E	Negative control to show that TET3G does not induce expression of the ZsGreen without dox; pTRE3G-ZsGreen1 is not inducible endogenously.
+A +D	Negative control to show that Alfred (from pTRE3G-ZsGreen1-Alfred) is not inducible by dox without TET3G.
+A	Negative control to show that Alfred (from pTRE3G-ZsGreen1-Alfred) is not inducible endogenously.
+D	Negative control to show dox does not affect regular cell growth.
Blank	Negative control to show regular growth of A549 cells.

Twenty-four hours post-transfection, the cells were imaged using fluorescence microscopy to detect ZsGreen, coloured green, and then incubated with MMNA for 20 minutes prior to imaging for MMNA fluorescence, coloured blue (Figure 4.1.2.). The cells containing pCMV-TET3G and pTRE3G-ZsGreen1 and induced by dox produced fluorescence from ZsGreen. Cells not induced by dox did not fluoresce green, nor did cells that did not contain pCMV-TET3G. Regardless of the treatment, all the cells fluoresced due to tagging by MMNA. Notably, some cells displayed brighter MMNA fluorescence, but this occurred in non-green cells as well as green cells. The pTRE3G-ZsGreen1-Alfred-transfected and dox-induced treatment, the dox-induced treatment, and non-transfected treatment also showed dimmer overall MMNA fluorescence compared to the other treatments.

4.1.3. Verification of absence of FRET

A test for FRET was conducted using MDA-MB-231 cells and MDA-MB-231-BR cells. The latter cell line is derived from the former, and expresses GFP. Since the excitation and emission curves of GFP are similar to ZsGreen, it was used as a control to test for absence of FRET between ZsGreen and MMNA. The same microscope settings used for ZsGreen were used for imaging GFP. Cells were plated, and then 24 hours later, incubated for 20 minutes with MMNA prior to imaging (Figure 4.1.3.). MDA-MB-231-BR cells that were green showed no increase in blue fluorescence compared to cells that were not green. MDA-MB-231 and MDA-MB-231-BR cells had no difference in blue fluorescence. Together, this suggested that FRET was not present, because the presence of green fluorescence did not appear to cause an increase in blue fluorescence.

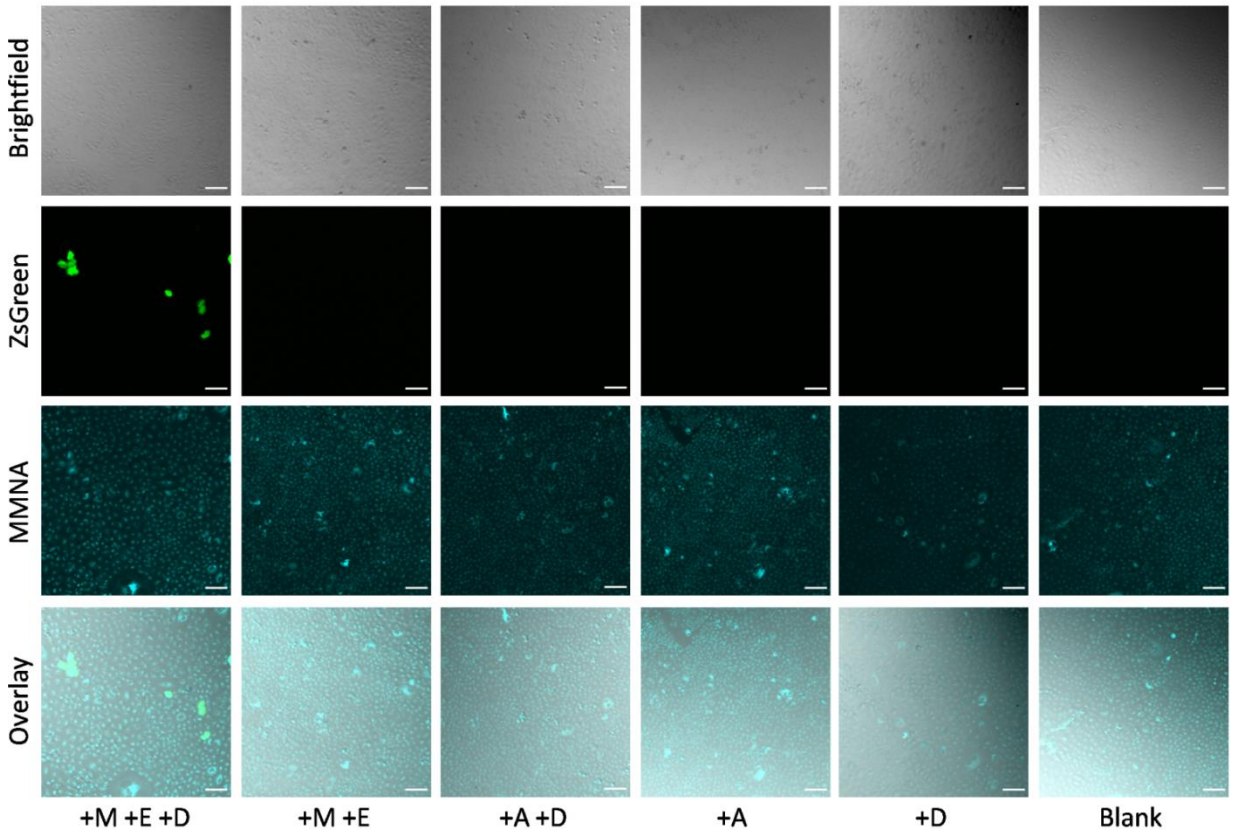

Figure 4.1.2. Controls testing functionality of the Tet-On® Inducible System. Plasmids (M: pCMV-TET3G; A: pTRE3G-ZsGreen1-Alfred; E: pTRE3G-ZsGreen1) were transfected into A549 cells, and then induced with doxycycline (D) for 24 hours. They were incubated with methyl 5-MeO-*N*-anthranilate (MMNA) for 20 minutes, and imaged at 10× for ZsGreen and MMNA. The brightfield image was taken separately, and included in the overlay with image processing. Scale bar = 100 µm.

4.1.4. Alfred expression in the Tet-On® Inducible System

Alfred expression was tested in A549 cells using the verified Tet-On® Inducible System. pCMV-TET3G and pTRE3G-ZsGreen1-Alfred were transiently co-transfected into cells. After 24 hours, the cells were incubated with dox for 24 hours. Then, the cells were incubated with MMNA for 20 minutes prior to imaging (Figure 4.1.4.). In the pCMV-TET3G and pTRE3G-ZsGreen1-Alfred transfected and dox-induced cells, the few cells showing green

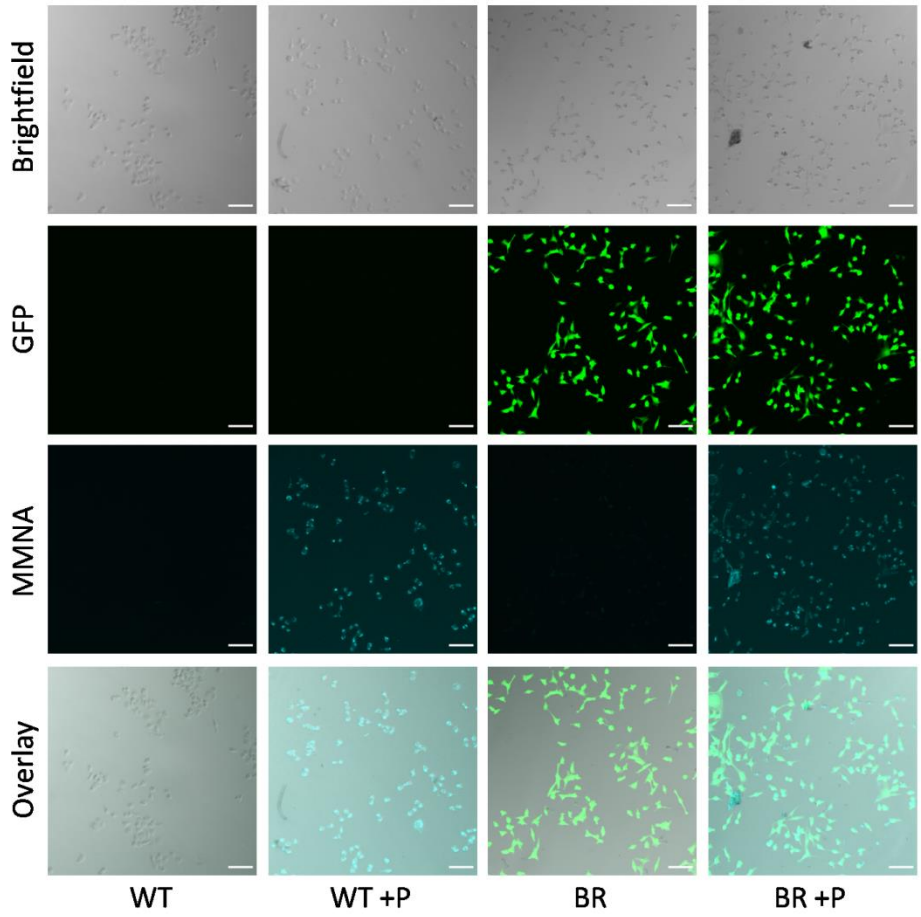

Figure 4.1.3. Fluorescent resonance emission transfer (FRET) of GFP and methyl 5-MeO-*N*-anthranilate (MMNA). MDA-MB-231 (wild type) and MDA-MB-231-BR (containing GFP) cells were imaged. They were incubated with MMNA for 20 minutes, and imaged at 10× for ZsGreen and MMNA, using the same settings for previous imaging. Scale bar = 100 µm.

fluorescence did not all show brighter MMNA fluorescence. There were also cells showing brighter blue fluorescence that did not fluoresce green. It should be noted that dead or dying cells (circular in shape) tend to exhibit bright MMNA fluorescence; these cells are disregarded in the analysis. In other words, aldehyde detection did not co-localize with ZsGreen signal in its location and its intensity among the cells, suggesting Alfred is not expressed or is not detected

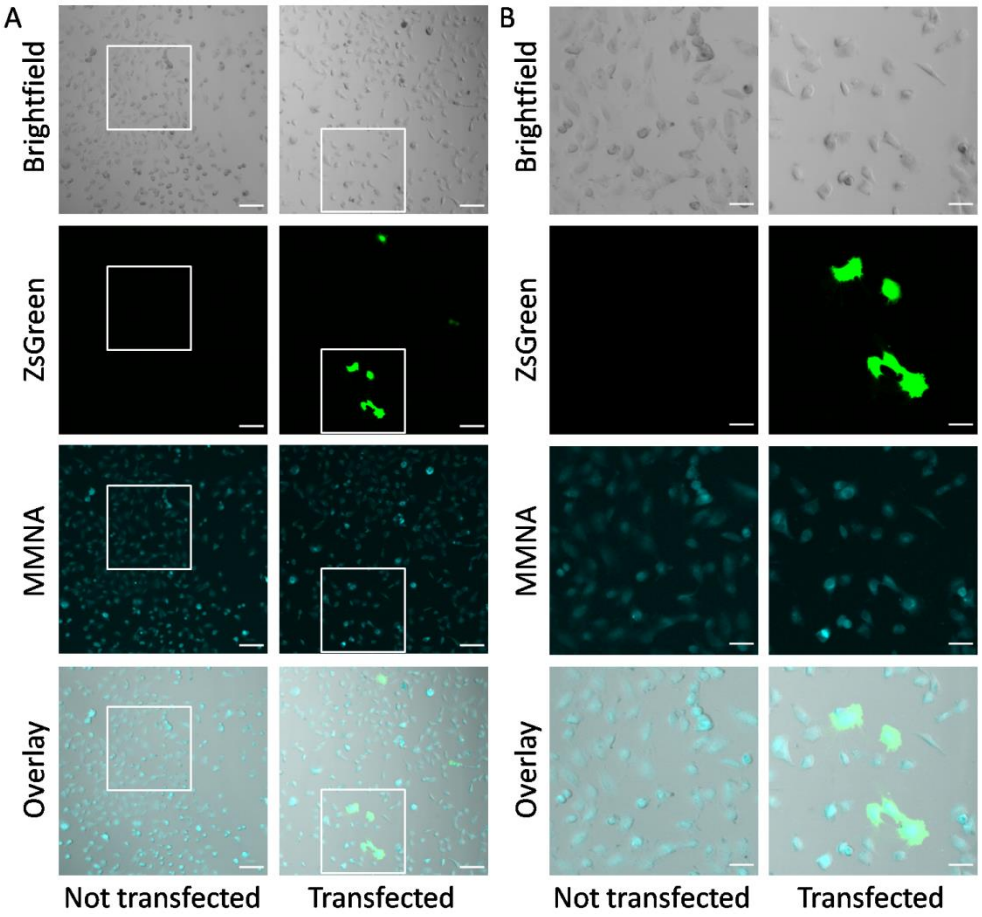

Figure 4.1.4. Alfred expression in the Tet-On® Inducible System. Plasmids (M: pCMV-TET3G; A: pTRE3G-ZsGreen1-Alfred) were transfected into A549 cells. They were induced with doxycycline for 24 hours, incubated with methyl 5-MeO-*N*-anthranilate (MMNA) for 20 minutes, and imaged for ZsGreen and MMNA. **A.** Image taken at 10×. Scale bar = 100 µm. **B.** 2.5× magnification of the area indicated in Panel A. Scale bar = 40 µm.

with the experimental conditions tested. In the magnified images (Figure 4.1.4.B), it can be seen that MMNA is dispersed across the cell, and in some cells, is perinuclear. This perinuclear localization is present in both Alfred-transfected cells and non-transfected cells, and has been observed as normal aldehyde distribution in previous studies.

4.2. Alfred expression with the Tet-On® Inducible System with overexpression of FGE

Since the expression of Alfred did not show an increase in blue fluorescence, it was postulated that Alfred is expressed, but proteins that were localized to the cell surface did not express the aldehyde. Thus, FGE was overexpressed to increase the thiol-to-aldehyde conversion.

4.2.1. Generation of a cell line stably expressing the TET3G transactivator protein

A kill curve was conducted on A549 cells using geneticin (G418). Cells were plated, and then a range of concentrations of G418 was added to assess the concentration at which 100% of the cells would die within 3-5 days. After 5 days of incubation in media containing G418, near 100% cell death occurred in cells incubated with 1100 µg/mL of G418 (Figure 4.2.1.). This was indicated by the lack of adherent live cells after washing away the lifted dead cells. There were still a few live cells present; however, this may be attributed to the confluency in the well, which will not be an issue during the selection process. Thus, 1100 µg/mL was the concentration used to create A549-TET3G, a stable A549 cell line containing pCMV-TET3G using G418 resistance as the selection marker (Figure 4.2.2.). Though few, there were cells present that exhibited green fluorescence. As transfection efficiency has previously been low, this was deemed sufficient for subsequent experiments.

4.2.2. Alfred expression in the A549-TET3G cell line

pTRE3G-ZsGreen1-Alfred was transiently transfected into A549-TET3G cells. Cells were transfected, and then imaged after 24 hours. After imaging the cells, they were incubated

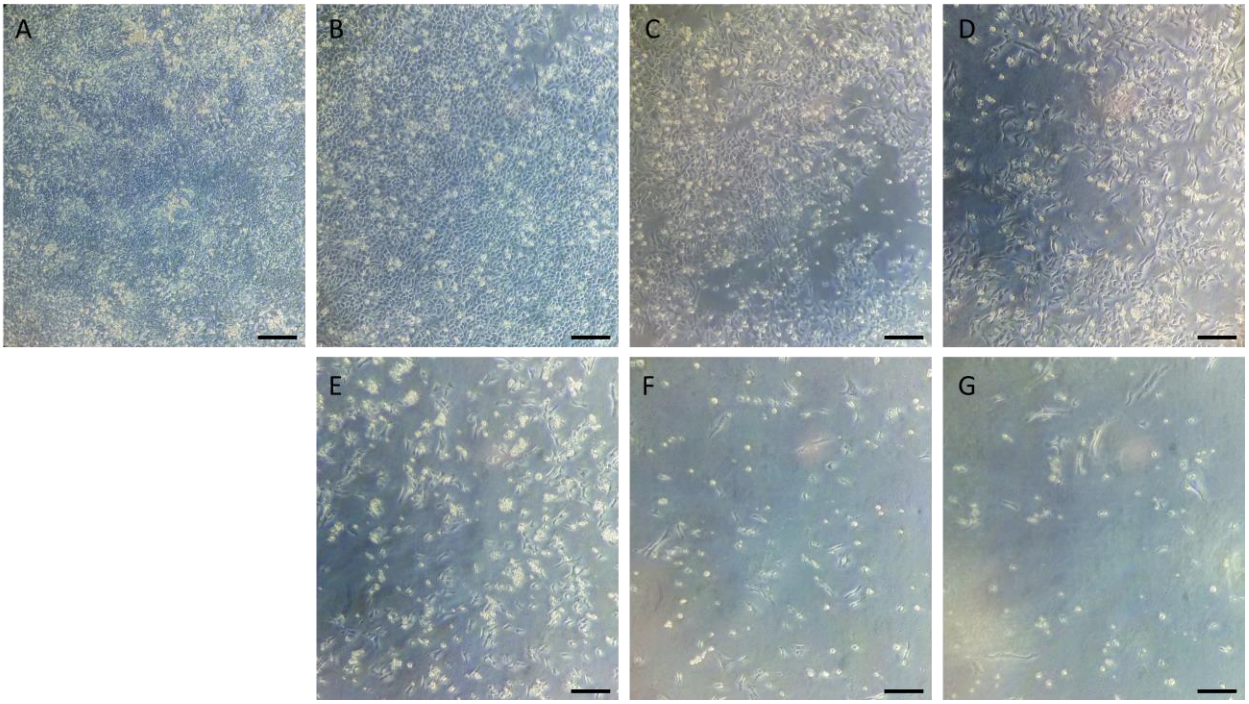

Figure 4.2.1. Determination of G418 concentration for selection of the A549 cell line. A549 cells were plated at 10000 cells/well of a 12-well plate, in triplicate. The cells were incubated in DMEM+ with varying concentrations of G418: **A.** 0 µg/mL, and increasing concentrations of 100 µg/mL, from **B.** 700 µg/mL to **G.** 1200 µg/mL. Cells were imaged 5 days after initially adding the antibiotic. Scale bar = 200 µm.

with MMNA and imaged again (Figure 4.2.3.). pTRE3G-ZsGreen1-Alfred-transfected cells can be identified by their green fluorescence. The faint green fluorescence did not always overlap with brighter blue fluorescence after MMNA was added, and MMNA localization is perinuclear within some cells, similarly to previous microscopy experiments. However, a new finding is observed in the magnified areas (Figure 4.2.3.B). Some cells show punctate fluorescent foci from MMNA within the bounds of the cell. However, these foci are not necessarily present in cells fluorescing green. This indicates that the bright foci may not be indicative of Alfred localization. Regardless, presence of Alfred did not result in increased aldehydes on the cell surface.

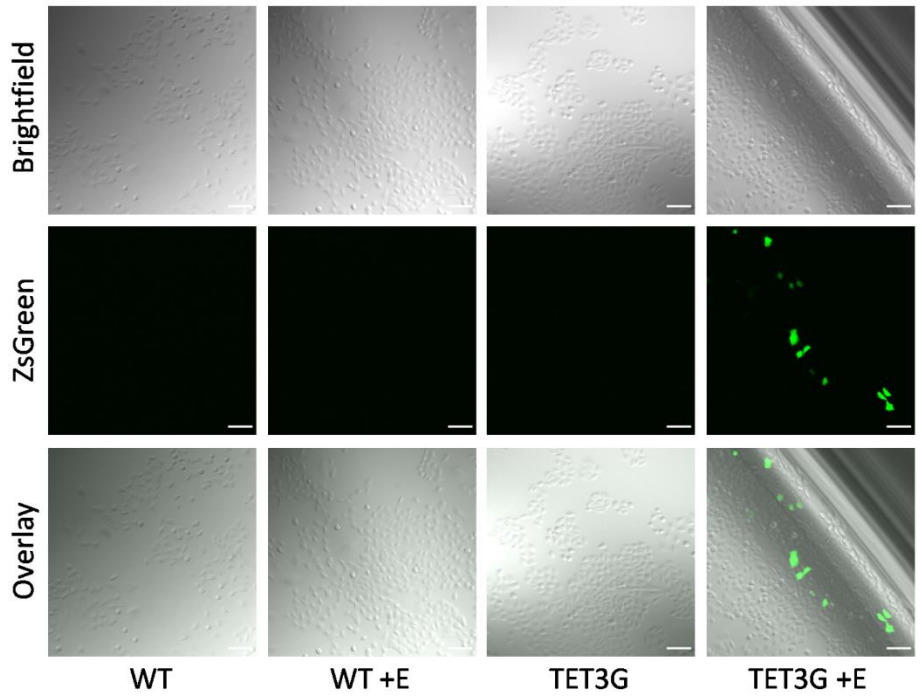

Figure 4.2.2. Verification of stable transfection of pCMV-TET3G into the A549 cell line. pTRE3G-ZsGreen1 (E) was stably transfected into wild type A549 cells (WT) or A549-TET3G cells (TET3G) using G418 selection. They were induced with dox for 24 hours, and then imaged at 10× for ZsGreen. Scale bar = 100 µm.

4.2.3. Alfred expression in the A549-TET3G cell line with overexpression of FGE

pTRE3G-ZsGreen1-Alfred and pCMV-FGE were transiently co-transfected into A549-TET3G cells. Since introduction of Alfred alone was insufficient to see an increase in aldehydes on the cell surface, FGE was overexpressed in order to increase production of cell surface aldehydes. Cells were imaged 24 hours post-transfection, and then imaged again after incubating with MMNA for 20 minutes (Figure 4.2.5.). Microscopy results were consistent with previous observations: diffuse MMNA fluorescence was shown in green-fluorescing cells, cells exhibited MMNA fluorescence in the perinuclear region of the cell (Figure 4.1.4.B), and punctate fluorescent foci were present in cells that did not fluoresce green (Figure 4.2.3.).

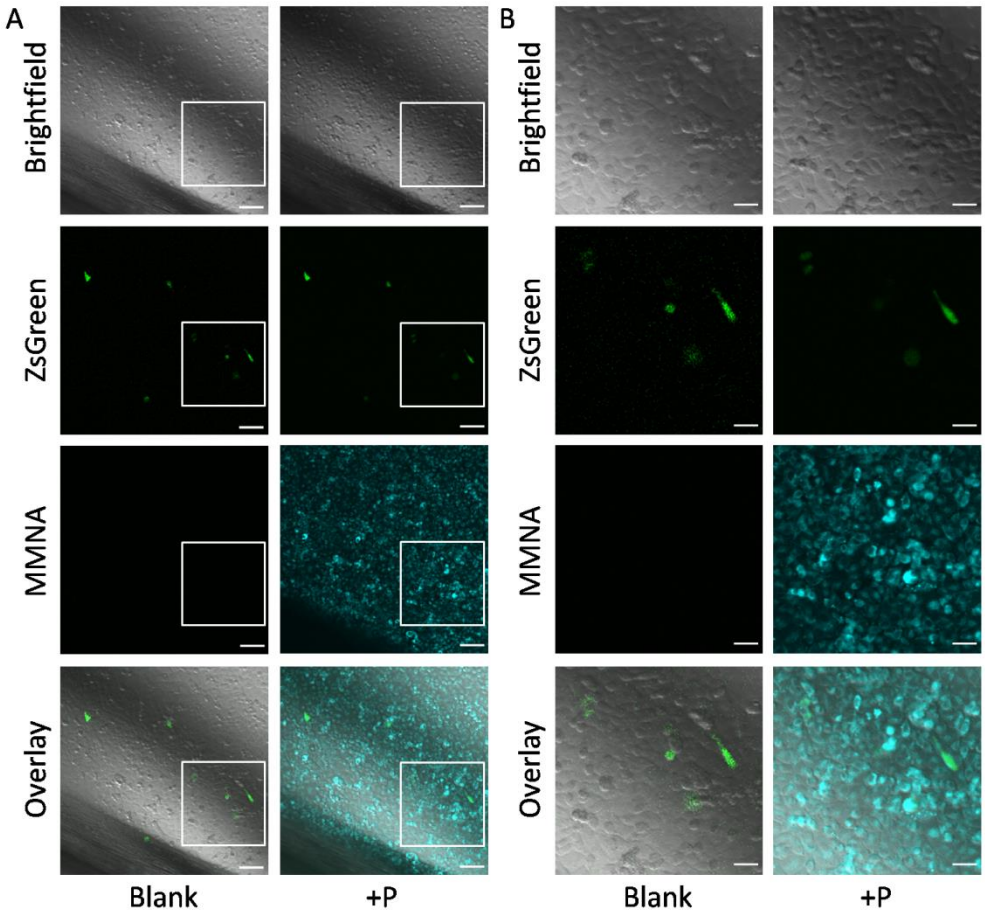

Figure 4.2.3. Transient Alfred expression in the Tet-On® Inducible System using the A549-TET3G cell line. pTRE3G-ZsGreen1-Alfred was transiently transfected into A549-TET3G cells. They were induced with doxycycline for 24 hours, incubated with methyl 5-MeO-*N*-anthranilate (MMNA) for 20 minutes, and imaged for ZsGreen and MMNA. **A.** Image taken at 10×. Scale bar = 100 µm. **B.** 2.5× magnification of the area indicated in Panel A. Scale bar = 40 µm.

Since an increased thiol-to-aldehyde conversion did not yield different results, it was again hypothesized that Alfred had not yet translocated to the cell surface due to the perinuclear localization of MMNA observed in Figure 4.1.4.B. Therefore, a time-course experiment was conducted, using the same procedure, to observe any changes after a longer period of induction.

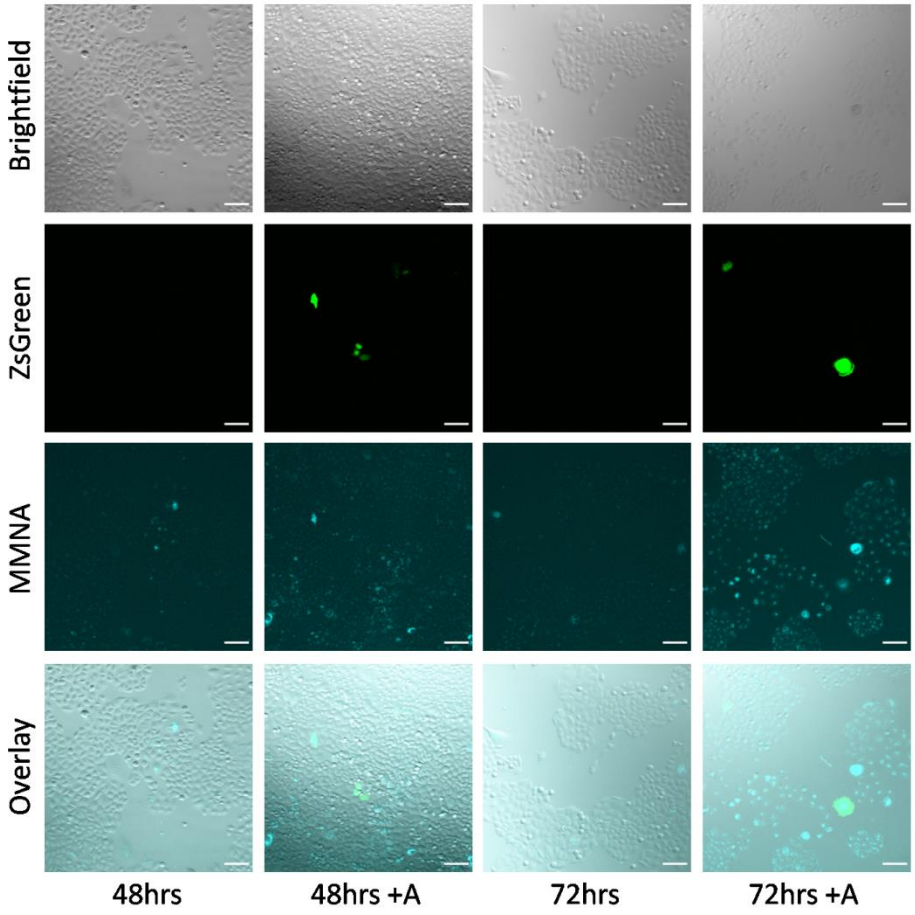

Figure 4.2.4. Time-course of transient Alfred expression in the Tet-On® Inducible System using the A549-TET3G cell line. pTRE3G-ZsGreen1-Alfred (A) was transiently transfected into A549-TET3G cells. They were induced with doxycycline for 48 or 72 hours, incubated with methyl 5-MeO-*N*-anthranilate (MMNA) for 20 minutes, and imaged at 10× for ZsGreen and MMNA. Scale bar = 100 µm.

Cells were imaged 48 or 72 hours post-induction with dox, and then imaged again after incubating with MMNA for 20 minutes (Figure 4.2.6.). The cells continued to exhibit similar patterns of MMNA fluorescence as seen in cells imaged 24 hours post-transfection (Figure 4.2.5.).

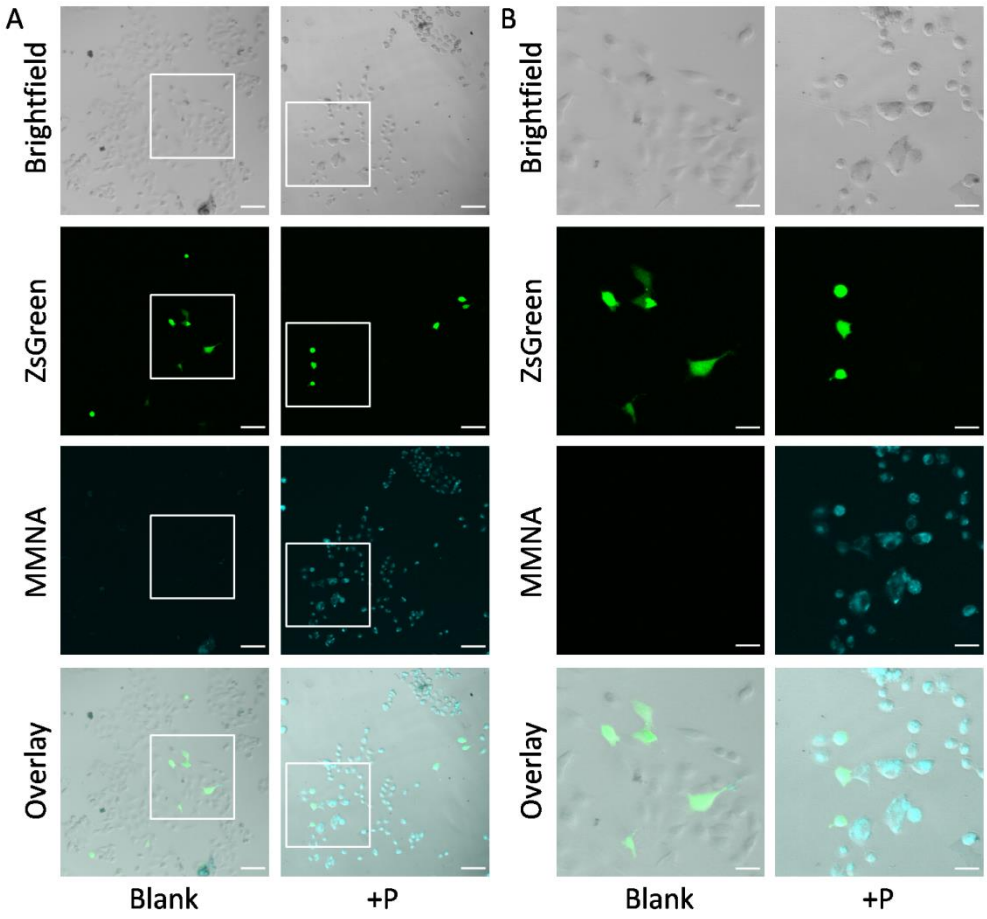

Figure 4.2.5. Transient Alfred expression in the Tet-On® Inducible System using the A549-TET3G cell line and overexpression of FGE. pTRE3G-ZsGreen1-Alfred and pCMV-FGE were transiently transfected into A549-TET3G cells. They were induced with doxycycline for 24 hours, incubated with methyl 5-MeO-*N*-anthranilate (MMNA) for 20 minutes, and imaged for ZsGreen and MMNA. A. Image taken at 10×. Scale bar = 100 µm. B. 2.5× magnification of the area indicated in Panel A. Scale bar = 40 µm.

4.3. Constitutive Alfred expression

Although the Tet-On® Inducible System controls showed that the system functioned as described, it was speculated that translation through the IRES may decrease Alfred expression to the point where its presence is unnoticeable. To eliminate this and other potential sources of problems, Alfred was introduced into the cells to be constitutively expressed instead of expression through dox-induced means.

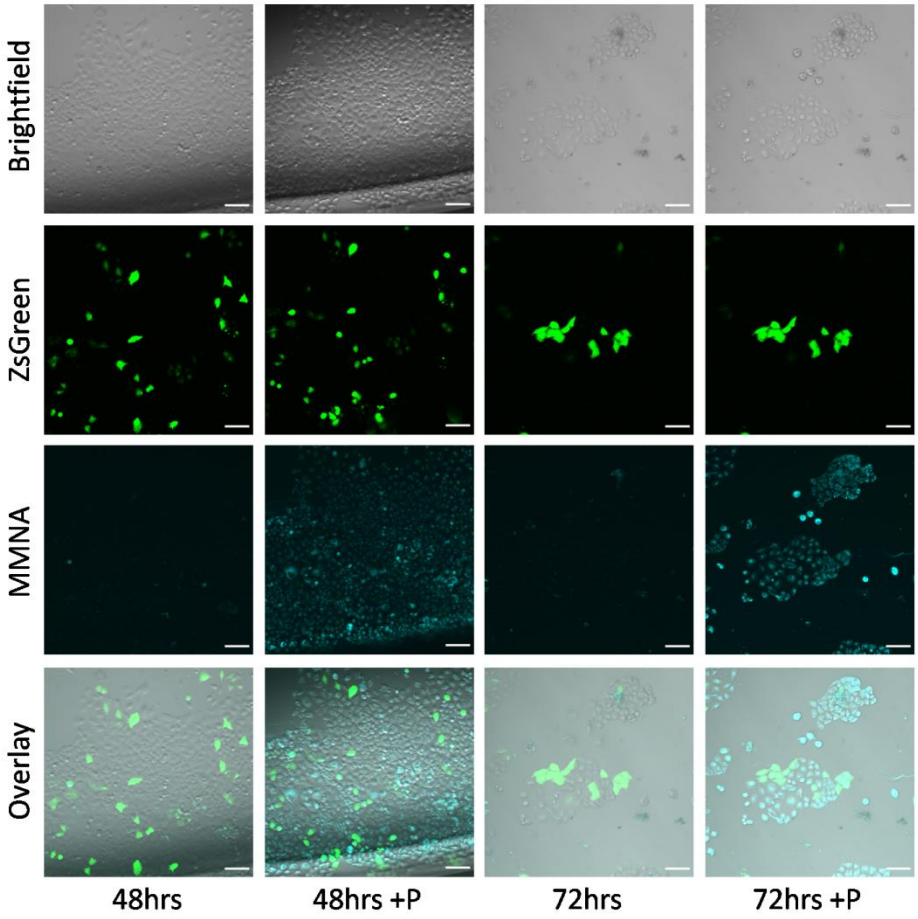

Figure 4.2.6. Time-course of transient Alfred expression in the Tet-On Inducible System using the A549-TET3G cell line and overexpression of FGE. pTRE3G-ZsGreen1-Alfred was transiently transfected into A549-TET3G cells. They were induced with doxycycline for 48 or 72 hours, incubated with methyl 5-MeO-*N*-anthranilate (MMNA) for 20 minutes, and imaged at 10× for ZsGreen and MMNA. Scale bar = 100 µm.

4.3.1. Creation of pCMV-Alfred

The FGE gene in pCMV-FGE was removed and replaced with the Alfred gene to create pCMV-Alfred and then screened with PCR and restriction enzyme digest. This would allow constitutive expression of Alfred in the same manner as FGE, since both proteins will originate from the same backbone vector.

The Alfred gene was obtained through PCR from pCMV-Alfred, and the FGE gene was removed from pCMV-FGE through PCR as well. The Alfred gene was then inserted into the empty pCMV vector. If the ligation was successful, digestion with NdeI and ScaI should yield 2 fragments, 927 bp and 4705 bp, which should sum up to the expected total plasmid length, 5632 bp. If the Alfred gene was not inserted into pCMV, the fragments should be 927 bp and 4405 bp. If the PCR on pCMV-FGE was unsuccessful, and vector retained the FGE gene, the fragments should be 927 bp and 5545 bp. The digests yielded fragments that were close to the expected length (Figure 4.3.1.), which does not confirm successful ligation. Therefore, the samples were sent for DNA sequencing, which resulted in a 100% alignment with the reference sequence (Appendix B).

4.3.2. Generation of a cell line stably expressing FGE

A549-FGE, a stable A549 cell line containing pCMV-FGE, was generated using G418 resistance as the selection marker using the same G418 concentration as determined previously (Figure 4.2.1.). After transfection, the cells were split into 10×10 cm dishes to allow dispersion of cells for individual surviving cells to grow into colonies. After selection for at least 2 weeks, 6 colonies were isolated using trypsin-soaked Whatman™ filter paper, and then grown to a sufficient number for testing in Western blot using anti-FGE antibody. Full length FGE is expected to appear at ~41 kDa; secreted FGE is cleaved to result in an N-terminally truncated ~37 kDa protein [23]. Notably, sample T4 showed 2 bands tagged with the anti-FGE antibody (Figure 4.3.2.), showing an increase in FGE content compared to the other samples. Since these

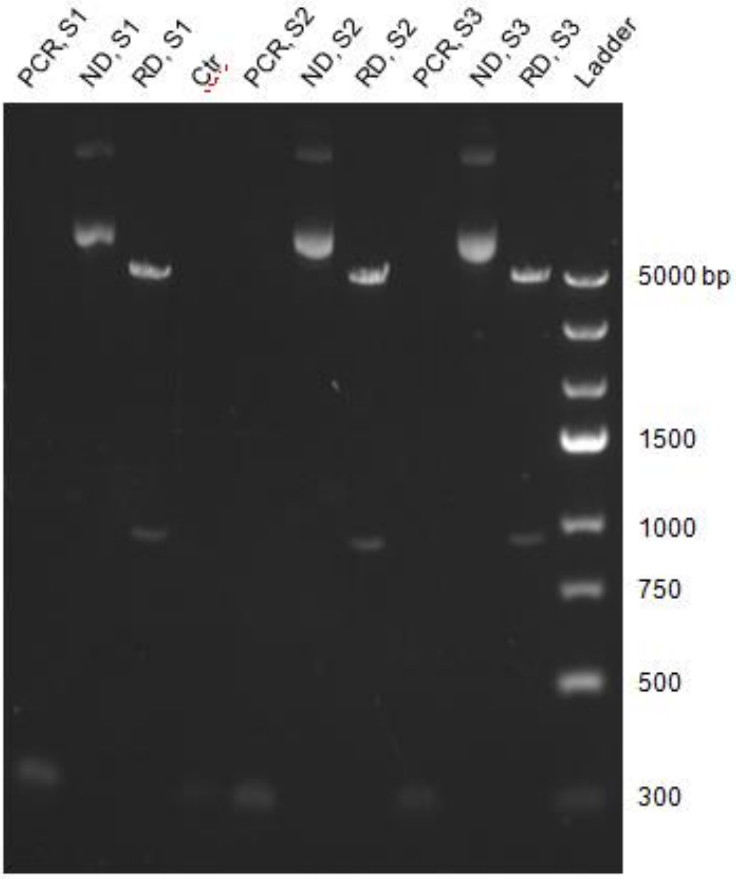

Figure 4.3.1. PCR, and NdeI and ScaI digest screening of pCMV-Alfred. Samples extracted from transformed bacterial cultures and were digested with Nde and ScaI restriction enzymes. PCR was conducted on the non-digested sample using sequencing primers. The PCR product, non-digested samples, and digested samples were screened with electrophoresis on a 1% agarose gel. The PCR fragment is located at ~300 bp. The non-digested sample is located at ~6000 bp. The digest fragments are located at ~5000 bp and ~900 bp. PCR: PCR product using sequencing primers; ND: non-digested; RD: restriction enzyme digest using NdeI and ScaI; S1-S3: samples 1-3; Ctr: negative control (water); Ladder: GeneRuler Express DNA Ladder. All non-digested samples are diluted 1:10.

lysates were obtained after antibiotic selection and selection and each colony was physically isolated, this indicates that the cells in each colony should be a pure cell line stably transfected with pCMV-FGE. Therefore, the T4 cell line was grown for use in subsequent experiments.

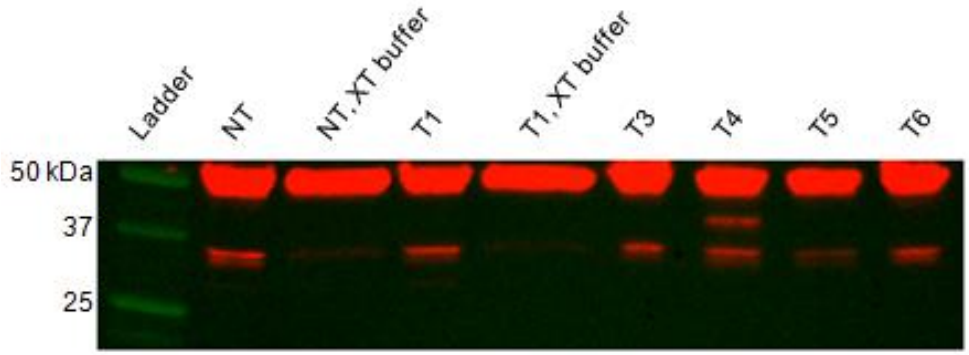

Figure 4.3.2. FGE content in wild type A549 or A549-FGE cell lines. Cell lysates were obtained from A549 cells, either from a wild type cell line (sample NT) or stably transfected cells (samples T1, T3-T6). Samples were prepared in Laemmli buffer, except lanes 3 and 5, which were prepared in XT buffer. Samples were run on SDS-page (4-20% polyacrylamide in a top-bottom gradient), and the proteins were tagged with anti-FGE antibody (anti-rabbit secondary) and anti-tubulin antibody (anti-rabbit secondary). FGE is located at ~41 kDa and ~37 kDa, and tubulin is the loading control, located at 51 kDa. Ladder: ProteinPlus All Blue ladder; NT: not transfected; T1-T6: transfected samples 1-6.

4.3.3. Alfred expression in the A549-FGE cell line

pCMV-Alfred was transiently transfected into wild type A549 or A549-FGE cells. Cells were imaged 24 hours post-transfection, and then imaged again after incubating with MMNA for 20 minutes (Figure 4.3.3.). There was no visible difference in MMNA fluorescence between non-transfected cells and Alfred-transfected cells, nor any difference between wild type A549 cells or A549-FGE cells. In addition, microscopy continued to yield results consistent with previous observations: diffuse MMNA fluorescence was shown in green-fluorescing cells, cells exhibited MMNA fluorescence in the perinuclear region of the cell (Figure 4.1.4.B), and punctate fluorescent foci were present in cells that did not fluoresce green (Figure 4.2.3.).

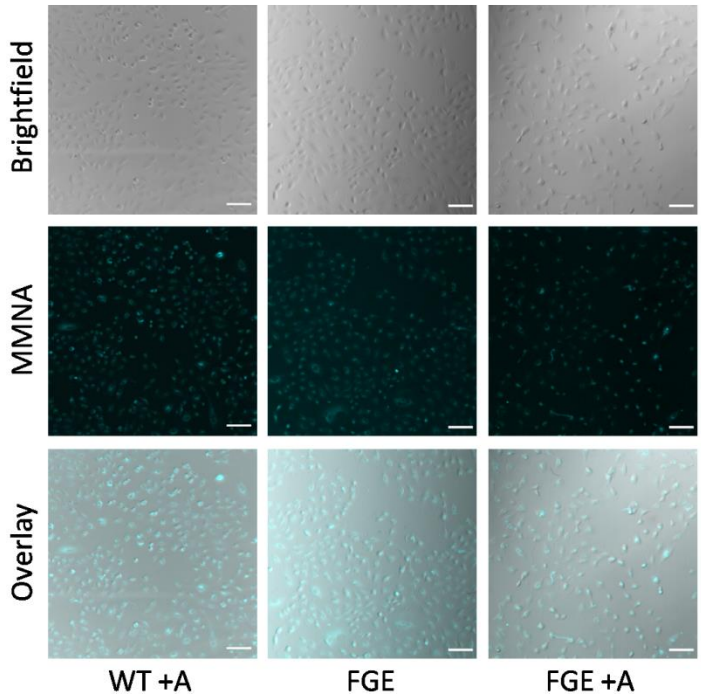

Figure 4.3.3. Constitutive Alfred expression in wild type A549 and A549-FGE cell lines. pCMV-Alfred (A) was transiently transfected into wild type A549 (WT) or A549-FGE (FGE) cells. Cells were imaged after 24 hours, incubated with methyl 5-MeO-*N*-anthranilate (MMNA) for 20 minutes, and imaged at 10× for ZsGreen and MMNA. Scale bar = 100 µm.

Next, a time-course experiment was conducted. Cells were imaged 48 or 72 hours post-transfection, and then imaged again after incubating with MMNA for 20 minutes. First, transfection efficiency was established by co-transfection with pCMV-Alfred and commercially available mCardinal-N1 plasmid. In this way, mCardinal fluorescence could guide imaging to locations with potentially higher Alfred transfection when imaged with MMNA. Imaging the same field of view (Figure 4.3.4.), the cells showing red fluorescence showed that MMNA fluorescence is not visibly different from previous observations of the perinuclear localization nor the punctate foci in non-transfected cells, and therefore mCardinal should not affect MMNA fluorescence. This also indicates an absence of FRET.

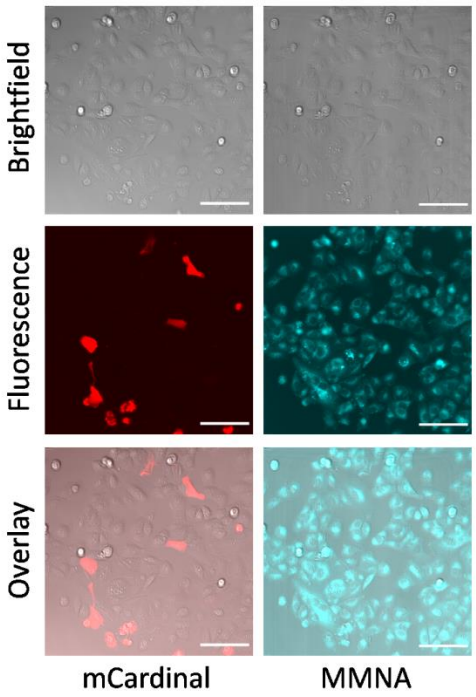

Figure 4.3.4. Transient Alfred expression in the vicinity of mCardinal expression in the A549-FGE cell line. pCMV-Alfred and mCardinal-N1 were transiently co-transfected into A549-FGE cells. They were incubated for 48 hours, incubated with methyl 5-MeO-*N*-anthranilate (MMNA) for 20 minutes, and imaged at 20× for fluorescence (red: mCardinal; blue: MMNA). Scale bar = 100 µm.

Then, the rest of the cells were imaged in the vicinity of higher mCardinal expression (Figure 4.3.5., 4.3.6.). Transfection efficiency is low, but present at all 3 timepoints. Cells seemed to have increased fluorescence over the course of 72 hours. However, there was also an increase in cell confluency and death over this time. In addition, the intensity and localization of MMNA fluorescence is consistent with previous observations (Figure 4.1.4.B, 4.2.3.) in both non-transfected and transfected cells, and in wild type A549 cells and A549-FGE cells.

4.4. Verification of Alfred expression

Though previously it was hypothesized that the treatment conditions or imaging parameters were not optimized, it was now postulated that Alfred may not be expressed at all.

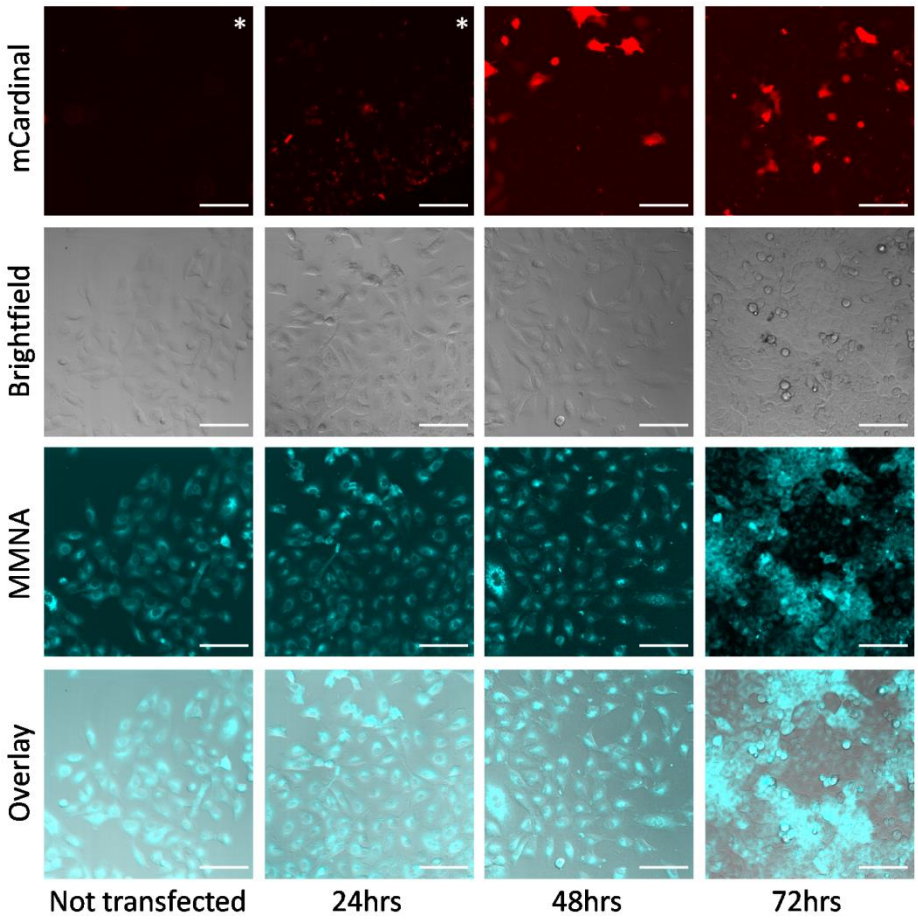

Figure 4.3.5. Transient Alfred expression in the wild type A549 cell line. mCardinal-N1 was transiently co-transfected with pCMV-Alfred into A549 cells. They were incubated for 24, 48, or 72 hours, then incubated with methyl 5-MeO-*N*-anthranilate (MMNA) for 20 minutes, and imaged at 20× for MMNA in an area where mCardinal fluorescence is higher. mCardinal is imaged separately from the other three channels, and is not included in the overlay. *indicates images taken in a separate field of view. Scale bar = 100 µm.

Wild type A549 and A549-FGE cells were transiently transfected with pCMV-Alfred and harvested to test for Alfred content. A549 does not endogenously express any c-Myc. In addition, the anti-Myc antibody 9E10 specifically does not target endogenous c-Myc, but does identify Myc-tagged proteins, so bands identified by anti-Myc antibody in Western blot should indicate the presence of Alfred (Figure 4.4.1.). Alfred is expected to appear at ~11 kDa, while the positive

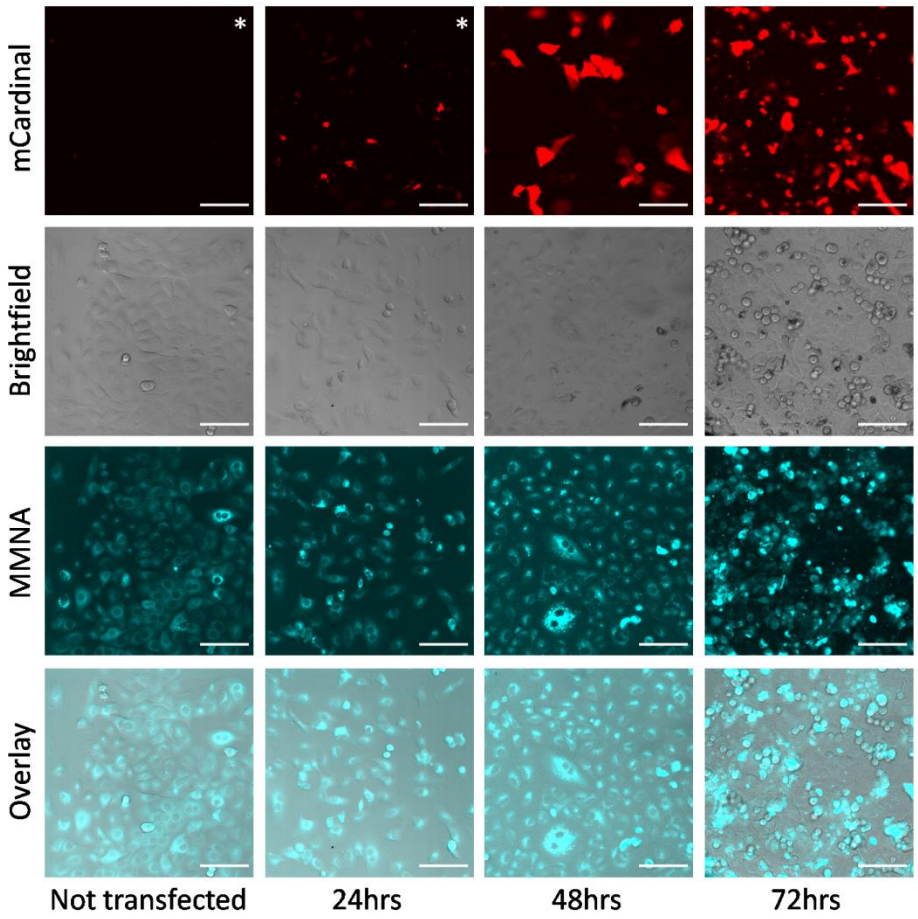

Figure 4.3.6. Transient Alfred expression in the A549-FGE cell line. mCardinal-N1 was transiently co-transfected with pCMV-Alfred into A549-FGE cells. They were incubated for 24, 48, or 72 hours, then incubated with methyl 5-MeO-*N*-anthranilate (MMNA) for 20 minutes, and imaged at 20× for MMNA in an area where mCardinal fluorescence is higher. mCardinal is imaged separately from the other three channels, and is not included in the overlay. *indicates images taken in a separate field of view. Scale bar = 100 µm.

control is expected to appear at ~50 kDa. Since the protein of interest is small, the samples were allowed to migrate ~2/3 of the way down the gel to prevent loss of the sample. The lanes containing the cell lysates do not contain a band when tagged with the anti-Myc antibody. However, the positive control shows a band at the expected weight, indicating that the antibody does work in identifying Myc-tagged proteins. It is important to note that the signal arising from the positive control may be so high that a very faint signal indicative of the presence of Alfred

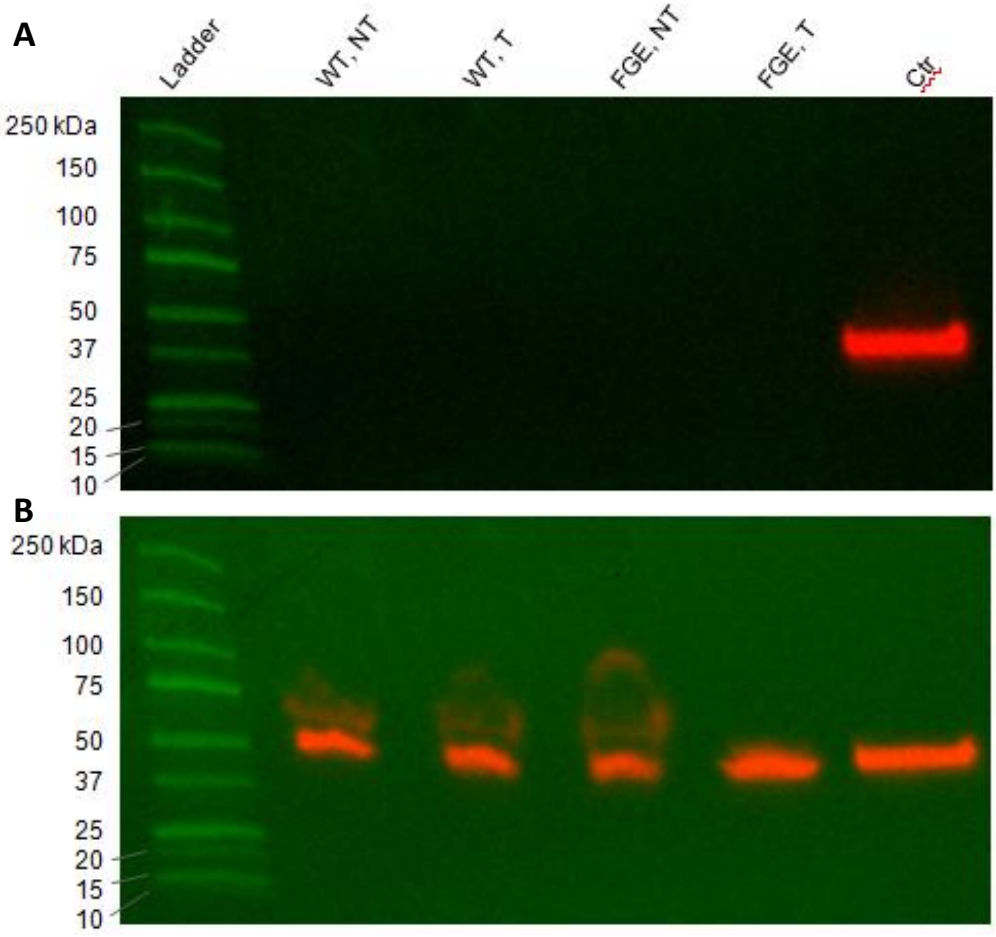

Figure 4.4.1. Alfred content in wild type A549 or A549-FGE cell lines transiently transfected with pCMV-Alfred. Cell lysates were obtained from wild type A549 cells or A549-FGE cells transiently transfected with pCMV-Alfred after 24 hours. A positive control was obtained from a bacterial culture stably expressing a Myc tagged protein. Samples were run on SDS-page (4-20% polyacrylamide in a top-bottom gradient), and the proteins were tagged with **A.** anti-Myc antibody (anti-mouse secondary) and then, without stripping the anti-mouse antibody, with **B.** anti-tubulin antibody (anti-rabbit secondary). Alfred is located at ~11 kDa, the positive control is located at ~50 kDa, and tubulin is the loading control, located at 51 kDa. Ladder: ProteinPlus All Blue ladder; WT: wild type cell line; FGE: FGE-stable cell line; NT: not transfected; T: transfected; Ctr: bacterial culture.

may be indiscernible, even with significant overexposure times. Regardless, since the presence of Alfred is unable to be confirmed, it is hypothesized that the cells imaged in previous figures do not show an increase in MMNA fluorescence because Alfred is not expressed.

4.5. Prediction of protein secondary structure

In order to assess if protein folding could interrupt the signal from Alfred, the secondary structure of Wu et al.'s construct [7] and that of Alfred was predicted using online software. Two software were used: Chou & Fasman Secondary Structure Prediction (CFSSP) (Figure 4.5.1., 4.5.2.); and the improved self-optimized prediction method (SOPMA) (Figure 4.5.3., 4.5.4.). Alignment of the secondary structures between Wu et al.'s construct [7] and Alfred show that these structures are similar, besides the gaps where Wu et al.'s construct [7] contained the

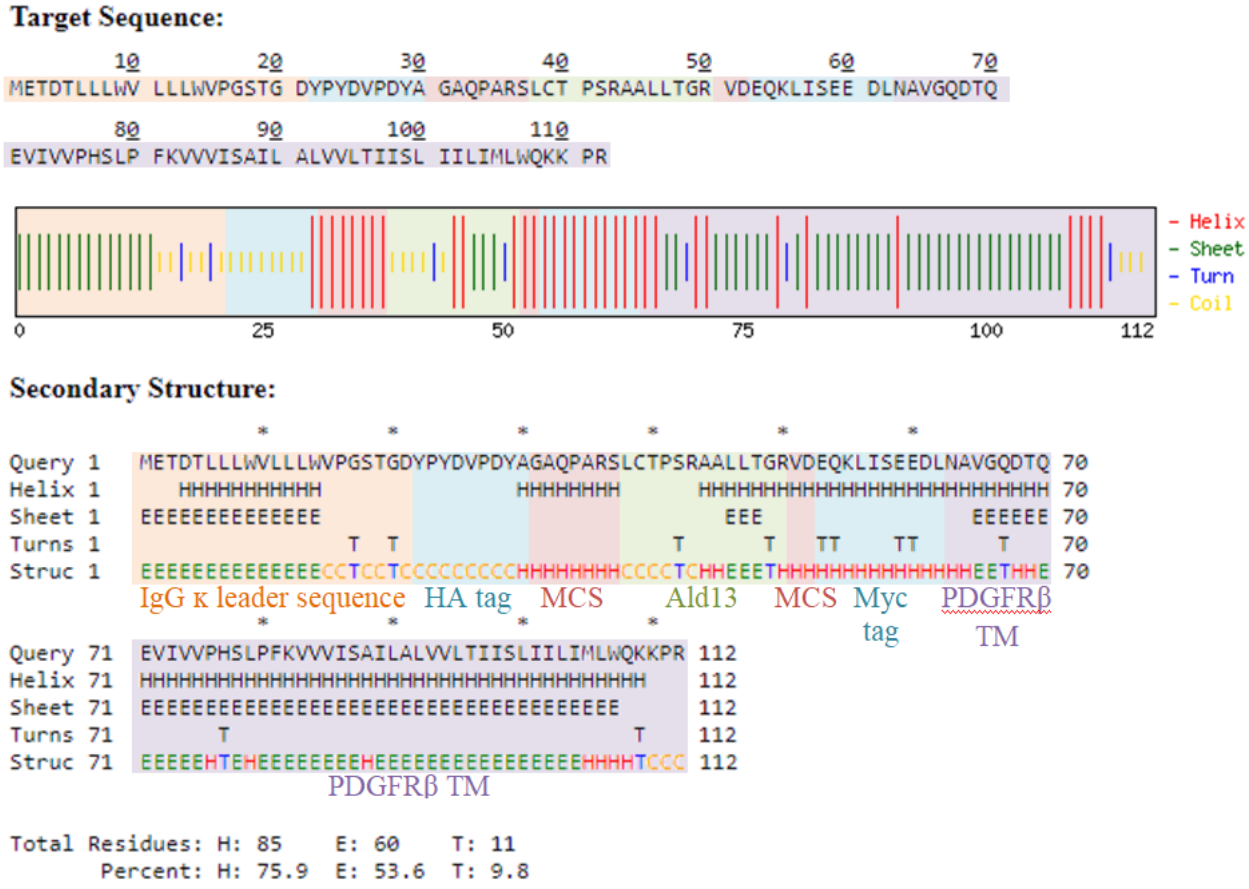

Figure 4.5.1. Secondary structure of Wu et al.'s construct [7] as predicted by Chou & Fasman Secondary Structure Prediction (CFSSP). The amino acid sequence was inputted into CFSSP available online. Regions of the protein are annotated in colour. Orange: immunoglobin κ (IgG κ) leader sequence; blue: hemagglutinin epitope (HA) tag or Myc tag; red: multiple cloning site (MCS); green: aldehyde tag (Ald13); purple: platelet derived growth factor receptor β transmembrane (PDGFRβ TM) domain.

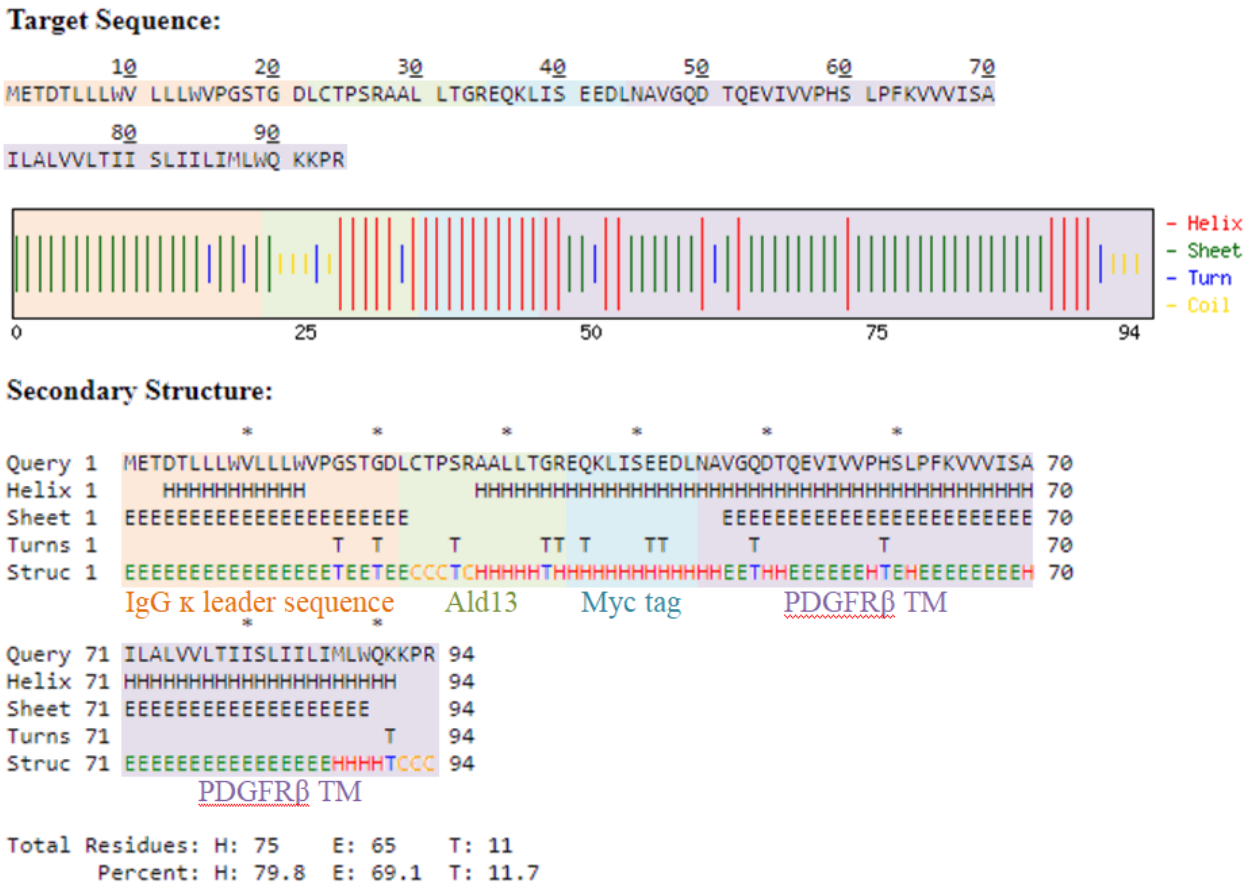

Figure 4.5.2. Secondary structure of Alfred as predicted by Chou & Fasman Secondary Structure Prediction (CFSSP). The amino acid sequence was inputted into CFSSP available online. Regions of the protein are annotated in colour. Orange: immunoglobin κ (IgG κ) leader sequence; green: aldehyde tag (Ald13); blue: Myc tag; purple: platelet derived growth factor receptor β transmembrane (PDGFRβ TM) domain.

hemagglutinin (HA) epitope sequence and the multiple cloning site (MCS) sequences (Figure 4.5.5., 4.5.6.) that are not present in Alfred. These software produced different predictions, but the alignment to Wu et al.'s construct [7] remains similar. However, one small discrepancy between the secondary structure of the proteins is identified by aligning the CFSSP predictions (Figure 4.5.5.). In the FGE substrate site (Ald13) within Alfred, there is a small segment of helix whereas this segment is a sheet in Wu et al.'s construct [7].

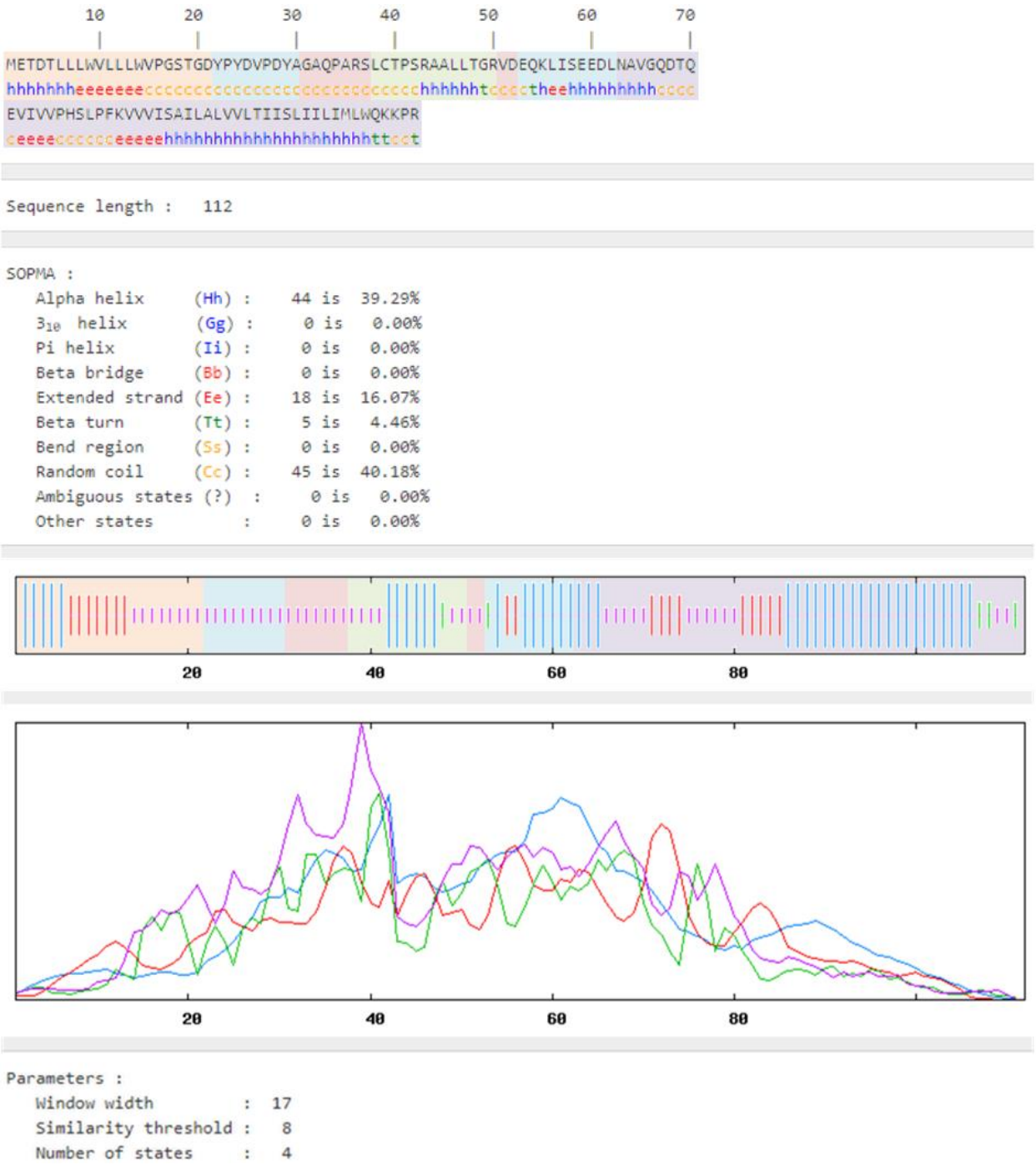

Figure 4.5.3. Secondary structure of Wu et al.'s construct [7] as predicted by the improved self-optimized prediction method (SOPMA). The amino acid sequence was inputted into SOPMA available online. Regions of the protein are annotated in colour. Orange: immunoglobin κ (IgG κ) leader sequence; blue: hemagglutinin epitope (HA) tag or Myc tag; red: multiple cloning site (MCS); green: aldehyde tag (Ald13); purple: platelet derived growth factor receptor β transmembrane (PDGFRβ TM) domain.

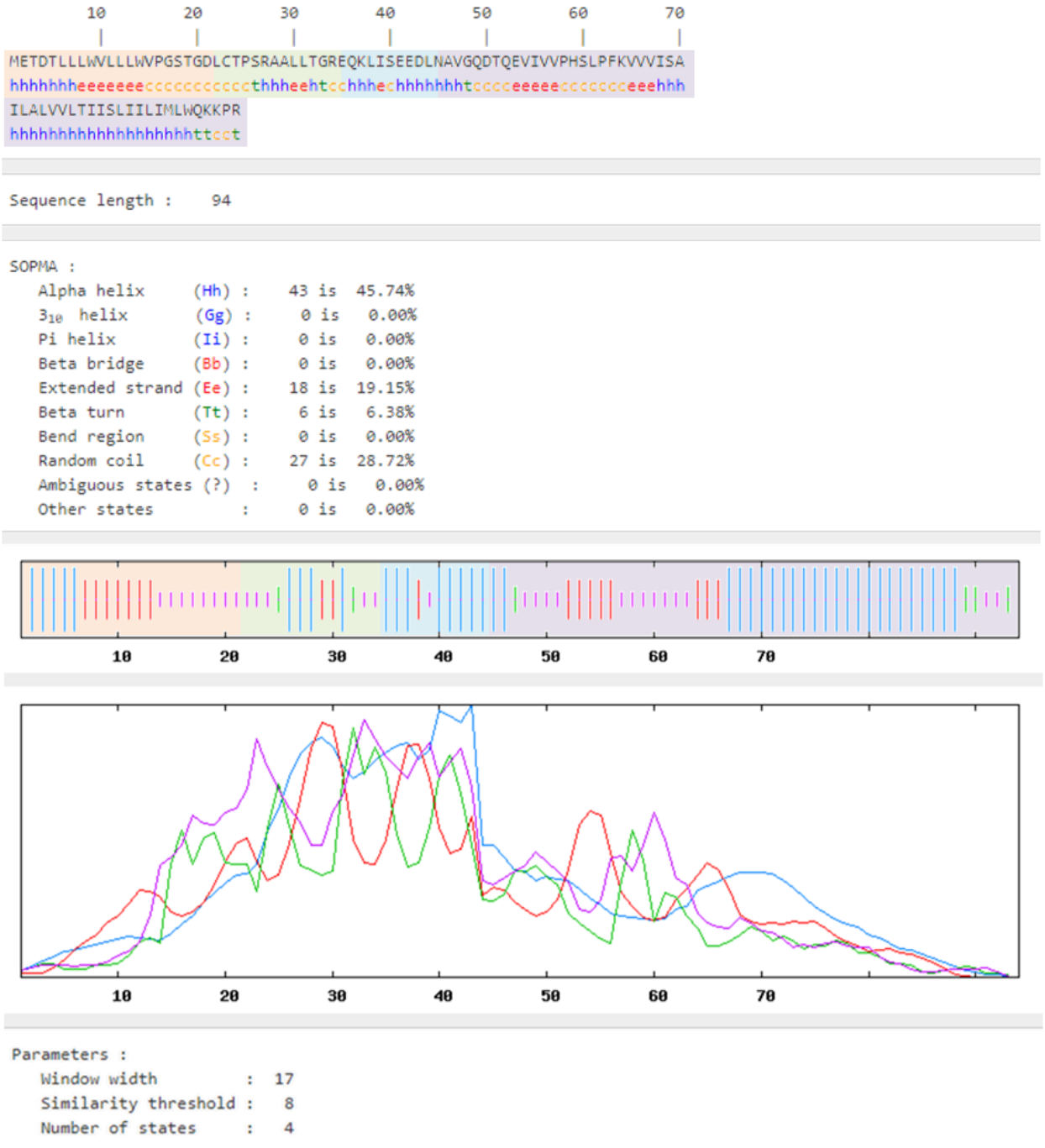

Figure 4.5.4. Secondary structure of Alfred as predicted by the improved self-optimized prediction method (SOPMA). The amino acid sequence was inputted into SOPMA available online. Regions of the protein are annotated in colour. Orange: immunoglobin κ (IgG κ) leader sequence; green: aldehyde tag (Ald13); blue: Myc tag; purple: platelet derived growth factor receptor β transmembrane (PDGFRβ TM) domain.

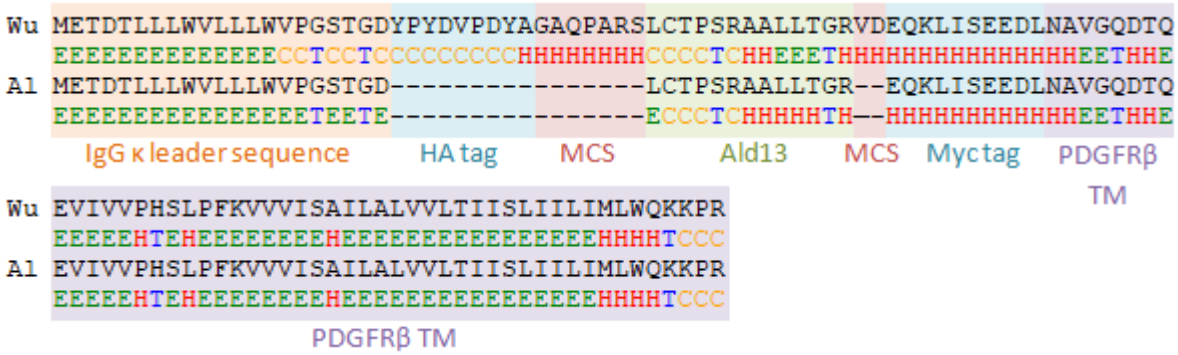

Figure 4.5.5. Alignment of the secondary structure of Wu et al.'s construct [7] and Alfred, as predicted by Chou & Fasman Secondary Structure Prediction (CFSSP). The amino acid sequence of Wu et al.'s construct (Wu) [7] and Alfred (Al) were aligned and matched with the corresponding secondary structure as predicted by CFSSP available online. Regions of the protein are annotated in colour. Orange: immunoglobin κ (IgG κ) leader sequence; green: aldehyde tag (Ald13); blue: Myc tag; purple: platelet derived growth factor receptor β transmembrane (PDGFRβ TM) domain. C: coil; E: sheet; H: helix; T: turn.

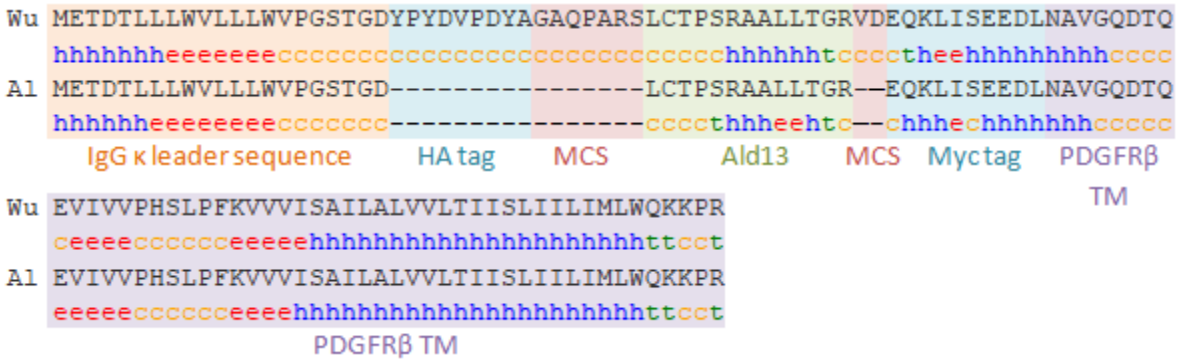

Figure 4.5.6. Alignment of the secondary structure of Wu et al.'s construct [7] and Alfred, as predicted by the improved self-optimized prediction method (SOPMA). The amino acid sequence of Wu et al.'s construct (Wu) [7] and Alfred (Al) were aligned and matched with the corresponding secondary structure as predicted by SOPMA available online. Regions of the protein are annotated in colour. Orange: immunoglobin κ (IgG κ) leader sequence; green: aldehyde tag (Ald13); blue: Myc tag; purple: platelet derived growth factor receptor β transmembrane (PDGFRβ TM) domain. c: random coil; e: extended strand; h: α-helix; t: β turn.

5. Discussion

This work aimed to create a novel reporter gene called Alfred, a transmembrane (TM) protein that displays an aldehyde on the cell surface. *En route* to the cell surface through the endoplasmic reticulum, the 13-mer formylglycine generating enzyme (FGE) substrate sequence allows a specific cysteine residue to be converted into a formylglycine residue by FGE, a sulfatase-modifying enzyme endogenous to the endoplasmic reticulum. Aldehyde probes such as methyl 5-MeO-*N*-aminoanthanilate (MMNA) can detect this displayed aldehyde, and as a result, the aldehyde should be visualized as punctate fluorescent foci on the cell surface when imaged with fluorescence microscopy. However, this was not reflected in the microscopy results. Using the Tet-On® Inducible System allowed expression of the gene of interest (GOI), Alfred, to be chemically controlled by the presence of doxycyclin (dox). ZsGreen was used as a positive control for the induction of transcription in the Tet-On® Inducible System. Fluorescent foci were not observed to be increased in cells fluorescing green. In addition, the foci were observed in non-transfected cells as well as Alfred-transfected cells. Microscopy results did not change in a time-course experiment, nor in cells overexpressing FGE. In the end, Alfred was undetectable by Western blot in cells transfected to express Alfred constitutively.

In summary, the results did not corroborate the expected outcome. The absence of Alfred should be investigated further to improve on the expression and detection of Alfred.

5.1. Microscopy results

pTRE3G-ZsGreen1-Alfred was constructed in order to express Alfred using the Tet-On® Inducible System. The system was tested with positive and negative controls, and its fidelity and

effect on A549 cells was established. The absence of FRET was also established using a GFP-expressing cell line and comparing it to the non-fluorescent parent cell line. Expression of Alfred with the Tet-On® Inducible System did not result in visible differences between co-transfected and non-transfected cells when tagged with MMNA; i.e. no difference in blue fluorescence between cells fluorescing green compared to non-fluorescent cells. To troubleshoot these results, a variety of changes were made. The A549-TET3G cell line was created in order to increase the number of cells expressing ZsGreen and Alfred. FGE was overexpressed in order to increase thiol-to-aldehyde conversion for increased detection of Alfred. Time-course studies were conducted to allow an increased amount of Alfred to reach the cell surface. Nonetheless, microscopy did not show any difference between Alfred-transfected cells and non-transfected cells in any of the tested conditions.

Alfred was then expressed constitutively to eliminate the possibility that the Tet-On® Inducible System reduced Alfred expression. Overexpression of FGE, time-course studies, and increased magnification still did not show the punctate fluorescent foci indicative of tagged TM proteins. Finally, Western blot was conducted on cell lysates that had been transfected with pCMV-Alfred, but Alfred (containing a Myc tag) was undetectable by anti-Myc antibody.

5.1.1. Initial testing of Alfred expression

The advantage of the Tet-On® Inducible System is that it allows controlled expression of the GOI and easy identification of transfected cells. It was established that the system worked as described: addition of dox resulted in expression of the reporter gene, ZsGreen (Figure 4.1.2). This inferred that the TET3G transactivator protein was expressed constitutively and that binding of dox lead to induction of p_{TRE3G} to express ZsGreen. FRET did not contribute to a decrease of

the MMNA signal in the presence of GFP (Figure 4.1.3.), and therefore Alfred detection by MMNA should not have been affected by FRET. Together, these controls established that the inability to obtain the expected results does not arise from failure of the Tet-On® Inducible System, nor from FRET.

In Figure 4.1.4., it can be seen that green fluorescence (originating from ZsGreen) and MMNA fluorescence did not always overlap. Also, the non-transfected control did not show an apparent difference from the transfected cells. Therefore, it can be inferred that the brighter MMNA fluorescence did not represent detection of Alfred, because cells that were not transfected (not green) exhibited similar patterns of MMNA fluorescence compared to the cells that were transfected. In addition, it was observed that aldehydes were localized in the perinuclear region of the cell, instead of in punctate foci scattered throughout the cell's surface. This is hypothesized to be in the translocation pathway of a TM protein *en route* to the cell membrane through the endoplasmic reticulum [12]. In other words, Alfred may still be residing within the endoplasmic reticulum at the time of imaging. Therefore, it was hypothesized that waiting a longer period of time before imaging may allow Alfred to reach the cell surface.

5.1.2. Time-course of Alfred expression

Since it was hypothesized that there was insufficient time for protein translocation, the cells were left to incubate for longer than 24 hours.

After transcription, the length of time required for Alfred to display an aldehyde on the cell surface is dependent on multiple factors including: the translation rate of Alfred; the IgG κ leader sequence targeting Alfred to the endoplasmic reticulum; and localization of Alfred from the endoplasmic reticulum to the cell surface by the PDGFRβ TM domain. In regular stress-free

conditions, the rate of protein expression is related to the length of the mRNA and cell volume [24]. In general, protein synthesis occurs on the scale of seconds to minutes [25]. Assuming Alfred has a similar half-life to PDGFRβ, the protein should not be degraded for 3 hours [26]. This should lead to accumulation of the protein within the cell and ideally remain at a reasonable level. The Tet-On® Inducible Systems User Manual recommends imaging the cells at 24 hours after transfection [5]; however other sources image proteins at 48 hours after transfection [7]. In previous studies, PDGFRβ was imaged at 48 hours post transfection in brain cells [27].

The rate at which Alfred is targeted to the endoplasmic reticulum may be limited by translation of the IgG κ signal peptide by the ribosome and its subsequent binding by signal binding proteins (SBPs) [12]. Previous studies have used the same IgG κ leader sequence, and achieved cleavage at the expected site (immediately after the last residue of the signal peptide) from cells harvested at 48 hours post-transfection [28], and therefore this timepoint may be better for viewing Alfred expression. Though not specifically studying the IgG κ signal peptide, other studies have been able to achieve expression of IgG in A549 cells [29].

Furthermore, since Alfred contains the PDGFRβ TM domain, if its localization in the cell membrane was similar to that of PDGFR, then it should localize efficiently [30]. This is further supported in previous studies where PDGFRβ has been found to be involved in the proliferation and migration of A549 cells [31], and is related to the growth of tumour mass in non-small-cell lung cancer (NSCLC) [31]. Therefore, the absence of Alfred from the cell membrane would not be due to failure of the ability of PDGFRβ to integrate into the phospholipid bilayer. However, if the IgG k leader sequence, aldehyde tag, and/or the Myc tag lead to protein folding, the tertiary structure of the PDGFRβ TM domain may inhibit the TM domain's residues to prevent the anchor sequence in the TM domain from being held properly within the bilayer [32]. This may

result in the polypeptide becoming completely inserted in the endoplasmic reticulum and eventually secreted from the cell as if it were a secreted protein [12] (discussed in Section 5.2.1.).

These issues were addressed in the subsequent experiments, after pCMV-TET3G was stably transfected into A549 cells (Figure 4.2.2.) to increase the proportion of transfected cells. The time-course experiment, where the cells were imaged at 24 hours (Figure 4.2.3.), 48 hours and 72 hours post-induction (Figure 4.2.4.), also exhibited similar fluorescence patterns that showed there was no increased MMNA fluorescence exclusively in cells that fluoresced green. Thus, it was hypothesized that time was not the issue.

5.1.3. Increase of thiol-to-aldehyde conversion

Given that aldehydes exist endogenously, perhaps the presented aldehyde from Alfred does not occur at a high enough concentration to be distinguished from the rest of the Aldehydic Load. One issue may be that there was insufficient FGE in the endoplasmic reticulum for noticeable thiol-to-aldehyde conversion. Therefore, overexpression of FGE should increase the signal from Alfred.

However, co-transfection with pCMV-FGE and pTRE3G-ZsGreen1-Alfred also did not show differences between the transfected samples and the non-transfected controls (Figure 4.2.5.). The time-course study also did not yield any visible differences (Figure 4.2.6.). Since the transfection rate of pCMV-FGE cannot be assessed, a cell line stably expressing FGE was created for subsequent experiments.

5.1.4. Constitutive Alfred expression

Alfred is expressed after the IRES in the Tet-On® Inducible System, and therefore the problem may be that Alfred is translated via cap-dependent translation using IRES2. In general, the gene after the IRES is expected to be expressed at ~20-50% of the gene before the IRES, which may be insufficient to produce a distinguishable signal [33]. Therefore, pCMV-Alfred was created (Figure 4.3.1.) to constitutively express Alfred.

Microscopy continued to show similar results (Figure 4.3.3., 4.3.5, and 4.3.6.). Although this strategy does not allow analysis of specific cells, comparison of Alfred-transfected cells to the non-transfected cells shows no difference, regardless of overexpression of FGE. Cells were co-transfected with mCardinal-N1 to assess transfection efficiency, and imaging occurred near cells exhibiting red fluorescence, in attempts to image an area where transfection may be increased compared to other areas. The increased magnification allowed better visualization of the localization of MMNA fluorescence within the cells; however, the punctate fluorescent foci are present in both non-transfected control cells and in Alfred-transfected cells. This lead to the hypothesis that the observed foci are not indicative of Alfred localization (discussed in Section 5.2.4). Altogether, imaging showed similar localization of MMNA between non-transfected and transfected cells, indicating that Alfred is not identified in the cells. This occurred with constitutive Alfred expression over a period of 72 hours, with and without overexpression of FGE.

5.1.5. Verification of Alfred expression

As a final investigation, lysates harvested from cells constitutively expressing Alfred were used for SDS-page and Western blotting. The presence of Alfred, which should be

identified by its myc tag, was unable to be confirmed. Tubulin was used as a loading control; however, while bacteria contain tubulin homologs with low sequence similarity [34], and while the exact sequence of the tubulin antibody used is proprietary, tubulin may not be an appropriate control indicative of the amount of protein loaded. Instead, Ponceau can be used to assess total protein content, and myc-tagged protein content can be normalized to this total. This would also allow a similar amount of protein content to be loaded, perhaps decreasing the signal from the bacterial sample, allowing the potentially faint signal arising from Alfred to present itself if it existed.

Regardless, the lack of (or perhaps, minimal) Alfred expression explains the lack of distinctive Alfred labelling in microscopy. FGE expression is under control of the same CMV promoter; in fact, the entire plasmid is identical except for the gene expressed by CMV. Since FGE expression was verifiable with Western blot (Figure 4.3.2.), it is unlikely that the lack of the Alfred protein is due to lack of transcription induced by the CMV promoter, even if the expression from these two promoters may not be equal. The CMV promoter has also been used in previous studies where multiple pCMV plasmids expressed genes at comparable levels in different cell lines [35,36]. Therefore, it can be hypothesized that the RNA transcript may be degraded prior to translation.

RNA degradation occurs naturally within eukaryotic cells as part of maintaining cellular homeostasis by removing non-functional or aberrant RNA molecules [37]. In particular, mRNA degradation is regulated by the presence of the 5’ cap and adjacent functional groups for exonuclease binding, methylation of bases to prevent endonuclease binding, secondary structure of the mRNA to increase stability, and presence of specific 3’ UTR motifs [37,38]. The Alfred transcript does not contain notable sequences resulting in mRNA degradation. Most notably,

restriction enzymes are not endogenous in eukaryotic cells, and thus the restriction sites present in Alfred should not be an issue within A549 cells to target the transcript for degradation [39]. One of the common sequence-specific 3' UTR motifs leading to mRNA decay is also not present in Alfred [38].

It can also be hypothesized that the Alfred protein is degraded quickly and not accumulated in the cell even if the protein is translated (Section 5.1.2. discusses expression of IgG and PDGFRβ). Protein degradation also occurs naturally to remove malfunctioning or damaged proteins, and degradation of endogenous and exogenous proteins is part of the adaptive immune response [40]. Proteins can be targeted for removal by use of a degron or ubiquitin tag to be identified for degradation by the proteasome [40]. Unstructured segments of the protein are also required for protein degradation, and serve as degradation initiation sites [40]; however, prediction of the secondary structure did not show unstructured regions within Alfred that are not present within Wu et al.'s construct [7] (Figure 4.5.2., 4.5.4.) [41]. Other protein-specific proteases may exist that can degrade Alfred; however the components of Alfred are modelled after the commercially available pDisplay plasmid gene product and have been previously tested to express the aldehyde tag [7,42].

5.2. Differences from previous studies

Wu et al.'s study [7] was a heavy influence on the design of this project. Hence, comparisons will be made between this project and that study in order to examine discrepancies that could result in different outcomes.

5.2.1. Presence of epitopes

In their study, Wu ligated the aldehyde tag (the 13-mer iteration) in between BglII and SalI of pDisplay (Invitrogen) [7,42]. As a result, the gene follows a hemagglutinin A (HA) epitope, an epitope containing a high proportion of charged resides [43], and originates from a viral glycoprotein currently studied as a vaccine target [44]. The gene is followed by the Myc epitope, which is derived from the c-Myc oncogene, a master regulator in cell growth regulation and metabolism [45]. The epitopes are not directly adjacent to the aldehyde tag, but rather the aldehyde tag is flanked by remaining nucleotides of the MCS; this whole segment is preceded by the T7 promoter site and an IgG κ signal peptide leader sequence.

The Alfred gene differs from this design in that it does not contain the HA epitope, and also the aldehyde tag is directly followed by the Myc epitope without arbitrary nucleotides in between. The rest of the gene design is similar to Wu et al.'s gene [7], in terms of arrangement and components, but the Alfred gene block had also been codon-optimized in order to achieve the maximum expression of the protein in human cells, and therefore the nucleotide sequence is different from that of Wu et al. [7], while the amino acid sequence remains the same.

Since both HA and Myc are small epitopes (9 and 10 amino acids, respectively) [43], they are generally hypothesized to not affect protein structure, function, localization, or stability [46]. Epitopes are mainly used for tagging the protein for identification in antibody-based technologies, including Western blotting [46]. It is unlikely that absence of the HA epitope (YPYDVPDYA) or MCS residues (GAQPA, VD), that were present in Wu et al.'s construct [7], would reduce or affect Alfred expression. The charged residues should not result in detrimental secondary structure, as its residues should not form α-helices capable of interacting with the single TM helix of the PDGFRβ TM domain [47,48]. Comparison of the secondary structure

between Wu et al.'s construct [7] and Alfred did not yield significant differences besides presence of the HA tag and MCS residues, as predicted by the improved self-optimized prediction method (SOPMA) available online, likely due to the alignment consensus-based algorithm (Figure 4.5.6.) [41]. This means that the secondary structure should not be the reason that Alfred would be unable to integrate into the phospholipid bilayer, since its structure is the similar to Wu et al.'s construct [7]. However, the structures are different as predicted by the Chou & Fasman Secondary Structure Prediction Server (CFSSP) (Figure 4.5.5.), notably that there is a small segment of helix within the Ald13 of Alfred and a small segment of helix within the HA tag/MCS of Wu et al.'s construct [7] (Figure 4.5.2., 4.5.4.) [49]. These differences were hypothesized to be insignificant. First, CFSSP was unable to identify the PDGFRβ TM domain as a helix [50]; rather it is predicted to be a sheet, and therefore this algorithmic method may not be suitable for this application. Second, each helix is very short, consisting of ~1.5-2 turns of the helix, so likely only 3-4 amino acids would be interacting with the PDGFRβ TM domain helix and should therefore not result in strong interaction between these two helices. Third, the helix within the Ald13 tag is only separated from the adjacent helix by 1 turn residue; most likely, it sterically requires too much strain to be able to interact strongly. These 3 factors resulting from CFSSP analysis, in conjunction with the prediction from SOPMA, lead to the hypothesis that the secondary structure should not lead to tertiary structures which render the TM domain unable to integrate into a phospholipid bilayer. In addition, even though epitope tags are hypothesized to not impact protein structure, a study shows that the tag resulted in a destabilization of the tagged protein, leading to reduced protein activity in the cell [51]. The results of the study also show that the instability arises from the spacer in between the protein and the added tag [51]. The study differs in that the protein used originates from *Saccharomyces cerevisiae*, and is tagged at

the C-terminal [51]. Although the studied protein differs, it is hypothesized that HA likely does not causes an overexpression of Wu et al.'s gene [7] that allows it to be detected where Alfred is not. The hypotheses regarding any changes that the presence or absence of the HA epitope can have on the secondary structure of Alfred can be tested with experimental procedures (discussed in Section 5.3.).

In addition, it is hypothesized that the presence of the tag is unlikely to serve an additional purpose for binding of the recognition site to FGE. The cocrystal structure of FGE-LCPTSRA has shown that FGE contains a shallow binding pocket that can accommodate up to 6 amino acids in the consensus recognition sequence [52], and therefore it likely does not need extra residues to help in holding the polypeptide. An example where extra residues are required would be similar to restriction enzymes requiring extra nucleotides adjacent to its cut site for efficient digestion [53]. However, this is likely not the case for FGE.

Interestingly, the Myc epitope tag may also not be completely reliable in this situation. A recent study shows that the detection of the Myc tag by the monoclonal anti-Myc antibody 9E10 (the antibody used in this study) is highly variable depending on the amino acid sequence adjacent to the tag [46]. However, since the Myc tag has shown itself to be reliable in other situations, so this was not considered when designing the Alfred gene. Since the amino acid sequence and not the nucleotide sequence is in question in this situation, the variability of the Myc antibody should be a non-issue; the two amino acids separating the aldehyde tag from the Myc epitope in Wu et al.'s gene [7], which are not present in Alfred, should not significantly affect the protein.

5.2.2. Probe choice

Wu et al. chose to use aminooxy Alexa Fluor 647 to detect their protein [7], whereas MMNA was chosen to detect Alfred. In this project, MMNA was chosen in order to ensure the reaction between *N*-aminoanthranilates and Alfred. This would open the doorway for future works, where Alfred-expressing tumours in living systems could be detected by similar probes for CEST-MRI or PET imaging [20,21].

Though the reaction is similar, the oxime and hydrazone products are slightly different in that the former contains an oxygen where the latter contains a nitrogen adjacent to the nitrogen that had acted as a nucleophile (i.e. is an α-nucleophile). In particular, the oxime is more stable than the hydrazone [54]. In both molecules, the nitrogen that had previously acted as a nucleophile can be protonated, and the adjacent carbon is highly electrophilic in the structure [54] (Figure 5.2.1.). This combination results in a higher possibility of hydrolysis [54].

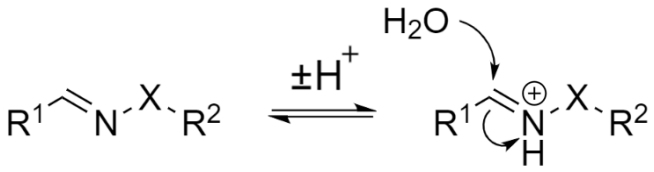

Figure 5.2.1. Protonation and hydrolysis of oximes and hydrazones. Protonation of the nitrogen results in a highly electrophilic adjacent carbon, allowing hydrolysis of the pi bond. X: oxygen (oximes) or nitrogen (hydazones).

The adjacent atom's (X as depicted in Figure 5.2.1.) pull on electrons will reduce the electron density around the nucleophilic nitrogen, increasing its likelihood of being protonated. Therefore, hydrolysis is less likely to occur in the oxime because the adjacent atom is oxygen, which is more electronegative than the nitrogen in the hydrazone [55]. As a result, aminooxy Alexa Fluor 647 may have a higher stability when bound to formylglycine than MMNA.

However, this issue was addressed in the design of MMNA, and the stability of MMNA should be sufficient for imaging purposes [14].

Another difference between the probes is that aminooxy Alexa Fluor 647 is highly polar, and therefore impermeant to the cell membrane [55]. By contrast, MMNA is permeable, which contributes to its ability to identify the Aldehydic Load of cells [14]. While the original purpose of Alfred was to act as a reporter gene to explore the reaction of Alfred with MMNA, aminooxy Alexa Fluor 647 may allow better imaging of cell-surface aldehydes by eliminating noise from detection of aldehydes in the cells.

5.2.3. Cell line

Wu et al. had used Chinese hamster ovary (CHO) cells and human embryonic kidney (HEK293) cells [7], whereas the cell line used in this study was lung cancer (A549) cells. A549 cells were originally chosen for this work as it is a cancer initiating cell line [56], which would allow for tumour growth in a living system for tumour imaging with CEST-MRI or PET tracers [20,21]. However, presence of aldehydes has previously been investigated as a potential diagnostic marker for lung cancer [57], and this may correlate with an increase in background signal in the detection of aldehydes. On the contrary, CHO and HEK293 are not cancer cell lines, and therefore may have a lower background signal arising from a lower Aldehydic Load. These cell lines may show the clearer and brighter signal arising from Wu et al.'s construct [7] where Alfred does not.

5.2.4. Analysis of results

In Wu et al.'s study, the flow cytometry results are represented in mean fluorescence units, measured in arbitrary units and compared as a ratio between the controls and the sample [7]. The sample consists of cells with a TM protein containing the aldehyde tag (TM-Ald13), and the controls are cells containing an unaltered TM protein (uTM) or a mutated TM-Ald13 with a C18A mutation that removes the cysteine of the aldehyde tag (TM-Ald13-C18A) [7]. Although the bar graph results show that TM-Ald13 tagging results in more than double the fluorescence exhibited by uTM and TM-Ald13-C18A tagging, data obtained from flow cytometry is usually displayed in a scatterplot graph, where points on the graph indicate the number of cells exhibiting levels of fluorescence intensity [7]. Therefore, it is unclear what the results are showing and how they were obtained. As a result, it is difficult to determine the difference in fluorescence intensity levels that should be expected. This finding also does not necessarily correlate with the microscopy images shown in their study. The cells with TM-Ald13 showed some fluorescence all along the cell with some brighter foci and some less bright foci [7]. The fluorescent foci were understood as detection of the aldehyde tag by an aldehyde probe, and the lighter fluorescence was understood as noise arising from regular cell-surface aldehydes (ex. acetylated cell surface saccharides) [7]. The signal presents as punctate foci because endocystosis may cause the cell surface-tagged protein to internalize as vesicles. This logic was applied similarly in this study. However, the cells containing uTM also showed a few brighter foci that may be comparable to some of the foci seen in the cells containing TM-Ald13 [7]. The signal from the aldehyde probe was greatly reduced in general as well, so that barely any signal is seen [7]. This does not allow the cells containing TM and TM-Ald13 to be compared, since the "background" is displayed at different levels. Therefore, the expected difference between the

fluorescence from uTM and TM-Ald13 tagging is unclear, and therefore may not be applicable to this study. Furthermore, it is not clear whether a single foci indicates the presence of a single TM-Ald13 protein, a dimer, or TM-Ald13 aggregates on the cell surface. This question arises from the presence of brighter and dimmer foci [7]; perhaps this implies that the brighter foci arise from tagging of multiple TM-Ald13, while the lighter foci are a result of fewer TM-Ald13. PDGFRβ aggregation has previously been observed in human skin fibroblasts [58], and dimerization of PDGFRβ is common [51]. Dimerization is due to interactions of the immunoglobulin-like domains in the extracellular domain of PDGFRβ, which are not present in TM-Ald13, and therefore the question of brighter and dimmer foci remains unanswered. Furthermore, a previous study visualized the PDGFR TM domain localization as fluorescence around the perimeter of the cell [59]. In that study, biotin ligase was used to bind a biotin analogue containing a ketone to a lysine within an accepter peptide [59]. This peptide was attached alongside cyan fluorescent protein to the extracellular N-terminal of the PDGFR TM domain, and then expressed in HeLa cells [59]. The ketone was reacted with biotin-hydrazide, and then tagged with streptavidin-Alexa Fluor 568, which is also impermeant to the cell membrane [55,59]. The application of the PDGFR TM domain to display a tag on the extracellular face of the membrane was similar between this study and Wu et al. [7]. However, the difference in fluorescence localization (i.e. punctate foci [7] *vs.* labelling at the periphery of the cell [59]) may indicate that further cellular processes are affecting transmembrane protein localization, warranting further investigation.

5.3. Future directives

Future experiments could improve the detection of Alfred or investigate the lack of Alfred expression. For example, reverse transcriptase PCR (RT-PCR) can be conducted to assess the hypothesis that the Alfred RNA transcript is degraded prior to translation. A positive result would show that Alfred is expressed and processed into mRNA but cannot dock the translation initiation factors in translation initiation, or that the mRNA is degraded prior to a significant amount of translation. A negative result would show that Alfred pre-mRNA is not processed into mRNA, or it is not expressed by its promoters. For the latter case, if the vector used was pCMV-Alfred, then the CMV promoter may be less expressed than other CMV promoters in the cells; however, this is unlikely, as the fluorescent foci in cells transfected with only pCMV-Alfred (results not shown) did not appear different from co-transfected cells; in other words, there is no visual difference in the frequency and brightness of the fluorescent foci. If the vector used was pTRE3G-ZsGreen-Alfred in the Tet-On® Inducible System, then transcription would have stopped prior to reaching the Alfred gene, which is downstream of the ZsGreen gene that was expressed and visualized in microscopy. To fix this problem, the order of ZsGreen and Alfred in pTRE3G-ZsGreen1-Alfred can be switched. This would also solve the issue of IRES-dependent translation reducing expression of the second gene, which is Alfred in this case. Expression of ZsGreen would be indicative of the level of translation reduction, and it can still allow identification of transfected cells; simultaneously, Alfred expression would be increased, and would at least ensure that transcription of Alfred had occurred. This had not been tested in the present work, as Alfred expression had been expressed constitutively to increase expression and eliminate the need to troubleshoot other aspects of the Tet-On® Inducible System. However, this may be an alternative to co-transfection with a second vector indicative of general

transfection efficiency (ex. mCardinal-N1), as it will continue to allow identification of transfected cells rather than areas where transfection may be increased.

The signal from Alfred may be ameliorated by using a different aldehyde probe. An alternative for MMNA could be aminooxy Alexa Fluor 647, which could be used as a positive control to ensure the aldehyde in Alfred can be detected, since it was shown to have detected Wu et al.'s gene [7], and has also been used in detecting formylglycine generated by FGE [60]. It can also reduce background signal resulting from Aldehydic Load within the cell, because aminooxy Alexa Fluor 647 does not enter the cell [55]. If the background signal was not present, the possible presence of fainter signals arising from low levels of Alfred expression may be distinguishable. In addition, assuming that Alfred was expressed and detected in RT-PCR, and that Western blotting showed translation of Alfred, failure to show punctate foci would indicate that punctate foci are not indicative of the presence of Alfred. Instead, cell surface probes may be seen as a brighter dotted outline on the cell's perimeter [61]. Once Alfred is optimized by the use of a cell surface probe, Alfred can be tested again with MMNA, which would allow applications with structurally similar probes that act as CEST-MRI transfer agents and PET tracers.

The experiments can also be conducted with a non-cancer cell line, like CHO or HEK293 to reduce the Aldehydic Load of the cells and also reduce the background signal, since increased aldehydes is associated with cancer [56]. In addition, these more widely studied cell lines have been shown to abundantly express the CMV promoter [36], and notably the IgG κ signal peptide has been optimized in CHO cells [62]. Importantly, FGE should be overexpressed in these cell lines, because CHO and HEK293 cells have lower endogenous FGE content than A549 cells [9]. HeLa cells are also a cancerous cell line, but it is also well-studied. These cells also express a

similar level of FGE compared to A549 cells, and therefore may be a good comparison for troubleshooting Alfred expression and detection [9].

In addition, the HA epitope and MCS residues may be included in Alfred. As previously discussed, the epitope was hypothesized to not be required as extra residues for efficient FGE activity, and to not affect the protein's tertiary structure or its function. However, it is possible that unspecific cleavage of the IgG κ signal peptide results in an elongated or truncated N-terminal [62]; that is, after cleavage of the IgG κ leader sequence, the cleavage site may be shifted so that the FGE recognition site is not directly at the N-terminal of Alfred, or it is cleaved when the signal peptide is cleaved. Therefore, adding the HA epitope and the MCS residues as a buffer between the IgG κ signal peptide and the FGE recognition sequence may decrease unspecific signal peptide cleavage resulting in a truncated protein and losing the aldehyde tag. Also, these residues and the MCS residues that were present in Wu et al.'s construct [7] may play a role in allowing FGE binding to the recognition site. Since the recognition site is very close to the membrane introduction of these 2 residues into Alfred may increase the thiol-to-aldehyde conversion by allowing FGE to bind to the recognition site, where previously the enzyme may not be able to reach close to the membrane. In addition, the HA tag may be an alternative epitope for Alfred identification in Western blotting, since the Myc tag had not identified the presence of Alfred. Tagging with a reliable epitope can also be beneficial for visualizing localization of Alfred over time using immunofluorescence microscopy, which could act as a guide for where to expect tagging of Alfred via aldehyde probes. Furthermore, concerns of the secondary structure affecting the tertiary protein structure and Alfred's ability to integrate into the phospholipid bilayer can be addressed by determining the secondary structure experimentally rather than prediction through software. Many methods exist to estimate the

secondary structure, such as circular dichroism, infrared, and Raman spectroscopy, as well as nuclear magnetic resonance and X-ray crystallography [63,64]. Using the experimental methods detailed in this report, a reasonable method to choose would be circular dichroism spectroscopy, since it does not require a high amount of sample [65] and Alfred expression may be very low. In addition, the technique is relatively simple, and the data it produces is easy to quickly analyze [63]. Analysis of the secondary structure may also be a simpler alternative compared to determining the tertiary structure, as the latter requires nuclear magnetic resonance or X-ray crystallography [66,67].

In summary, RT-PCR could be used in troubleshooting to identify the step in the central dogma where Alfred fails to be expressed and allow for better gene design. The signal from Alfred may be ameliorated by using a different probe, such as aminooxy Alexa Fluor 647, to reduce background signal from cellular Aldehydic Load. A different cell line can be used for testing Alfred expression to address the possibility that Alfred is incompatible with A549 cells, and increase Alfred expression using previously established findings to optimize the gene design. The gene design may also be improved by adding the HA epitope, to mimic a similar gene in previous studies, to prevent truncation of the FGE recognition site, and to allow immunofluorescence microscopy to verify localization of Alfred.

6. Conclusion

Alfred was designed to be a reporter gene displaying aldehydes on the cell surface. This gene contained the IgG κ leader sequence, the formylglycine generating enzyme (FGE) substrate sequence, a Myc epitope tag, and the platelet derived growth factor receptor β transmembrane (PDGFRβ TM) domain (Figure 1.5.1.). Notably, the FGE substrate sequence allows the generation of an aldehyde from a formylglycine generating enzyme (FGE) recognition site, which allows a critical cysteine residue to be converted into a formylglycine reside; effectively, this is a thiol-to-aldehyde conversion (Figure 1.2.1.). This gene was ligated into the Tet-On® Inducible System pTRE3G-ZsGreen1 after the IRES, into the MCS (Figure 1.4.1., 1.5.1., 3.1.1.). Gene expression was induced in the presence of doxycyclin (dox) by the TET3G transactivator protein, which is expressed constitutively by pCMV-TET3G (Figure 1.4.2.). The Tet-On® Inducible System was used to express Alfred in A549 cells. After 24 hours of induction by dox, the cells were imaged and probed with methyl 5-MeO-*N*-aminoanthanilate (MMNA) for fluorescence microscopy imaging. The positive and negative controls established the fidelity of the Tet-On® Inducible System, and also established imaging parameters for subsequent experiments (Figure 4.1.2.). A test for fluorescent resonance energy transfer (FRET) was conducted to ensure that fluorescence from GFP, which has similar excitation and emission curves to ZsGreen, does not contribute to fluorescence from MMNA (Figure 4.1.3.). Transfected cells, indicated by green fluorescence originating from ZsGreen, did not show an increased amount of MMNA tagging, indicated by blue fluorescence, and cells that exhibited brighter MMNA fluorescence were not limited to cells that fluoresced green (Figure 4.1.4.). Instead some cells were observed to show MMNA fluorescence in the perinuclear region inside the cell, and

thus it was hypothesized that there was insufficient time for Alfred to localize to the cell membrane. The A549-TET3G cell line was generated to increase the proportion of cells that were transfected (Figure 4.2.2., 4.2.3.). The images showed no visual difference in MMNA fluorescence localization between Alfred-transfected cells and non-transfected cells. Notably, the cells did exhibit punctate fluorescent foci (Figure 4.2.3.), which were postulated to indicate presence of Alfred in the cell membrane, but the fluorescent foci were present in both Alfred-transfected cells and non-transfected cells. A time-course experiment was conducted to allow for a longer incubation time, and the cells were imaged at 48 and 72 hours (Figure 4.2.4.). Still, the cells did not show differences between Alfred-transfected cells and non-transfected cells, so cells were additionally transfected with FGE to increase the production of aldehydes available for MMNA detection (Figure 4.2.5., 4.2.6.). Overexpression of FGE continued to show no difference in fluorescence of MMNA between Alfred-transfected and non-transfected cells. Alfred was constitutively expressed to increase accumulation of the protein, but this also did not change fluorescence patterns in transfected cells as compared to non-transfected cells (Figure 4.3.3., 4.3.5., 4.3.6.). The expression of Alfred was verified using the Myc tag *via* Western blot; however the lack of bands present demonstrated the absence of Alfred (Figure 4.4.1.). Finally, the secondary structure of Alfred was predicted using online software to show that the secondary structure should not have affected integration of Alfred into the phospholipid bilayer (Figure 4.5.1., 4.5.2., 4.5.3., 4.5.4.).

Since the Alfred design was modeled after Wu et al.'s construct [7], the differences between these studies were identified as potential sources of problems in this work. First, the HA epitope was present in Wu et al.'s construct [7], but not in Alfred (Section 5.2.1.). This was hypothesized to not be an issue, since epitopes generally do not interfere with protein structure

and function. In addition, the predicted secondary structure does not show significant structures that may affect protein folding, and by extension, presence of absence of the HA epitope should not affect the integration of Alfred into the phospholipid bilayer. Second, Wu et al.'s work used aminooxy Alexa Fluor 647 [7], which is a hydroxylamine, as the aldehyde probe as opposed to MMNA, which is a hydrazine (Section 5.2.2.). This probe does not enter cells. Therefore, the probe does not detect the Aldehydic Load within the cells, which may reduce background fluorescence and allow better detection of the aldehyde tag in Alfred. Furthermore, previous studies have shown that the oxime product resulting from the hydroxylamine-aldehyde reaction is more stable than the hydrazone product resulting from the hydrazine-aldehyde reaction. This should not be an issue, since this issue was previously addressed in the design of MMNA. Also, microscopy with MMNA is conducted within the suggested timeframe to prevent a decrease in the signal resulting from hydrolysis of the probe. Third, Wu et al.'s study used Chinese hamster ovary (CHO) cells and human embryonic kidney (HEK293) cells [7], which are not a cancer cell line unlike the A549 lung cancer cell line used in this study (Section 5.2.3.). However, increased aldehydes have previously been associated with lung cancer, and this may have resulted in increased background fluorescence signal from MMNA. Fourth, the punctate fluorescent foci may not truly indicate presence of Alfred. The foci were present in aldehyde-tagged cells and also in non-tagged cells at much lower levels. In addition, the background fluorescence is not comparable, and therefore the method of image analysis may not be indicative of the aldehyde tag on the cell surface.

To resolve the issues identified in this study, some future directives were postulated (Section 5.3.). The interruption of the Alfred gene-to-protein pathway may be identified by RT-PCR. This may show that the mRNA transcript is not present in significant amounts,

showing mRNA degradation. RT-PCR also may show that the transcript is present by expression with the Tet-On® Inducible System. In this case, the locus of ZsGreen and Alfred may be switched, to avoid the reduction of protein expression after the IRES. Other studies may use aminooxy Alexa Fluor 647 or a non-cancer cell line to reduce background signal arising from detection of endogenous aldehydes. This will yield a clearer signal arising from Alfred, and may also identify whether punctate fluorescent foci are indicative of the localization of Alfred. Finally, the HA epitope may be incorporated into Alfred, to eliminate the possibility that the aldehyde tag is removed from the protein by cleavage. The epitope may also serve as an additional method of tagging Alfred, since the Myc tag was unable to identify Alfred. Concerns regarding the secondary structure and interference of proper integration of the PDGFRβ TM domain into the phospholipid bilayer can be addressed with experimental characterization of the secondary structure using methods such as circular dichroism spectroscopy.

This study was unable to create a reporter gene displaying an aldehyde on the cell surface. However, this project warrants future investigation, as this is an invisible protein prior to tagging, allowing for detection by aldehyde probes with a variety of usages such as fluorescence, CEST-MRI, and PET-based imaging. Reporter genes are also crucial in experiments as they can serve as positive or negative controls. Besides imaging, it can also allow creation and optimization of novel aldehyde probes. The importance of reporter genes should not be underestimated, and their development will continue to evolve alongside the development of molecular biology and its techniques.

7. References

1. Moriya, T., Yamaoka, T., Wakayama, Y., Ayukawa, S., Zhang, Z., Yamamura, M., Wakao, S., Kiga, D. Comparison between Effects of Retroactivity and Resource Competition upon Change in Downstream Reporter Genes of Synthetic Genetic Circuits. *Life (Basal).* 2019, **9**(1), 30.

2. Zhang, J., Campbell, R.E., Ting, A.Y., Tsien, R.Y. Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Cell Biol.* 2002, **3**(12), 906-918.

3. Tyagi, S. Imaging intracellular RNA distribution and dynamics in living cells. *Nat. Methods.* 2009, **6**(5), 331-338.

4. Nordgren, I.K., Tavassoli, A. A bidirectional fluorescent two-hybrid system for monitoring protein-protein interactions. *Mol. BioSyst.* 2014, **10**(3), 485-490.

5. Tet-On® 3G Inducible Expression Systems User Manual. *Clontech® Laboratories, Inc.* 2014. *Manual*.

6. Terskikh, A., Fradkov, A., Ermakova, G., Zaraisky, A., Tan, P., Kajava, A.V., Zhao, X., Lukyanov, S., Matz, M., Kim, S., Weissman, I., Siebert, P. "Fluorescent Timer": Protein That Changes Color with Time. *Science.* 2000, **290**(5496), 1585-1588.

7. Wu, P., Shui, W., Carlson, B.L., Hu, N., Rabuka, D., Lee, J., Bertozzi, C.R. Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. *Proc. Natl. Acad. Sci. U. S. A.* 2008, **106**(9), 3000-3005.

8. Barfield, R.M., Rabuka, D. Noncanonical Amino Acids Methods and Protocols. *Humana Press*, New York, NY. 2018. *Book.*

9. SUMF1 Gene. GeneCards. Website.

10. Tuley, A., Lee, Y-J., Wu, B., Wang, Z., Liu, W.R. A Genetically Encoded Aldehyde for Rapid Protein Labelling. *Chem. Commun. (Camb).* 2014, **50**(56), 7424-7426.

11. Minnihan, E.C., Yokoyama, K., Stubbe, J. Unnatural amino acids: better than the real things? *F1000 Biol. Rep.* 2009, **1**, 88.

12. Lodish. H., Berk, A., Zipursky, S.L., et al. Section 17.3 Overview of the Secretory Pathway. *Molecular Cell Biology. 4th edition*. W.H. Freeman, New York. 2000. *Book*.

13. ChemTube3D. Oxime formation. *University of Liverpool. Website*.

14. Suchy, M., Lazurko, C., Kirby, A., Dang, T., Liu, G., Shuhendler, A.J. Methyl 5-MeO-*N*-aminoanthanilate, a minimalist fluorogenic probe for sensing cellular aldehydic load. *Org. Biomol. Chem*. 2019, **17**, 1843-1853.

15. Lazurko, C., Radonjic, I., Suchy, M., Liu, G., Rolland-Lagan, A-G., Shuhendler, A. Fingerprinting Biogenic Aldehydes through Pattern Recognition Analyses of Excitation-Emission Matrices. *ChemBioChem*. 2019, **20**(4), 543-554.

16. TetOff® and Tet-On® Gene Expression Systems User Manual. *Clontech Laboratories, Inc*. 2012. *Manual*.

17. Pacheco, A., Martinez-Salas, E. Insights into the Biology of IRES Elements through Riboproteomic Approaches. *J. Biomed Biotechnol*. 2010. doi:10.1155/2010/458927

18. Martinez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J., Embarek, A.M. Insights into Structural and Mechanistic Features of Viral IRES Elements. *Front. Microbiol.* 2018, **8**, 2629.

19. Witherell, G.W., Wimmer, E. Encephalomyocarditis Virus Internal Ribosomal Entry Site RNA-Protein Interactions. *J. Virol.* 1994, **68**(5), 3183-3192.

20. Dang, T., Suchy, M., Truong, Y.J., Oakden, W., Lam, W.W., Lazurko, C., Facey, G., Stanisz, G.J., Shuhendler, A.J. Hydrazo-CEST: Hydrazone-Dependent Chemical Exchange Saturation Transfer Magnetic Resonance Imaging Contrast Agents. *Chemistry*. 2018, **24**(36), 9148-9156.

21. Kirby, A., Suchy, M., Brouwer, A., Shuhendler, A. Mapping aldehydic load in vivo by positron emission tomography with [^{18}F]NA_3BF_3. *Chem. Commun.* 2019, **55**, 5371-5374.

22. Useful Numbers for Cell Culture. *ThermoFisher Scientific*. *Website.*

23. Ennemann, E.C., Radhakrishnan, K., Mariappan, M., Wachs, M., Pringle, T.H., Schmidt, B., Dierks, T. Proprotein Convertases Process and Thereby Inactivate Formylglycine-generating Enzyme. *J. Biol. Chem.* 2013, **288**(8), 5828-5839.

24. Riba, A., Di Nanni, N., Mittal, N., Arhné, E., Schmidt, A., Zavolan, M. Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proc. Natl. Acad. Sci. U.S.A.* 2019, **116**(30, 15023-15032.

25. Alberts, B., Johnson, A., Lewis, J., Raff. M., Roberts, K., Walter, P. From RNA to Protein. *Molecular Biology of the Cell. 4th edition*. Garland Science, New York. 2002. *Book.*

26. Keating, M.T., Williams, L.T. Processing of the Platelet-derived Growth Factor Receptor. Biosynthetic and degredation studies using anti-receptor antibodies. *J. Biol. Chem.* 1987, **262**(16), 7932-7937.

27. Renner, O., Tsimpas, A., Kostin, S., Valable, S., Petit, E., Schaper, W., Marti, H.H. Time- and cell type-specific induction of platelet-derived growth factor receptor-β during cerebral ischemia. *Brain Res. Mol.* 2003, **113**(1-2), 44-51.

28. Wang, X., Liu, H., Yuan, W., Cheng, Y., Han, W. Efficient production of CYTL1 protein using mouse IgGκ signal peptide in the CHO cell expression system. *Acta. Biochim. Biophys. Sin. (Shanghai).* 2016, **28**(4), 391-394.

29. Qiu, X., Zhu, X., Zhang, L., Mao, Y., Zhang, J., Peng, H., Li, G., Lv, P., Li, Z., Sun, X., Wu, L., Zheng, J., Deng, Y., Hou, C., Tang, P., Zhang, S., Zhang, Y. Human Epithelial Cancers Secrete Immunoglobulin G with Unidentified Specificity to Promote Growth and Survival of Tumor Cells. *Cancer Res.* 2003, **63**(19), 6488-6495.

30. Huang, F., Yao, Y., Wu, J., Zhang, Y., Yu, L., Pu, X., Xia, L. Blocking platelet-derived growth factor-d inhibits human lung cancer progression. *Int. J. Clin. Exp. Pathol.* 2017, **10**(3), 3109-3115.

31. Wang, P., Song, L., Ge, H., Jin, P., Jiang, Y., Hu, W., Geng, N. Crenolinab, a PDGFR inhibitor, suppresses lung cancer cell proliferation and inhibits tumor growth in vivo. *Onco. Targets Ther.* 2014, **7**, 1761-1768.

32. Lodish. H., Berk, A., Zipursky, S.L., et al. Section 17.5 Insertion of Membrane Proteins into the ER Membrane. *Molecular Cell Biology. 4th edition*. W.H. Freeman, New York. 2000. *Book.*

33. Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E., Hayakawa, T. IRES-Dependent Second Gene Expression Is Significantly Lower Than Cap-Dependent First Gene Expression in a Bicistronic Vector. *Mol. Ther.* 2000, **1**(4), 376-382.

34. Pilhofer, M., Ladinsky, M.S., McDowall, A.W., Petroni, G., Jensen, G.J. Microtubules in *Bacteria*: Ancient Tubulins Build a Five-Protofilament Homolog of the Eukaryotic Cytoskeleton. *PLoS Biol.* 2011, **9**(12).

35. Aneja, M.K., Geiger, J., Imker, R., Uzgun, S., Kormann, M., Hasenpusch, G., Maucksch, C., Rudolph, C. Optimization of *Streptomyces* bacteriophage φC31 integrase system to prevent post integrative gene silencing in pulmonary type II cells. *Exp. Mol. Med.* 2009, **41**(12), 919-934.

36. Raup, A., Jérôme, V., Freitag, R., Synatschke, C.V., Müller, A.H. Promoter, transgene, and cell line effects in the transfection of mammalian cells using PDMAEMA-based nano-stars. *Biotechnol. Rep. (Amst).* 2016, **11**, 53-61.

37. Chanfreau, G.F. RNA Modification. *Academic Press,* Cambridge, MA. 2017. *Book.*

38. Geissler, R., Simkin, A., Floss, D., Patel, R., Fogarty, E.A., Scheller, J., Grimson, A. A widespread sequence-specific mRNA decay pathway mediated by hnRNPs A_1 and A_2/B_1. *Genes. Dev.* 2016, **30**(9), 1070-1085.

39. Luo, G.-Z., Wang, F., Weng, X., Chen, K., Hao, Z., Yu, M., Deng, X., Liu, J., He, C. Characterization of eukaryotic DNA N^6-methyladenine by a highly sensitive restriction enzyme-assisted sequencing. *Nat. Commun.* 2016, **7**, 11301.

40. Schrader, E.K., Harstad, K.G., Matouschek, A. Targeting proteins for degradation. *Nat. Chem. Biol.* 2009, **5**(11), 815-822.

41. Geourjon, C., Deléage, G. SOPMA: Significant Improvements in Protein Secondary Structure Prediction by Consensus Prediction From Multiple Alignments. *Comput. Appl. Biosci.* 1995, **11**(6), 681-684.

42. pDisplay™ Mammalian Expression Vector. *ThermoFisher Scientific. Website*.

43. Epitope Tags in Protein Research. Tag Selection & Immunotechniques. *Sigma Life Science.* 2012. *Manual.*

44. Raymond, D.D., Bajic, G., Ferdman, J., Suphaphiphat, P., Settembre, E.C., Moody, M.M., Schmidt, A.G., Harrison, S.C. Conserved epitope on influenza-virus hemagglutinin head defined by a vaccine-induced antibody. *Proc. Natl. Acad. Sci. U.S.A.* 2018, **115**(1), 168-173.

45. Miller, D.M., Thomas, S.D., Islam, A., Muench, D., Sedoris, K. c-Myc and Cancer Metabolism. *Clin. Cancer Res.* 2012, **18**(20), 5546-5553.

46. Schüchner, S., Behm, C., Mudrak, I., Ogris, E. The Myc tag monoclonal antibody 9E10 displays highly variable epitope recognition dependent on neighboring sequence context. *Sci. Signal*. 2020, **13**(616).

47. Chen, P.H., Chen, X., He, X. Platelet-derived growth factors and their receptors: structural and functional perspectives. *Biochim. Biophys. Acta.* 2013, **1834**(10, 2176-2186.

48. Counterman Burba, A. E., Lehnert, U., Yu, E.Z., Cerstein, M. Helix Interaction Tool (HIT): a web-based tool for analysis of helix-helix interactions in proteins. *Bioinformatics*. 2006, **22**(22), 2735-2738.

49. Ashok Kumar, T. CFSSP: Chou and Fasman Secondary Structure Prediction server. WIDE SPECTRUM: Research Journal. 2013, **1**(9), 15-19.

50. Chen, P.H., Unger, V., He, X. Structure of full-length human PDGFRβ bound to its activating ligand PDGF-B as determined by negative-stain electron microscopy. *J. Mol. Bio.* 2015, **427**(24), 3921-3934.

51. Saiz-Baggetto, S., Méndez, E., Quilis, I., Igual, J.C., Bañó, M.C. Chimeric proteins tagged with specific 3xHA cassettes may present instability and functional problems. *PLoS One*. 2017, **12**(8).

52. Appel, M.J., Bertozzi, C.R. Formylglycine, a Post-Translationally Generated Residue with Unique Catalytic Capabilities and Biotechnology Applications. *ACS Chem. Biol.* 2015, **10**(1), 72-84.

53. Primer Design for Restriction Enzyme Cloning (E6901). *New England BioLabs. Inc. Website.*

54. Kalia, J., Raines, R.T. Hydrolytic Stability of Hydrazones and Oximes. *Angew. Chem. Int. Ed. Engl.* 2008, **47**(39), 7523-7526.

55. The Molecular Probes Handbook. *ThermoFisher Scientific*. *Manual.*

56. Meng, X., Wang, X., Wang, Y. More than 45% of A549 and H446 cells are cancer initiating cells: Evidence from cloning and tumorigenic analyses. *Oncol. Rep.* 2009, **21**(4), 995-1000.

57. Fuchs, P., Loeseken, C., Schubert, J.K., Miekisch, W. Breath gas aldehydes as biomarkers of lung cancer. *Int. J. Cancer.* 2010, **126**(11), 2663-2670.

58. Wiseman, P.W., Höddelius, P., Petersen, N.O., Magnusson, K.-E. Aggregation of PDGF-β receptors in human skin fibroblasts: characterization by image correlation spectroscopy (ICS). *FEBS Lett.* 1997, **401**(1), 43-48.

59. Chen, I., Howarth, M., Lin, W., Ting, A.Y. Site-specific labelling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods.* 2005, **2**, 99-104.

60. Rush, J.S., Bertozzi, C.R. New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes *in Vitro and in Vivo*. *J. Am. Chem. Soc.* 2008, **130**(37), 12240-12241.

61. Quinn, D.J., McFerran, N.V., Nelson, J., Duprex, W.P. Live-cell visualization of transmembrane protein oligomerization and membrane fusion using two-fragment haptoEGFP methodology. *Biosci. Rep*. 2012, **32**(3), 333-343.

62. Haryadi, R., Ho, S., Kok, Y.J., Pu, H.X., Zheng, L., Pereira, N.A., Li, B., Bi, X., Goh, L.-T., Yang, Y., Song, Z. Optimization of Heavy Chain and Light Chain Signal Peptides for High Level Expression of Therapeutic Antibodies in CHO Cells. *PLoS One*. 2015, **10**(2).

63. Pelton, J. T., McLean, L.R. Spectroscopic Methods for Analysis of Protein Secondary Structure. *Anal. Biochem.* 2000, **277**(2), 167-176.

64. Greenfield, N.J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* 2006, **1**(6), 2876-2890.

65. Kumagai, P.S., DeMarco, R., Lopes, J.L.S. Advantages of Synchrotron Radiation Circular Dichroism Spectroscopy to Study Intrinsically Disordered Proteins. *Eur. Biophys. J.* 2017, **46**(7), 599-606.

66. Cavalli, A., Salvatella, X., Dobson, C.M., Vendruscolo, M. Protein structure determination from NMR chemical shifts. *PNAS*. 2007, **104**(23), 9615-9620.

67. Smyth, M.S., Martin, J.H.J. x Ray crystallography. *Mol. Pathol.* 2000, **53**(1), 8-14.

Appendix A. pTRE3G-ZsGreen1-Alfred plasmid sequence

CTCGAGTTTACTCCCTATCAGTGATAGAGAACGTATGAAGAGTTTACTCCCTA
TCAGTGATAGAGAACGTATGCAGACTTTACTCCCTATCAGTGATAGAGAACGTATAA
GGAGTTTACTCCCTATCAGTGATAGAGAACGTATGACCAGTTTACTCCCTATCAGTG
ATAGAGAACGTATCTACAGTTTACTCCCTATCAGTGATAGAGAACGTATATCCAGTT
TACTCCCTATCAGTGATAGAGAACGTATAAGCTTTAGGCGTGTACGGTGGGCGCCTA
TAAAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGCAATTCCACAACACT
TTTGTCTTATACCAACTTTCCGTACCACTTCCTACCCTCGTAAAGTCGACATGGCTCA
GTCAAAGCACGGTCTAACAAAAGAAATGACAATGAAATACCGTATGGAAGGGTGCG
TCGATGGACATAAATTTGTGATCACGGGAGAGGGCATTGGATATCCGTTCAAAGGG
AAACAGGCTATTAATCTGTGTGTGGTCGAAGGTGGACCATTGCCATTTGCCGAAGAC
ATATTGTCAGCTGCCTTTATGTACGGAAACAGGGTTTTCACTGAATATCCTCAAGAC
ATAGCTGACTATTTCAAGAACTCGTGTCCTGCTGGTTATACATGGGACAGGTCTTTTC
TCTTTGAGGATGGAGCAGTTTGCATATGTAATGCAGATATAACAGTGAGTGTTGAAG
AAAACTGCATGTATCATGAGTCCAAATTTTATGGAGTGAATTTTCCTGCTGATGGAC
CTGTGATGAAAAAGATGACAGATAACTGGGAGCCATCCTGCGAGAAGATCATACCA
GTACCTAAGCAGGGGATATTGAAAGGGGATGTCTCCATGTACCTCCTTCTGAAGGAT
GGTGGGCGTTTACGGTGCCAATTCGACACAGTTTACAAAGCAAAGTCTGTGCCAAG
AAAGATGCCGGACTGGCACTTCATCCAGCATAAGCTCACCCGTGAAGACCGCAGCG
ATGCTAAGAATCAGAAATGGCATCTGACAGAACATGCTATTGCATCCGGATCTGCAT
TGCCCTGACGGCCGGCCCCTCTCCCTCCCCCCCCCCCTAACGTTACTGGCCGAAGCC
GCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACCATATTGCCGTC

TTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTCCTAG

GGGTCTTTCCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGC

AGTTCCTCTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCA

GCGGAACCCCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAG

ATACACCTGCAAAGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGTG

GAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCA

GAAGGTACCCCATTGTATGGGATCTGATCTGGGGCCTCGGTACACATGCTTTACATG

TGTTTAGTCGAGGTTAAAAAAACGTCTAGGCCCCCCGAACCACGGGGACGTGGTTTT

CCTTTGAAAAACACGATGATAATATGGCCACAACCGGGCCGGATATCACGCGTATG

GAAACCGACACACTGCTGCTGTGGGTGCTGCTTCTTTGGGTGCCCGGATCTACCGGC

GATCTGTGTACACCTTCTAGAGCCGCTCTGCTGACCGGCAGAGAGCAGAAGCTGATC

TCCGAAGAGGACCTGAATGCCGTCGGCCAGGACACCCAAGAAGTGATCGTCGTCCC

TCACAGCCTGCCTTTCAAGGTGGTGGTCATCAGCGCCATTCTGGCCCTGGTGGTGCT

GACCATCATCAGCCTGATCATCCTGATTATGCTGTGGCAGAAGAAGCCCCGCTGAGG

ATCCAATGTAACTGTATTCAGCGATGACGAAATTCTTAGCTATTGTAATACTCTAGA

GGATCTTTGTGAAGGAACCTTACTTCTGTGGTGTGACATAATTGGACAAACTACCTA

CAGAGATTTAAAGCTCTAAGGTAAATATAAAATTTTTAAGTGTATAATGTGTTAAAC

TACTGATTCTAATTGTTTGTGTATTTTAGATTCCAACCTATGGAACTGATGAATGGGA

GCAGTGGTGGAATGCCTTTAATGAGGAAAACCTGTTTTGCTCAGAAGAAATGCCATC

TAGTGATGATGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAAAGAAGAG

AAAGGTAGAAGACCCCAAGGACTTTCCTTCAGAATTGCTAAGTTTTTTGAGTCATGC

TGTGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATTTACACCACAAAGGAAAAAGC

TGCACTGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTATAAGTAGGCA

TAACAGTTATAATCATAACATACTGTTTTTTCTTACTCCACACAGGCATAGAGTGTCT

GCTATTAATAACTATGCTCAAAAATTGTGTACCTTTAGCTTTTTAATTTGTAAAGGGG

TTAATAAGGAATATTTGATGTATAGTGCCTTGACTAGAGATCATAATCAGCCATACC

ACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGA

AACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTA

CAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTC

TAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGCGGCTCTAGAGCT

GCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTC

CGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATC

AGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAA

AGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT

GCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTC

AAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTG

GAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGC

CTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGT

TCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCC

GACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGAC

TTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG

CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAG

TATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCT

CTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGC

AGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGT

CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAA

AAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAA

GTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTA

TCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGAT

AACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAG

ACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC

GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGC

CGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATT

GCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTT

CCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCT

CCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGG

TTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGT

GACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTG

CTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGT

GCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTT

GAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACT

TTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG

AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTG

AAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAA

AAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCT

AAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCT

TTCGTCTTCAAGAATTC

Appendix B. pCMV-Alfred plasmid sequence

CTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGT

TGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCA

CTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGT

GTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAA

GACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAG

AACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCG

CGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCG

CCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCA

AGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGA

CCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGAC

GGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAA

ACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGC

CGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATT

AATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGG

CAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCC

CAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACC

ATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATT

CTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGC

CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAA

AAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGG

ATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGT

GGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCG

CCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGT

CCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACG

ACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTG

GCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGC

CGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGC

TACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGA

TGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCG

CCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGT

CGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTC

TGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTT

GGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGT

GCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGAC

GAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCAAC

CTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGA

ATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAG

TTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATA

GCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTC

CAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTG

GCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCAC

ACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGC

TAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGT

GCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGG

CGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAG

CGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAAC

GCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGG

CCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATC

GACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT

CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACC

TGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA

TCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT

TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG

ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGT

ATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAA

GAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTG

GTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCA

AGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTA

CGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGA

TTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCA

ATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAG

GCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCG

TGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATAC

CGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGA

AGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAAT

TGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTT

GCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCT

CCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCG

GTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCA

CTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCT

TTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGAC

CGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTT

TAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTAC

CGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCAT

CTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAA

AAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAAT

ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTAT

TTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTG

ACGTCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTG

CTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCT

GAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTG

CATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAG

ATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTC

ATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCC

GCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCC

CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTA

AACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGA

CGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGA

CTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGG

TTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGT

CTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTT

CCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACG

GTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTG

GCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCATGGAA

ACCGACACACTGCTGCTGTGGGTGCTGCTTCTTTGGGTGCCCGGATCTACCGGCGAT

CTGTGTACACCTTCTAGAGCCGCTCTGCTGACCGGCAGAGAGCAGAAGCTGATCTCC

GAAGAGGACCTGAATGCCGTCGGCCAGGACACCCAAGAAGTGATCGTCGTCCCTCA

CAGCCTGCCTTTCAAGGTGGTGGTCATCAGCGCCATTCTGGCCCTGGTGGTGCTGAC

CATCATCAGCCTGATCATCCTGATTATGCTGTGGCAGAAGAAGCCCCGCTGA

Appendix C. Plasmids and primers

Table C.1. Plasmids purchased and created and their functions.

Name	**Main characteristics and purpose**
pCMV-TET3G	Contains constitutively expressed transactivator protein TET3G to activate pTET3G promoters in the presence of doxycyclin; one of the plasmids in the Tet-On® Inducible System
pTRE3G-ZsGreen1	Contains the pTRE3G promoter induced by TET3G-dox; expresses ZsGreen in the presence of doxycyclin; one of the plasmids in the Tet-On® Inducible System
pTRE3G-ZsGreen1-Alfred	Contains the pTRE3G promoter induced by TET3G-dox; expresses ZsGreen and Alfred in the presence of doxycyclin; one of the plasmids in the Tet-On® Inducible System
pCMV-FGE	Contains constitutively expressed formylglycine generating enzyme (FGE)
pCMV-Alfred	Contains constitutively expressed Alfred
mCardinal-N1	Contains constitutively expressed mCardinal

Table C.2. PCR primers

Plasmid created	Plasmid used as template	Primer name	Sequence
pCMV-Alfred	pCMV-FGE	CA_FGE_F	GCCCCGCTGACTCGAGTCTAGAGGGCCC
		CA_FGE_R	CGGTTTCCATGCTAGCCAGCTTGGGTCTC
	pTRE3G-ZsGreen1-Alfred	CA_Alf_F	GCTGGCTAGCATGGAAACCGACACACTGC
		CA_Alf_R	TAGACTCGAGTCAGCGGGGCTTCTTCTG

Table C.3. Sequencing primers

Plasmid sequenced	Primer name	Sequence
pTRE3G-ZsGreen1-Alfred	TA_SEQ_F1	CCTCGGTACACATGCTTTACAT
pCMV-Alfred	CA_SEQ_F1	GTTTGACTCACGGGGATTT
	CA_SEQ_R1	CTGCTATTGTCTTCCCAATC