Engineering Fluorescent Proteins for Next-Generation Bioimaging Applications

Athena Zitti

A Thesis submitted to Cardiff University for the degree of Doctor of Philosophy

Division of Molecular Bioscience

School of Biosciences

Cardiff University

June 2025

Acknowledgements

I'd first like to express my sincere gratitude to my supervisor Prof Dafydd Jones for all the guidance and support he provided me over the years and for all the opportunities he offered me through this journey. Thank you for believing in me back in 2020, for helping me get this studentship and reaching the finish line! I would like to also give a massive thank you to Prof Pete Watson, for all his support both academically and emotionally especially in the last year of my PhD. I would like to thank my secondary supervisor Prof Paola Borri for helping me understand the concepts of Raman Spectroscopy and the physics aspect of this project. I would also like to extend my gratitude to Dr David Regan for all the Raman imaging data collected for this thesis and all his help with understanding and analysing the data. A massive thank you goes towards Dr Pierre Rizkallah for all his help with solving my crystal structure and his nostalgic storytelling. Thank you to Dr Georgina Menzies for all her help with the Molecular Dynamics aspect of this thesis.

Big shoutout to Dr Ozan Aksakal, who from the beginning of our journey has stood by me, pushing me to keep going and not give up, and who's never stop helping me and supporting me until the last moment. I would also like to thank Dr Rochelle Ahmed for being there in my early days and giving me the skillset that helped shaped me in the researcher I am today. I would like to thank past members of the DDJ group for their help and support: Dr. Rebecca Gwyther, Dr Lainey Williamson and Dr Krithika Ramakrishnan. Thank you to Dr Drew Mack for his constant support, daily jokes and encouragement. Massive thank you to my first ever master students who became more than just members in the lab, Hannah Read and Kai Pawson. Big shoutout to Dr Emily Heath for all the support and encouragement over the years, from our early days of volunteering, our demonstrating hours and the constant push to keep going.

The biggest thank you goes to the colleagues who became friends that became family. The ones that carried me through the last year of my PhD and especially through the writing of this thesis. Adam Cutts, James Davies, John Mcklarnon and Danoo Vitsupakorn. I couldn't have reached the finish line without your unlimited support, the coffee runs, the plushies and most importantly last push to keep going!

I would also like to extend my gratitude to all the collaborators who have supported my work in this project. Thank you to Dr Magdalen Lipka-Loyds in the MDI, Dr Halina Mikolajek at DLS for the crystallisation and x-ray crystallography. Thank you to Dr Andreas Zumbusch, Dr Andrea Pruccoli at Konstanz University for the ep-RS imaging of the mCherry variants.

Thank you Dr Michalis Panayides for riding along on this crazy rollercoaster, from our weekly lunches to our endless phone calls of despair and always being my ray of sunshine on a dark day. Dr Christina Yiallouridou thank you for being my mother figure away from home and supporting me since the first day I met you. Dr Gianna Gavriel, thank you for always reassuring me that everything will be okay in the end. Massive shoutout to my best friend Eleftheria Yiannakou for always being there for me, through my breakdowns to my biggest achievements, thank you for your endless support and encouragement.

Finally, I would like to extend my biggest gratitude to my family and friends for supporting me throughout my academic years. Thank you, mum and dad, but most of all thank you to my sister for keeping up with me and cheering me on through this journey.

Publications

Zitti, A. and Jones, D. 2023. Expanding the genetic code: a non-natural amino acid story. *The Biochemist* 45(1), pp. 2–6. doi: 10.1042/bio_2023_102.

Ahmed, R. D., Jamieson, W. D., Vitsupakorn, D., **Zitti, A.,** Pawson, K. A., Castell, O. K., Watson, P. D., & Jones, D. D. (2025). Molecular dynamics guided identification of a brighter variant of superfolder Green Fluorescent Protein with increased photobleaching resistance. Communications Chemistry, 8(1), 174. doi: 10.1038/s42004-025-01573-4

Zitti, A., Aksakal, O., Vitsupakorn, D., Rizkallah P. J., Mikolajek, H., Platts, J., Menzies, G. E., & Jones, D. D. (2025). Structure, Function and Dynamics of mCoral, a pH-Responsive Engineered Variant of the mCherry Fluorescent Protein with Improved Hydrogen Tolerance. International Journal of Molecular Sciences (IJMS), 27(1), 154. doi: 10.3390/ijms27010154

Protein structure:

Zitti, A., Rizkallah, P. J., Jones, D. D. (2025). mCoral: a mCherry variant with cysteine replacing methionine in the chromophore. PDB: 9TAD, doi: 10.2210/pdb9tad/pdb

Abstract

Modern biology and biomedical research have been transformed by advances in optical techniques that enable the sensitive detection and spatially resolved characterization of biomolecules. Among these, fluorescence microscopy has emerged as a pivotal tool for visualizing cellular structures and dynamics, allowing researchers to observe biological systems with high specificity and temporal resolution. Having the ability to selectively label cellular components with fluorophores has revolutionized the fields of cell biology, neurobiology and medical diagnostics.

Central to the success of fluorescent-based approaches has been the engineering of fluorescent proteins. From the discovery of the first fluorescent protein, the green fluorescent protein (GFP), to its subsequent optimization that led to a diverse family of genetically encoded fluorescent reporters and later the discovery and development of a repertoire of different coloured proteins, there is a constant need for fluorescent proteins (FPs) with better features to fit different analysis requirements.

This thesis explores the engineering of fluorescent proteins (FPs) to enhance their utility in bioimaging and Raman-based spectroscopy. Focusing on mCherry, mNeptune, and superfolder GFP (sfGFP), the work employs site-directed mutagenesis and non-natural amino acid (nnAA) incorporation to modulate fluorescence properties and introduce vibrational signatures detectable by Raman spectroscopy.

In Chapter 3, nnAAs—p-cyano-phenylalanine (pCNPhe) and p-ethynyl-phenylalanine (pCCPhe)—were incorporated into specific residues of mCherry and mNeptune to introduce C≡N and C≡C vibrational bonds. Structural modelling guided residue selection near the chromophore to maximize Raman coupling. Although protein expression was successful, many mutants lacked colour, suggesting disrupted chromophore maturation. Nonetheless, this work laid foundational strategies for genetically encoded Raman-active probes.

Chapter 4 details the creation of a novel mCherry variant (mCherryM66C), where cysteine was introduced at position 66 of the chromophore. Spectral analysis showed a 21 nm blue shift in excitation and a 25 nm shift in emission, along with a modest increase in

quantum yield. mCherryM66C displayed dual pH-dependent transitions (pKa = 5.7 and 8.8), consistent with a three-state chromophore model involving sequential protonation. While Raman signal was reduced compared to wild-type mCherry, the variant's pH sensitivity and environmental responsiveness suggest potential use as a biosensor. Chapter 5 investigates the sfGFP H148S mutant (YuzuFP), developed through molecular dynamics simulations to enhance chromophore hydrogen bonding. Replacing histidine with serine resulted in increased brightness (~1.5-fold), improved absorbance, and significantly enhanced photobleaching resistance (~3-fold). Despite minimal change in pKa, YuzuFP maintained fluorescence more effectively across a wide pH range, affirming its stability and suitability for imaging applications.

Collectively, this work demonstrates the effectiveness of combining computational design with biochemical engineering to create next-generation FPs with enhanced spectral properties and environmental sensitivity. These engineered variants hold promise for advanced imaging techniques, including multiplexed fluorescence and Raman-based biosensing.

Table of contents

Acknowledgements i

Publications iii

Abstract iv

Table of Contents vi

1. Introduction 1

1.1. Fluorescence Microscopy 1

1.1.1. Principles of Fluorescence 2

1.1.2. Fluorescent Proteins 4

1.1.2.1. Green Fluorescent Protein (GFP) 7

1.1.2.2. Red Fluorescent Proteins (RFPs) 10

1.1.2.2.1. mCherry and mNeptune 10

1.1.3. Limitations of Fluorescence Microscopy 13

1.2. Raman Spectroscopy 15

1.2.1. Principles of Raman Spectroscopy 15

1.2.2. Electronic pre-resonant Raman scattering 18

1.2.3. Genetically encoded Raman-active probes 20

1.3. Molecular Biosensors 24

1.4. Aims and Objectives 25

2. Materials and Methods 28

2.1. Materials 28

2.1.1. Laboratory Chemicals 28

2.1.2. Nucleic acid modifying enzymes 33

2.1.3. Bacterial growth media 34

2.1.4. Bacterial strains 34

2.2. Molecular Biology 35

2.2.1. Side-directed mutagenesis (SDM)

2.2.1.1. Vectors for protein expression 36

2.2.1.2. PCR product purification 39

2.2.1.3. Plasmid DNA purification 40
2.2.1.4. DNA phosphorylation and ligation 40
2.2.1.5. Non – canonical amino acids preparation 41
2.2.2. Agarose gel electrophoresis 41
2.2.3. Gel excision 41
2.2.4. E. coli transformation 42
2.3. Recombinants protein expression and purification 42
2.3.1. Recombinant protein production 42
2.3.2. Fluorescent protein purification 43
2.4. Protein Analysis 43
2.4.1. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) 43
2.4.2. Protein concentration 44
2.4.3. Spectral properties analysis 44
2.4.4. Biochemical analysis 45
2.4.4.1. pH profile 45
2.4.4.2. Oxidation potential 45
2.4.4.3. Mass Spectrometry 46
2.5. Protein Crystallography 46
2.5.1. mCherry^{M66C} crystallization 46
2.5.2. X-ray diffraction of mCherry^{M66C} 47
2.6. Computational modelling analysis 48
2.6.1. *In silico* modelling 48
2.6.2. Computational requirements 48
2.6.3. Parametrisation of mCherry 48
2.6.4. Input structure preparation 50
2.6.5. Energy minimisation 51
2.6.6. Two-step equilibrium 51
2.6.7. MD production run 52

2.6.8. Calculating Root Mean Square Deviations (RMSD) 52
2.6.9. Calculating Root Mean Square Fluctuations (RMSF) 53
2.6.10. Hydrogen bonding and pairwise distance analysis 54
2.7. Cell Imaging 55
2.7.1. Mammalian cell constructs 55
2.7.2. Transfection of mammalian constructs and cell imaging 56
2.8. Imaging Techniques 58
2.8.1. Single molecule analysis 58
2.8.2. SRS Imaging 61
3. Incorporation of Raman active bonds in Far – Red Fluorescent Proteins 63
3.1. Introduction 63
3.2. Results and Discussion 65
3.2.1. *In silico* structural analysis of mCherry and mNeptune and target residue selection 65
3.2.2. Mutant construction and production 68
3.2.3. Spectral analysis of incorporating non naturals in mCherry and mNeptune 71
3.2.4. Epr-SRS analysis of wild type mNeptune and mNeptune^{Y67pCCPhe} 74
3.3. Conclusion 76
4. Functional and structural characterisations of mCherry^{M66C} 77
4.1. Introduction 77
4.2. Results and Discussion 78
4.2.1. Incorporation of Cysteine in residue 66 79
4.2.2. Spectral characterisation of mCherry^{M66C} 80
4.2.3. Epr-SRS analysis of mCherry and mCherry^{M66C} 82
4.2.4. Responses of mCherry^{M66C} to changes in environment 85
4.2.4.1. The effect of pH on spectral properties 86
4.2.4.2. The effect of H_2O_2 on spectra properties 93
4.2.4.3. Mass Spectrometry of proteins in the presence of H_2O_2 99

4.2.5. Structure determination of mCherry 103
4.2.5.1. X-ray crystallography of mCherry^{M66C} 104
4.2.5.2. Structure analysis of mCherry^{M66C} 106
4.2.5.3. Comparison of mCherry^{M66C} to mCherry 107
4.2.6. Computational analysis of mCherry^{M66C} and mCherry 109
4.2.6.1. Molecular Dynamics analysis 109
4.3. Conclusion 121
5. Characterisation of superfolder GFP^{H148S} 123
5.1. Introduction 124
5.2. Results and Discussion 126
5.2.1. Engineering of H148S rational 126
5.2.2. The effects of H148S on the spectral properties of sfGFP 132
5.2.3. Photostability of YuzuFP 136
5.2.4. Crystallization limitations 145
5.2.5. Long time scale molecular dynamics 145
5.2.6. The importance of water molecules 151
5.3. Conclusion 155
6. Discussion 157
6.1. General Overview 157
6.2. Engineering of mCherry and mNeptune as Raman probes 157
6.3. Development of an mCherry-based fluorescent biosensor 158
6.4. Development of YuzuFP 160
6.5. Synopsis and future work 161
7. Bibliography 163

1 Introduction

The ability to understand the processes that occur within the cell has revolutionised our understanding of biology. Critical to this development have been the development of cell imaging techniques. At the forefront is fluorescence microscopy due to its high sensitivity but other approaches are emerging, such as Raman-based methods (Xiong et al. 2019; Singh et al. 2023) that overcome some of the inherent problems with fluorescence approaches. Probe development has also been critical as it allows certain cell components and now individual proteins to be labelled and monitored. One of the most important probe classes are the fluorescent proteins (FPs) as they allow genetic encoding of a fluorescent probe so allowing target-specific labelling and tracking (Thorn 2017). There are still room for FP improvement so they are brighter, the signal last for longer and can report actively on cellular events. Approaches such a Raman microscopy currently lack such probes, but inherent chemistry of biomolecules allows detection of molecule classes (e.g. proteins, lipids) (Dodo et al. 2022). However, the development of genetically encoded Raman tags would greatly facilitate its expansion in cell imaging.

1.1. Fluorescence Microscopy

The ability to observe biological structures and dynamics at the cellular and molecular levels has been a pivotal factor in advancing biomedical research. Among the various techniques developed for this purpose, fluorescence microscopy has emerged as one of the most powerful and versatile tools (Lakowicz 2006). Its widespread application is rooted in the fundamental principles of fluorescence and relies on the unique photophysical properties of fluorophores and chromophores. These are molecules specific class of chromophores are capable of undergoing electronic excitation and subsequently

release energy as emitted light once they return to the ground state. Fluorescence has become a vital tool in biology, especially for real-time, live-cell imaging, where its sensitivity and specificity provide unique insights into dynamic biological processes.

1.1.1 Principles of Fluorescence

Fluorescence is a photophysical phenomenon where a molecule absorbs light at a specific wavelength and subsequently emits light at a longer wavelength, allowing for the high sensitivity and selective detection of biological specimens (Valeur and Berberan-Santos 2012). Fluorescence occurs when a fluorophore absorbs a photon, and an electron is promoted from its ground state (S0) to an excited state (S1). The electron may undergo internal conversion, releasing some energy vibrationally, before returning to the ground state and emitting a photon at a lower energy (longer wavelength) (Figure 1.1). The difference between the excitation and emission wavelengths is called the Stokes shift, and it allows emitted light to be distinguished from the excitation source. Chromophores Organic fluorophores in particular, have the ability to undergo excitation and emission due to the behaviour of their in their conjugated bond system. These cChromophores contain extended π-electrons in conjugated double bonds systems leading to that get delocalisation of their π electron network ed leading to aso lowering the energy gap between the ground and excited states. As a result,The energy gap of this electronic transistion means that the chromophore can absorbs photons within the visible region of the electromagnetic spectrum (Tsien 1998).

The spectral properties of a chromophore are critical to its performance in imaging applications. Specifically, the excitation maximum (λ_{max}) and emission maximum (λ_{EM}) which define the wavelengths at which a chromophore absorbs and emits light most efficiently, respectively. The λ_{max} corresponds to the peak of the absorption spectrum, while the λ_{EM} indicates the peak fluorescence emission. The difference between these two wavelengths, known as the Stokes shift, plays a crucial role in minimizing spectral overlap between excitation and emission signals, thereby enhancing detection sensitivity and signal to noise ratio (Lakowicz J.R. 2006). Accurate characterisation of λ_{max} and λ_{EM} is

essential for the optimal selection of excitation sources and emission filters in fluorescence-based techniques.

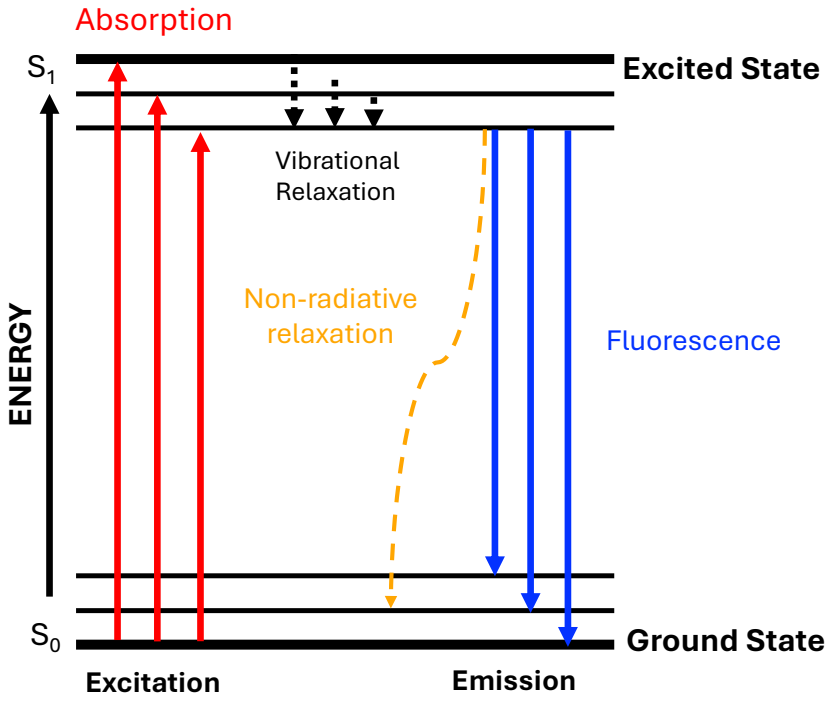

Figure 1.1: Jablonski diagram summarising the mechanism of fluorescence. Excitation of electrons from the ground state (S_0) when light (black arrow) is absorbed (red arrows) to an excited state (S_1). Vibrational relaxation (black dashed arrows) and non-radiative relaxation (orange arrows) returning the electrons to the ground state and emitting a photon (blue arrows).

Moreover, fluorescence efficiency is usually quantified using the quantum yield (QY), which represents the ratio of emitted to absorbed photons. A higher quantum yield (approaching 1) indicates a more efficient fluorophore while a value of 0 denotes no emission, indicating that all the absorbed energy is lost through non-radiative processes such as internal conversion or intersystem crossing to a potential non-fluorescent triplet state (Valeur and Berberan-Santos 2012; Würth et al. 2013). Quantum yield is not just a simple measure of fluorescence efficiency, it is in itself influenced by many factors such as the molecular structure of a chromophore as well as its surrounding environment (e.g.

hydrogen-bonds, solvent and pH) (Follenius-Wund et al. 2003; Cranfill et al. 2016). In parallel, the fluorescence Fluorescence lifetime is another important probe attribute which is theas it is a measure of the average time a molecule spends in its excited state before returning to the ground state, providinges complimentary insight into the fluorophore's behaviour and environment (Houston et al. 2024). Another factor of fluorescence to consider is the molar extinction coefficient of a molecule usually of a chromophore. Also referred to as the molar absorption coefficient, is a measure of how strongly a fluorescent protein absorbs light at a given wavelength. Together with quantum yield, these two factors are used to calculate the brightness of a fluorescent probe.

1.1.2 Fluorescent proteins

Many biological molecules are fluorescent. For example, many proteins are inherently fluorescent due to tryptophan (Toseland 2013). However, given that many different proteins that comprise the proteome contain tryptophan, the information gained is quite limited to gross cellular features. Small molecule fluorescent dyes have been developed to looks at particular cellular locations. For example, DAPI (4',6-diamidino-2-phenylindole) specifically labels DNA so allowing identification of the nucleus (Kapuscinski 1995). However, dyes have to be added to the cell so can be an invasive and lack specificity (Kim et al. 2021).

Fluorescent proteins (FPs) have revolutionised the field of fluorescence microscopy, enabling real-time, non-invasive visualization of dynamic biological processes in living cells and organisms. Since the discovery and cloning of Green fluorescent protein (GFP) from *Aequorea Victoria*, FPs have become an indispensable tool in modern cell biology (Shaner et al. 2005; Rodriguez et al. 2017). The critical

Regardless of their originating species or degree of genetic engineering, all FPs are ~27 kDa and are composed of an extremely rigid 11-stranded β-barrel that encloses an internal chromophore that forms spontaneously through a series of post-translation modifications (Figure 1.2) (Ormö et al. 1996). As briefly described beforechromophore is

responsible for the protein's fluorescence, absorbing light at a specific excitation wavelength and emitting it at a longer emission wavelength.

Regardless of the source of the FP, the mechanism of chromophore formation is understood to be similar across all fluorescent proteins (Remington 2006).

It is formed through the covalent rearrangement of three amino acids in the presence of molecular oxygen. This amino acid triad, also known as the XYG motif, usually consists of a variable residue (X), a tyrosine and a highly strictly conserved glycine (Subach et al. 2011). Although tyrosine is also highly conserved within nature like glycine, it can be replaced with other aromatic amino acids without disrupting the maturation of the chromophore (Heim et al. 1994). In GFP this motif consists of S65-Y66-G67 while in DsRed(from the coral Discosoma sp.) it consists of QM66-Y67-G68, with the conjugated double bond system expanding into the peptide bond with F65, giving mCherry its red region spectral properties. The molecular structures are shown below in Figure 1.2.

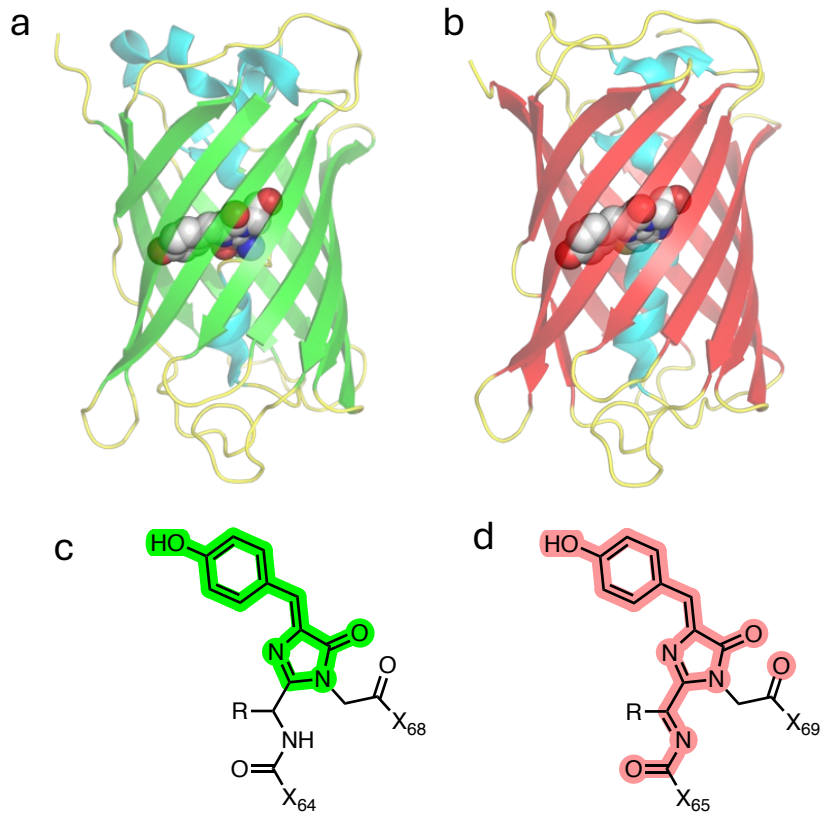

Figure 1.2: Basic structure of fluorescent proteins. Representative crystal structures of (a) green FP (EGFP, PDB: 4EUL (Arpino et al. 2012)) and (b) red FP (mCherry, PDB:2H5Q (Shu et al. 2006)). Chromophore structure of (c) green FP from *A. victoria* (avGFP) and (d) DsRed. The green and red highlighted regions represent the conjugated bond system central to the fluorescence properties.

The mechanism of the GFP chromophore maturation takes place in three steps: cyclization, oxidation and dehydration. The autocatalytic cyclization occurs between the glycine's amide nitrogen and the carbonyl carbon of serine, dehydration forms the double bond, and oxidation forms the β-methylene bridge between the phenolic and imidazolinone rings. Compared to GFP, multiple studies have suggested that the mechanism of chromophore formation of mCherry and other proteins descending from DsRed, is more complex (Subach and Verkhusha 2012) leading to an extended conjugated double bond system which gives RFPs their longer absorbance/emission wavelengths (Remington 2006). Chromophore maturation in mCherry, requires an additional oxidative step that leads to the formation of an acylimine linkage in the polypeptide backbone (Craggs 2009; Wachter et al. 2010).

Since its discovery, GFP has been extensively engineered to create a pallet of FPs with variable excitation and emission properties such as yellow (YFP) and cyan (CFP) fluorescent protein. In a similar manner red-shifted proteins such as mCherry and mNeptune, are derivatives of extensive engineering from their original proteins DsRed and eqFP578 respectively.

1.1.2.1 Green Fluorescent Protein (GFP)

Osamu Shimomura discovered that upon irradiation of the *Aequorea Victoria* jellyfish with ultraviolet light, it fluoresced green (Shimomura et al. 1962). Two proteins were isolated and purified from *A.victoria,* the chemiluminescent aequorin and the green fluorescent protein. Aequorin was the first to be isolated and studied, however, unlike the green fluorescence originally observed in the jellyfish tissue, its isolated form emitted a blue light instead. This led to the second protein being identified, which was responsible for the green fluorescence and thus was named green fluorescent protein notably known as GFP. GFP is a 27 kDa protein made up by a single polypeptide chain composed of 238 amino acids that folds into the characteristic β-barrel form that all FPs are known to acquire, enclosing its functional centre known as the chromophore (Yang et al. 1996) (Figure 1.2).

The mechanism behind the final green fluorescence was later suggested to be a transfer of energy between the two proteins (Morin and Hastings 1971) where it was found that aequorin's peak emission is at 470 nm which was close to one of the two excitation peaks of GFP. As a result, GFP could be excited and emit at 508 nm which resulted in its greenish colour observed in the tissues (Johnson et al. 1962). When the protein was isolated and purified, allowing its spectral properties (absorbance and fluorescence spectrum) to be recorded, that it confirmed the energy transfer theory as GFP was found to absorb the blue emission of aequorin and then convert it to the green light that was originally observed.

The mechanism for this was described by Morise *et al.* 1974 where it was suggested that this energy transfer between the two proteins was facilitated by Förster resonance

energy transfer (FRET). Moving forward, GFP was favoured for imaging instead of aequorin. Aequorin is chemiluminescent via calcium activated activation meaning it requires a co-factor, while GFP undergoes an autocatalytic cyclisation (described in 1.1.2) which give rise to its fluorescence in the absence of a co-factor making it ideal for tagging as it can be expressed in any cell type or organism without the need to modify the cellular environment.

Although the original GFP from *Aequorea victoria* (avGFP) is still the underlying protein scaffold by which many FPs have been engineered, the first truly useful engineered version of GFP that is still used today is enhanced GFP (EGFP). It is considered a better derivative as compared to avGFP, as there is a shift from the neutral protonated form of the chromophore (CroOH, λ_{max} 395 nm) to the deprotonated charge phenolate form ($CroO^-$, λ_{max} 485 nm). Many more derivatives have been engineered since then including superfolder GFP (sfGFP) which is the one primarily used in this thesis.

Superfolder GFP was the end result of a series of directed evolution approaches by (Pédelacq et al. 2006). The engineering began with cycle-3 GFP (αGFP), a variant of GFP with four mutations compared to the avGFP background (Q80R, F99S, M153T and V163A) (Crameri et al. 1996). Following that (Waldo et al. 1999) replaced phenylalanine at residue 64 with a leucine, and serine 65 to threonine. This resulted in the subtle rearrangement of the hydrophobic core packing close to the chromophore, including a reduction in surface exposed hydrophobic residues and an effect in the local hydrogen bond network in the vicinity of the chromophore, respectively (Arpino et al. 2012). Moving forward, a series of mutations were introduced into the αGFP background including S30R, Y39N, N105T, Y145F, I171V and A206V to give rise to sfGFP. These mutations conferred a significantly increased resistance to denaturation and misfolding, particularly under conditions that challenged protein stability, such as fusion to aggregation-prone partners or expression in oxidizing environments like the periplasm of *E. coli* (Aronson et al. 2011).

While sfGFP has a faster rate of protein folding and chromophore maturation, and is very stable compared to avGFP, it has similar fluorescent properties (brightness and photobleaching resistance) to that of EGFP (Cranfill et al. 2016). Green-yellow versions of avGFP with improved brightness such as Venus (Nagai et al. 2002) have previously been

engineered, but at the expense of having their resistance to photobleaching reduced (Cranfill et al. 2016). A key determinant of GFPs fluorescent properties is the histidine at position 148. Together with a structurally conserved water molecule, H148 directly interact with the phenolate group of the chromophore and is thought to play a key role in stabilising $CroO^-$ in the ground state (Seifert et al. 2003; Shu et al. 2007; Campanini et al. 2013). It has also been observed that, other non-avGFPs, have a non-ionisable polar residue instead of a histidine (Ong et al. 2011). Proteins like Dronpa which has a higher quantum yield (0.85) to that of sfGFP, has a serine at the equivalent position to H148 (residue 142) (Ando et al. 2004). In a similar manner, mCherry and other DsRed derived FPs have a serine at position 146 which again is the equivalent to that of H148 (Figure 1.3). Serine seems to be a favourable alternative to histidine in other natural FPs, hence in this thesis we will explore the effect of modifying H148S in sfGFP.

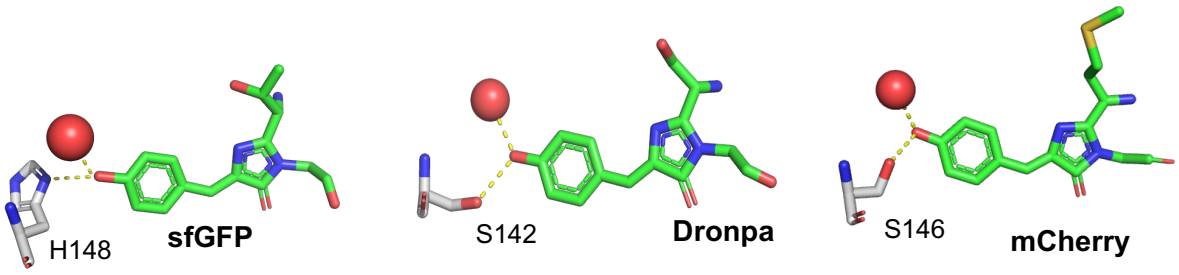

Figure 1.3: Residues equivalent to H148 that interact with the chromophore phenol group. S142 in Dronpa (PDB: 2IE2) and S146 in mCherry (PDB: 2H5Q). Figure adapted from Ahmed et. al (2025).

1.1.2.2 Red Fluorescent Proteins (RFPs)

Red fluorescent proteins (RFPs) are a subclass of fluorescent proteins that emit light in the red to far-red region of the visible spectrum. They are derived from naturally occurring anthozoa species such as *Discosoma* and *Entacmaea* and their development has significantly expanded the utility of fluorescence microscopy, especially for multicolour imaging, deep tissue visualization and long-term live-cell imaging due to their spectral properties and reduced phototoxicity (Piatkevich et al. 2010).

The first widely used RFP, DsRed, was derived from the coral *Discosoma sp.*(Matz et al. 1999). Similar to avGFP, DsRed had its limitation which restricted its application, including being an obligate tetramer, slow maturation time and propensity for aggregation, making them challenging for applications requiring sub-cellular localization (Campbell et al. 2002a). To overcome these constraints, intensive engineering efforts, such as rational protein engineering and directed evolution, gave rise to a new generation of monomeric RFPs such as mRFP1, mCherry (from DsRed) (Campbell et al. 2002; Shaner et al. 2004) and mNeptune (from eqFP578) (Lin et al. 2009). Although alternatives to DsRed such as eqFP578 with better properties (faster maturation and brighter) existed, it still had similar downsides due to its tetrameric oligomerisation and slow maturation making it inappropriate to use for live cell imaging or monitoring rapid reactions (Merzlyak et al. 2007). Hence again the need for new engineered variants like mCherry and mNeptune.

1.1.2.2.1 mCherry and mNeptune

An important aspect of this thesis covers the engineering of primarily mCherry and mNeptune. The mCherry variant is one of the most popular well-characterized RFP variants, developed as a monomeric derivative of DsRed. Through extensive rational protein engineering and directed evolution, mCherry was optimized for high brightness, rapid maturation and photostability (Shaner et al. 2004). Its red emission wavelength (610 nm) makes it ideal to be used in conjunction with other spectrally distinct fluorescent proteins like EGFP (Green) or mKate2 (Far-red), for multi-colour labelling experiments, minimizing spectral overlap and cellular autofluorescence. In addition to its improved emission characteristics, mCherry exhibits a high extinction coefficient of ~72000 $M^{-1}cm^{-1}$ and a decent quantum yield of 0.22 (Shaner et al. 2004). Functionally, mCherry has proven valuable across many biological applications, from cellular and developmental studies to *in vivo* imaging of whole organisms (Shu et al. 2006). Its relatively low pKa (~4.5) makes it highly advantageous for physiological studies allowing it to remain fluorescent across a range of cellular environments (Shaner et al. 2004). Moreover, its high resistance to

photobleaching allows for long-term imaging, making it ideal for time-lapse studies and super resolution microscopy (Shaner et al. 2007).

In a similar manner to mCherry, mNeptune was optimised compared to its predecessors, TagRFP and eqFP578 as a near-infrared fluorescent protein (NIR) (Shcherbakova and Verkhusha 2013). Through extensive protein engineering Lin et al. (2009) successfully shifted TagRFP's λ_{EM} to 650 nm, generating a bright and stable FP suitable for deep-tissue imaging. Due to its balance between brightness, photostability and monomeric structure, mNeptune makes an ideal protein for long duration live imaging and *in vivo* studies (Shaner et al. 2007).

Table 1.1: Spectral properties of mCherry, mNeptune and other red fluorescent proteins. Data acquired from FPbase.

FP	λ_{max} (nm)	λ_{EM} (nm)	ε ($mM^{-1}cm^{-1}$)	QY	Brightness	pKa
DsRed	558	583	72.5	0.68	49.3	N/A
mRFP1	584	607	50	0.25	12.5	4.5
mCherry	587	610	72	0.22	15.84	4.5
eqFP578	552	578	102	0.54	55.08	N/A
TagRFP	555	584	100	0.48	48.0	3.8
mNeptune	600	650	67	0.2	13.4	5.4

Like all fluorescent proteins, mCherry and mNeptune adopt the β-barrel fold, enclosing the chromophore, providing a stable environment (Figure 1.4). In RFPs, the chromophore undergoes an additional oxidation step that dehydrogenates the Cα-N bond on the variable residues of the XYG motif, forming an acylimine group. This additional group is what leads to the extended conjugated π-electron system of the chromophore. This essentially lowers the energy gap between the ground and excited states due to better electron delocalization, resulting in the red-shifted excitation and emission wavelengths observed compared to GFPs (Gross et al. 2000; Subach and Verkhusha 2012). The overall structure of both proteins is important for stabilising the chromophore and tuning the

spectral properties of the protein. Aside from the chromophore, interactions with neighbouring residues and water molecules (similar to GFP), are important for the overall optical properties of the protein.

One of the important mutations on converting DsRed to useful versions such as mCherry is the targeting of the chromophore forming residue, Q66. In mCherry and many other DsRed derived RFPs, Q66 is mutated to methionine. The mutation is thought to contribute towards optimising spectral properties and maturation (Shaner et al. 2004). In mNeptune and its original parent eqFP578, the equivalent residue is M64.

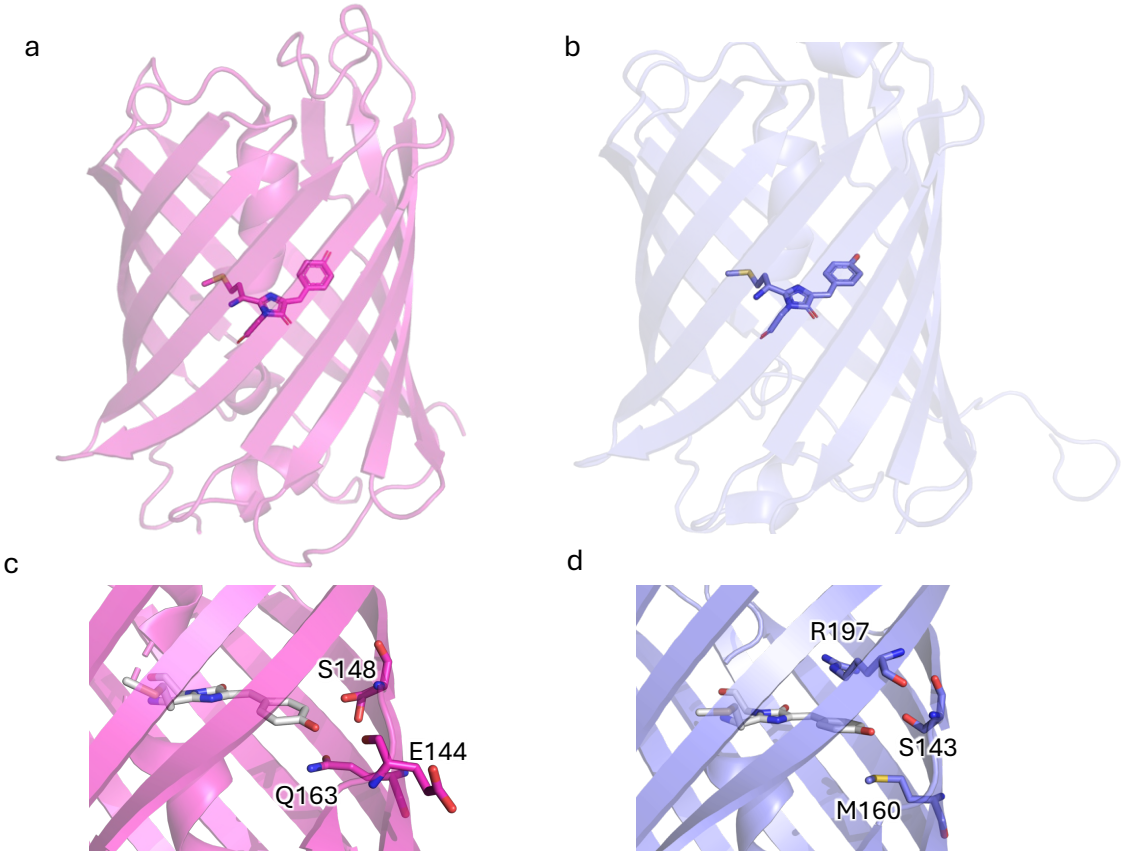

Figure 1.4: mCherry and mNeptune structure. β-barrel structure of a) mCherry (PDB: 2H5Q) and b) mNeptune (PDD: 3IP2). Top-view of c) mCherry and d) mNeptune showing the interaction between the chromophore phenol oxygen and neighbouring residues.

1.1.3 Limitations of Fluorescence Microscopy (FM)

While fluorescence microscopy has revolutionised cellular and molecular biology, fluorescence imaging still has its limitations. Some of these challenges are particularly important when attempting to extend fluorescent techniques to more complex, longer

timescale or multiplexed imaging applications. Such examples include chemical damage of the fluorescent protein via photobleaching after prolonged excitation (Grigorenko et al. 2015). This degradation of fluorophores over time restricts the duration of imaging, and particularly in time-lapse or long-term live cell experiments (Sarangi 2022). Aside from photobleaching, cells themselves can suffer from photoxicity, where constant light excitation can lead to damage of cellular structures or alter physiological processes, potentially introducing artifacts during live imaging studies (Icha et al. 2017).

Another critical constraint is the spectral overlap between fluorophores which limits the multiplexing potential of FM. The visible spectrum spans the 380 – 750 nm wavelength range and each emission peak spans around ~40 – 50 nm. In practice we could apply between up to 5 fluorescent proteins to a single sample before the different colours become unresolvable (Orth et al. 2018). It thus becomes limiting in experiments where there is a need to analyse biochemical pathways which could involve more molecules of interest. Increasing the number of different FPs used at a time risks the potential of inducing a FRET effect. Having a FRET effect is fine when looking at protein-protein interactions but within the context of multiplexing capabilities, it will limit the ability to image the localisation of multiple separate proteins.

Autofluorescence, which is the natural emission of light by biological molecules can further complicate imaging by introducing background noise obscuring specific signals. This is particularly problematic in tissues with high endogenous fluorescence as well as thicker or more complex biological tissues (Jun et al. 2017). Limited penetration depth is another issue that can come with autofluorescence, as it prevents the target signal from being identified and accumulates with depth contributing to a higher background signal. In thick tissues, light scattering and absorption reduce the intensity and focus of both the excitation and emission light. This not only limits how deep the microscope can effectively image but also diffuses autofluorescent signals, making it harder to distinguish them (Dang et al. 2019). The discovery and use of red and far-red fluorescent proteins has helped deal with the issue of autofluorescence and depth of imaging, however, it can still pose a limitation when in need of the use of different coloured FPs.

Quantitative analysis in fluorescence microscopy is inherently difficult as it involves measuring a variable (fluorescence intensity) that is affected by many experimental and biological parameters like the aforementioned. Therefore, with advancing research there is a constant need to advance the tools required. All these limitations, highlight the need for complementary imaging techniques such as Raman spectroscopy that can offer sharper spectral peaks and access to other detectable areas of the electromagnetic spectrum, as well as the improvement of current FPs, with improved photobleaching resistance.

1.2 Raman Spectroscopy

1.2.1 Principles of Raman Spectroscopy

Raman scattering is an inelastic light scattering phenomenon where incident photons first discovered by C. V. Raman and K. S. Krishnan in 1928, earning them the Nobel Prize in Physics in 1930, interact with vibrational modes of a molecular system, yielding a shift in the energy of the scattered photons (Long 2002). This effect provides a powerful spectroscopic technique for probing molecular vibrations and gaining insights into the structure and dynamics of matter (Raman and Krishnan 1928) and has since evolved into the powerful spectroscopic technique known as Raman Spectroscopy. The technique is widely used across physics, chemistry, materials science, biology and engineering as it has the ability to provide molecular-level information.

At its core, Raman scattering occurs when incident photons interact with the molecular vibrations of a sample, resulting in a shift in the energy of the scattered photons. To induce this event, a monochromatic light from a laser is usually directed onto the sample and the photons interact with the sample. While most photons are elastically scattered (Rayleigh scattering), a small fraction undergoes inelastic scattering, leading either to a gain (anti-Stokes) or loss (Stokes) in photon energy (Figure 1.5). These energy shifts correspond to the vibrational energy levels of the molecules and are characteristic of their chemical structure and environment. Each molecule has its own unique fingerprint, or spectrum according to the chemical bonds present within it (Smith and Dent 2005). Different vibrational modes, including, stretching, bending, or torsion of bonds, can be

directly identified through a Raman spectrum, which consists of sharp well-defined peaks of Raman scattered intensity against light wavelength or frequency, corresponding to these specific vibrations.

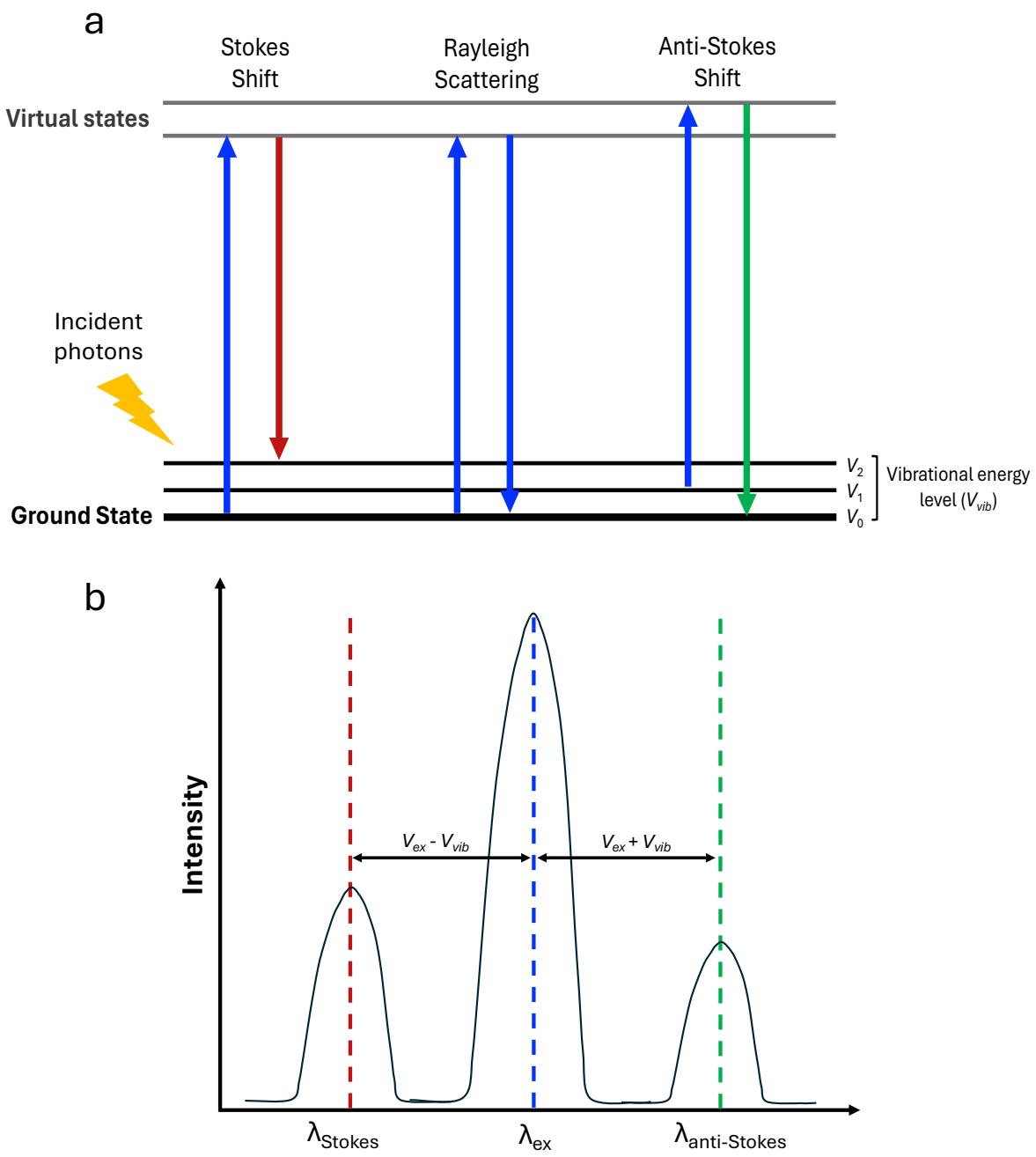

Figure 1.5: Raman Scattering and shift types. a) Energy diagram showing the different types of Raman scattering after excitation by incident photons. Stokes (red arrow) scattering occurs when the photon loses energy to the molecule populating a higher vibration state (V_2), whereas anti-Stokes scattering (green arrow) gains energy from the excited vibrational state of a molecule (V_1). Rayleigh scattering (blue arrows) illustrate the elastic scattering where photon energy remains unchanged. b) Raman spectrum showing the difference in intensity of each shift.

Raman scattering offers critical insights into the vibrational energy of molecules through energy transfer between quantised vibrational energy levels (Campanella et al. 2021). The molecular energy levels are typically categorized into a ground state and an excited state. The vibrational states are denoted as v_0, v_1 *and* v_2 and they represent the ground state and the first two excited states respectively. Figure 1.5 shows the concept of virtual states, which are non-resonant, transient energy levels depicted as grey lines. They do not correspond to actual molecular levels but rather serve as intermediate states during scattering. When a photon interacts with a molecule, it can temporarily elevate the system to a virtual state. The molecule then relaxes to a lower vibrational state, by emitting a scattered photon with either lower energy (Stokes) or higher energy (anti-Stokes) than the incident photon. Under standard temperature conditions, most molecules reside in the ground state (v_0), making transitions to higher vibrational states more likely than from already excited states. Consequently, stokes scattering is more probable and typically appears with greater intensity in Raman spectra compared to anti-stokes scattering (Keresztury 2006).

A Raman shift can be induced when incident light interacts with the vibrational modes of a molecule. This shift refers to the difference in frequency between the incident and scattered photons and is typically expressed in wavenumbers (cm^{-1}) rather than wavelength. Wavenumbers are a unit of frequency and thus are directly proportional to vibrational energy of chemical bonds whereas wavelength is a unit of the physical distance and is inversely to both the frequency and wave energy. Moreover, the Raman spectrum spans a broad range of wavenumbers, with each region offering specific insights into molecular structure and environment. One of the most informative regions is the "fingerprint" region, typically ranging from 400 to 800 cm^{-1}. This region is rich in vibrational modes that are highly specific to functional groups and chemical bonds within a molecule. For example, characteristic stretches such as C=C (1500-1900 cm^{-1}), C=N (1610-1680 cm^{-1}) and C=O (1680-1800 cm^{-1}) are commonly observed in organic compounds (Handzo et al. 2022). These vibrational signatures are unique to each molecule; thus, the fingerprint region serves as a molecular identifier, and is widely used in the analysis of proteins, lipids

as well as nucleic acids without the need for labels or chemical modifications (Lindley et al. 2021; Zhang et al. 2023). On the other hand, the biologically "silent" region, found typically between 1800-2400 cm^{-1}, is largely free from endogenous Raman signal of any biological molecules. This spectral gap provides an ideal window for introducing biorthogonal Raman tags, such as nitrile (2150-2250 cm^{-1}) and alkyne (2100-2150 cm^{-1}) groups. These tags produce sharp, distinct Raman peaks with minimal background interference, significantly enhancing signal to noise ratios in biological imaging and molecular tracking (Vardaki et al. 2024).

1.2.2 Electronic pre-resonant Raman scattering

Raman Scattering vibrational microscopy is an emerging analytical tool for imaging as it provides a label-free approach. Spontaneous Raman is a type of light scattering phenomenon where two photons are involved as described in Section 1.2.1. Different molecules will scatter incident photons in a unique manner upon interaction due to the specific chemical composition. As a result, different peaks are observed in the Raman spectra that correspond to specific chemical bonds, hence offering molecular specificity and in extension information about the chemical composition of the sample of interest (Zhang et al. 2010).

Although a very powerful method, it still has its limitations, with the primary one being that the Raman effect has an extremely weak signal especially when compared to absorption/fluorescence signals (Evans and Xie 2008). As a result, Raman microscopy images require high laser powers and very long integration times (100 ms to 1 s per pixel), which then limit the level of application when it comes live cell imaging (van Manen et al. 2005). Increasing the incident power of the laser will improve the signal intensity, but this is not an ideal solution as overtime this will again lead to photodamage of the sample. However, Raman spectroscopy is particular useful when complemented with fluorescent microscopy as it provides key chemical information about target molecules due to its chemical-bond specific vibrational transitions (Suhalim et al. 2012).

An alternative method that has also been previously employed to boost Raman signal is to selectively target molecules with strong electronic transitions such as small molecule chromophores, that will electronically enhance the vibrational Raman response (Szymanski 1970) but at the same time removing its for label-free benefit.

Electronic pre-resonant Raman Scattering (epr-RS) has been an emerging solution in enhancing the Raman scattering cross-section by enhancing the Raman polarizability of the molecule without necessarily needing to increase laser strength or acquisition time. There are two primary ways of implementing epr-RS; through Stimulated Raman Scattering (SRS) and Coherent anti-stokes Raman Scattering (CARS). In both methods, two lasers are used; a Pump beam and a Stokes beam, unlike in spontaneous Raman scattering that only uses one. The frequency of the two is adjusted so that when they interact with each other, a vibrational coherence is created in the sample, which in turn creates a coherent scattered signal allowing for a brighter and more directional scattered light to be detected.
CARS occurs when the Pump beam and the Stokes beam energy difference matches the target molecule's molecular vibrational mode. This vibrational excitation interacts coherently with a second pump photon, generating a new photon at the anti-stokes frequency which is blue shifted from the Pump and Stokes frequencies allowing easy detection (Evans and Xie 2008). On the other hand, SRS is where the Pump beam is at a higher energy level to that of the Stokes beam. Once the energy difference between the two beams matches the molecular vibration of the sample, it is excited into a vibrational state leading to a small intensity change. At this point there is a stimulated Raman gain in the Stokes beam and an analogous stimulated Raman loss in the pumps beam, which results in an enhanced Raman response (Wei and Min 2018). These principles have been previously demonstrated by characterising 28 different commercial dyes and used to prove that utilising the properties of SRS could help overcome fluorescent microscopy limitations (Wei et al. 2017) leading up to the main aim of one of the themes explored in this thesis. Utilising the super-multiplexing technique described by Wei et al. (2017) with fluorescent proteins instead of dyes will allow for the specific targeting of proteins as these novel probes will allow direct molecular tagging of target proteins.

1.2.3 Genetically encoded Raman-active probes

Current technology has proved the multiplexing capabilities with enhanced Raman response of commercial dyes. However, fluorescent dyes are limited in their use, as they often require sample fixation, which can then limit the ability for live-cell imaging of dynamic biological processes. In addition, to that, unlike fluorescent proteins, synthetic dyes are supplied to the sample externally, often suffering from non-specific binding and reduced cell viability over time (Zanetti-Domingues et al. 2013).

To address the limitation of synthetic Raman-active dyes, this thesis explores the use of genetically encoded Raman probes, following a similar strategy to that used for fluorescent protein tags. We specifically focused on monomeric red and far-red fluorescent proteins, that have previously been characterised for their optical properties (Luker et al. 2015). Both mNeptune and mCherry were chosen for this, due to their high molar absorption coefficients (≥ 67 000 $M^{-1}cm^{-1}$), which indicate strong electronic transitions that enhance pre-resonant Raman scattering microscopy (Campanella et al. 2021; Choorakuttil et al. 2023). Additionally, they are both widely used in florescent microscopy, and are also suitable for Raman-based imaging, due to the characteristic vibrational modes of their chromophores (Pruccoli et al. 2023).

The main goal for this was utilise the current conjugated system of the chromophore, with a focus on C=C and C=N stretching vibrations detectable around 1500 - 1900 cm^{-1}. Although, other biomolecules with C=C and C=N could be detected, the electronic pre-resonance enhancement effect of the chromophore should significantly amplify the Raman signal, hence making it more distinct compared to the other biomolecules. An alternative route to this, is to engineer the FPs to contain bonds that are visible within the biologically silent window instead. Doing that will prevent any limitation of mixing signals between biomolecules.

In order to do that, we need to incorporate bonds that resonate within the biologically silent window, and this can be done via the incorporation of non-natural amino acids (nnAAs) within the target FP. Non-natural amino acids, are modified non-proteinogenic amino acids not utilised by biology but can be incorporated into proteins

through genetic engineering. Expanding the available chemistry allows biochemists to introduce novel functional groups within their target proteins and in turn advance synthetic biology beyond the standard 20 amino acids (Zitti and Jones 2023). This is an important step forward in allowing researchers to study protein structure, function and overall chemistry while at the same time creating biomolecular probes with unique chemical properties.

The first step to incorporating these nnAAs, is by reprogramming the genetic code through amber codon suppression. This is done by repurposing the amber stop codon (TAG), which normally terminates mRNA translation on the ribosomes, to instead incorporate a nnAA at a target site in the protein. This is done using an orthogonal tRNA/aminoacyl-tRNA synthetase pair, that are engineered to specifically recognise the amber stop codon and their corresponding nnAA without interfering with the host cell's natural translation machinery (Blight et al. 2004; Miyake-Stoner et al. 2010) (Figure 1.6).

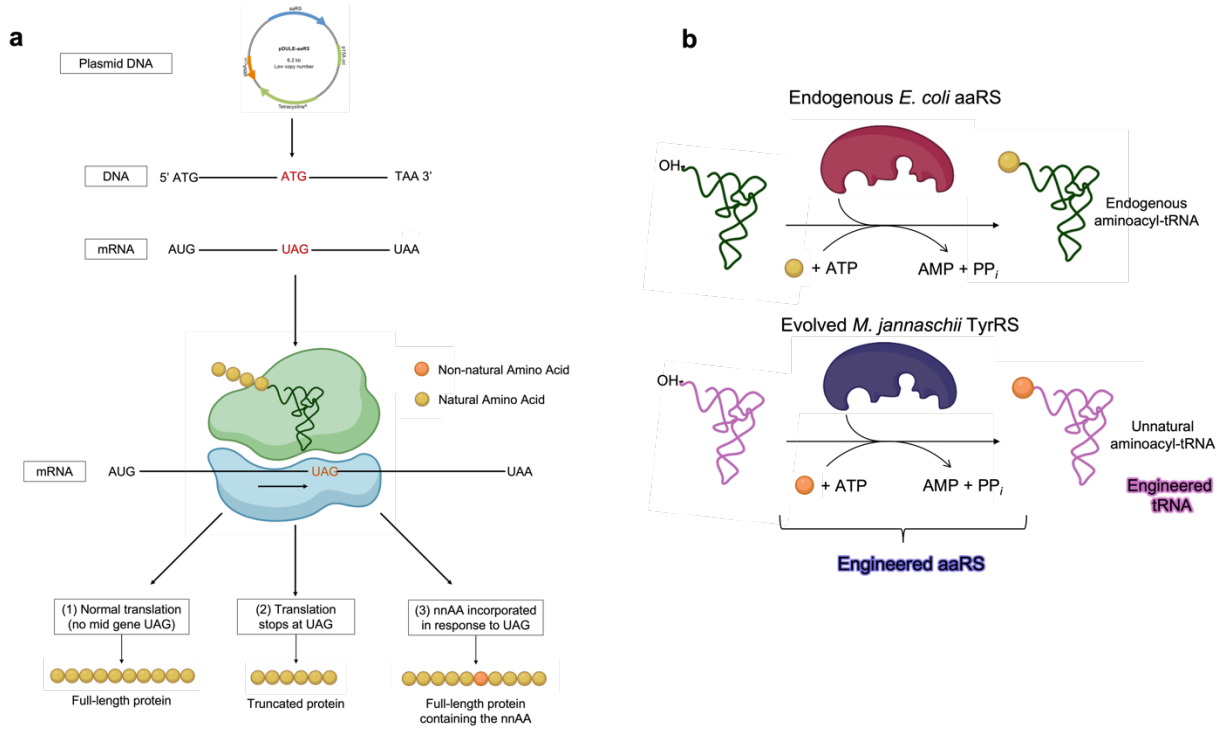

Figure 1.6: Non-natural amino acid incorporation: a) Different types of translation in (1) normal translation, (2) in the absence of nnAA and (3) in the presence of nnAA. b) Mechanism of incorporation in the presence of natural orthogonal tRNA/aminoacyl-tRNA synthetase pair compared to the engineered tRNA/aminoacyl-tRNA synthetase pair. Figure adapted from (Zitti and Jones 2023).

For the purpose of this thesis, the two nnAAs chosen were the para-cyano-phenylalanine (pCNPhe) and the para-ethynyl-phenylalanine (pCCPhe), due to their unique functional groups i.e. the nitrile and ethynyl respectively (Figure 1.7). Incorporating these nnAAs, aims to introduce unique Raman-active bonds that are not present in other biomolecules, which in turn allows them to be probed in Raman spectroscopy. Incorporating these Raman-active bonds will allow imaging of target proteins within the biologically silent window of the Raman spectrum, and more specifically between 2150-2250 cm^{-1} for pCNPhe and between 2100-2150 cm^{-1} for pCCPhe.

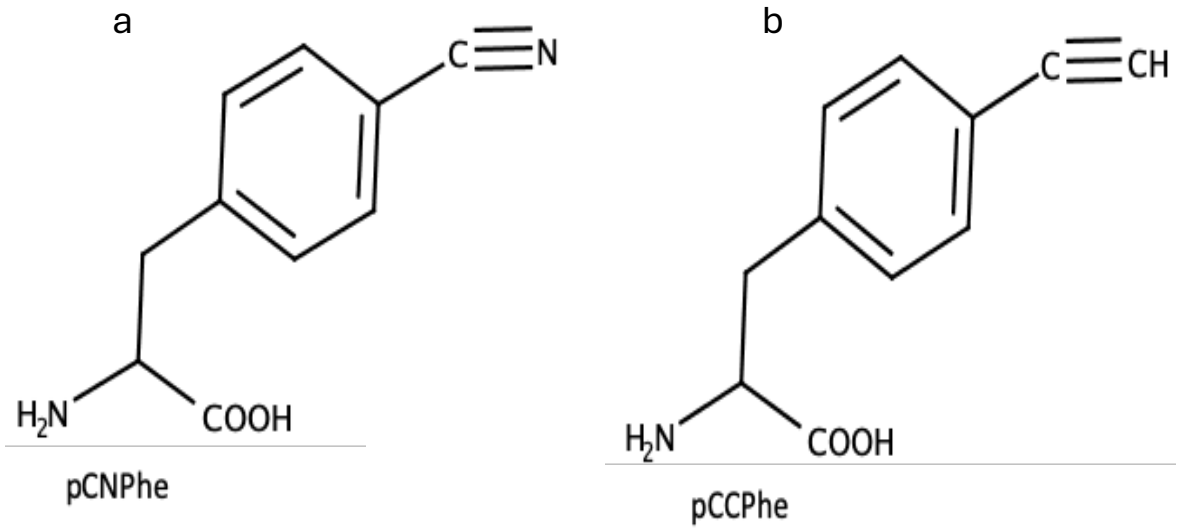

Figure 1.7: Chemical structure of the non-natural amino acids used in this thesis. a) pCNPhe (Pearson et al. 1996) and b) pCCPhe (Stokes et al. 2000).

For the context of this thesis, two approaches were explored to enhance the Raman-active bonds within the non-naturals through epr-RS. The first one, involves the incorporation of the nnAAs within the chromophore of the target fluorescent proteins. By employing this method, we aimed to replace the tyrosine at residue 67 of the chromophore with our target nnAAs. Similar approach to this was previously carried out by (Reddington et al. 2015; Hartley et al. 2016a) which showed that nnAAs can successfully be incorporated without compromising the function of the target fluorescent protein. Moving on, the second approach was by incorporating the nnAAs at key residues in close proximity around the

chromophore environment. The reason for this was to ultimately allow for through-space enhancement of the Raman signal. This would potentially be achieved through electronic transition coupling from the conjugated bond system of the chromophore to the Raman-active bonds present in the nnAAs. This coupling occurs when the electronic transition dipole moment of the chromophore interacts with the vibrational modes of the chemical bonds (Dietzek et al. 2018). This has been shown previously when engineering green FPs to yellow FPs, with π stacking of a second aromatic group with the chromophore results in a red shift in spectral properties (Ormö et al. 1996; Nagai et al. 2002). Vibronic coupling mechanisms that modulate electronic energy levels can effectively bridge electronic and vibrational states, thereby enhancing the Raman scattering signal. By strategically positioning non-canonical amino acids around the chromophore environment, it is possible to influence adjacent Raman-active vibrational modes through electronic transitions of the chromophore (García-González et al. 2024).

1.3 Molecular Biosensors

Fluorescent protein-based molecular biosensors have emerged as indispensable tools in modern biology, offering real-time, non-invasive, and highly specific detection of molecular events within living systems. Derived primarily from GFP and its engineered variants, such biosensors leverage the intrinsic fluorescence of proteins to monitor biochemical changes with high spatial and temporal resolution. By coupling the fluorescence of these proteins to conformational changes, molecular interactions, or the binding of specific analytes, such sensors enable highly sensitive and selective monitoring of cellular activity in real time (Fu et al. 2005; Greenwald et al. 2018).

Several design strategies have been developed to harness the optical properties of fluorescent proteins for biosensing. FRET sensors, for example, utilize the distance-dependent energy transfer between a donor and acceptor fluorophore to detect molecular interactions and conformational changes, making them ideal for tracking dynamic cellular signalling events such as kinase activity or fluctuations in ion concentrations (Hochreiter et al. 2015; Miyawaki and Niino 2015). Single fluorescent protein indicators, such as GCaMP

for calcium detection, encode conformationally sensitive domains within a single FP scaffold to produce robust fluorescence changes upon analyte binding (Tian et al. 2009). Other approaches, including bimolecular fluorescence complementation (BiFC) (Kerppola 2008), enable the detection of protein–protein interactions through the reassembly of split FP fragments (Kim et al. 2021), while environment-sensitive variants like pHluorin report changes in pH or ionic environments (Sankaranarayanan et al. 2000).

The benefits of these tools are substantial. By allowing for genetically targeted expression, FP-based biosensors enable highly specific and minimally invasive observations within live cellular or tissue contexts. Compared to traditional synthetic dyes, these genetically encoded sensors can be expressed stably and targeted to precise cellular compartments, making long-term studies feasible. Additionally, using proteins as biosensor components has many advantages. Proteins are low cost in design and synthesis and can easily have their properties altered and improved by genetic manipulation (Strianese et al. 2012). However, despite their versatility, they face several challenges in a similar way to that of fluorescent microscopy, like photobleaching and interference from autofluorescence.

Today, FP-based molecular biosensors play a pivotal role across disciplines, from elucidating fundamental cellular signalling pathways and protein dynamics to facilitating high-throughput screening for drug discovery. In fields as diverse as neuroscience, immunology, metabolism, and synthetic biology, these sensors enable direct observations of cellular behaviour with remarkable precision (Lavis 2021). The future of the field is poised for further expansion, with advances such as near-infrared and red-shifted FP variants allowing deeper tissue imaging, the engineering of sensors for new analytes, and the integration of molecular sensors with optogenetic tools for spatiotemporal control (Rodriguez et al. 2017). Together, these advances highlight the transformative potential of FP-based molecular biosensors as a central technology for understanding and manipulating biological systems.

In a similar manner to the rest of the rationale of this thesis, engineering of fluorescent proteins to improve their spectral properties or alter their properties to allow

sensitivity to changes in their environment that can be reflected in their spectral characteristics can be a pillar in generating an improved generation of fluorescent-based biosensors with more targets like changes in pH (Burgstaller et al. 2019) or ion concentration (Wu et al. 2022) as well as increasing the repertoire of available FPs aside from primarily using GFP.

1.4 Aims and Objectives

The overarching theme of this thesis is the use of protein engineering to advance the understanding of fluorescent protein structure-function relationship and their underlying properties after engineering, as well as to generate novel fluorescent proteins with the ultimate goal to be used as novel imaging probes. The thesis is broken down into three chapters, each addressing a unique research objective that ultimately builds up to the wider aim which is to improve current imaging methods as well as generate FPs with better spectral properties.

In Chapter 3, the focus is on engineering mCherry and mNeptune for the incorporation of the nnAAs within or in proximity to the conjugated system of the chromophore to create an enhanced Raman signal. This chapter establishes a foundation on elucidating how drastic changes within the chromophore influence the spectral properties and overall function of the target proteins. Limitations faced during the research contacted as part of this chapter, allowed for the design of the study described in chapter 4.

In Chapter 4, the focus shifts from the incorporation of non-natural amino acids to the incorporation of a cysteine within the chromophore. The aim was again to create a probe with an enhanced Raman signal, allowing a comparison between the two types of probes and making a probe more appropriate for *in vivo* live imaging. Spectral analysis will be used to identify the success of the probe for Raman spectroscopy, and analysis of the structure will help to further understand how these spectral properties emerge. Chapter 4 also demonstrates that introducing a cysteine into the chromophore generates useful spectral switching properties allowing detection of distinct pH profiles. This variant also

has improved resistance to ROS such as hydrogen peroxide. The structure of the variant was determined by X-ray crystallography and molecular dynamics performed to understand the molecular basis of the observed spectral properties.

Finally, Chapter 5 moves to a different topic, and mostly focuses on the engineering of a key residue for sfGFP's spectral properties, H148. This chapter builds on research contacted by previous members of the Jones lab, to further understand and create a holistic image on the importance of residue 148 on the spectral and physical properties of sfGFP. Computational methods employed, helped aid in the identification of the most viable residues for mutagenesis as well as provide dynamic inside on how the chromophore will interact with the change on the residue and the rest of its environment. Analysis of spectral properties together with MD analysis helped elucidate the underlying mechanism of why we see enhancement in sfGFP fluorescence and improved photobleaching ability.

2 Materials and Methods

2.1Materials

2.1.1Laboratory Chemicals

A range of chemicals were used and are listed throughout the thesis. Stock solutions were prepared by dissolving chemicals in deionised water (dH_2O) unless specified otherwise. All laboratory chemicals used are detailed on Table 2.1.

Table 2.1: Routinely Commonly used laboratory Chemicals chemicals and reagents.

Chemical	Specification/Preparation	Product Source
Acetic Acidacid	≥ 99.5 % purity	Scientific Laboratory Supplies (SLS)
Acrylamide: Bis-Acrylamide	40 % (w/v) Acrylamide: Bis-Acrylamide 29:1	Geneflow
Agarose	Molecular Biology Grade	Appleton Woods
Ammonium persulphate (APS)	Prepared as 10 % (w/v) stock	Melford
L-Arabinose	≥ 98 % purity	Sigma-Aldrich
B-mercaptoethanol	≥ 99 % purity	Sigma-Aldrich
Bromophenol blue	Technical grade	Sigma-Aldrich
Coomassie blue	Coomassie Brilliant Blue R250	Thermo Fisher Scientific
Deoxyadenosine triphosphate (dATP)	Prepared as a 10 mM stock	Promega
Deoxycytidine triphosphate (dCTP)	Prepared as a 10 mM stock	Promega
Deoxyguanosine triphosphate (dGTP)	Prepared as a 10 mM stock	Promega
Deoxythymidine triphosphate (dTTP)	Prepared as a 10 mM stock	Promega
DTT (dithiothreitol)	Technical grade	Thermo Fisher Scientific
Ethanol	≥ 99.8 % purity	VWR

Ethidium Bromide	Prepared as 10 mg mL^{-1} stock; 95 % purity	Sigma-Aldrich
D-Glucose	Anhydrous	Melford
Glycerol	Analytical grade	Thermo Fisher Scientific
Glycine	Glycine free base	Melford
Hydrogen peroxide	9.8 M (30 % solution)	Fisher Bioreagents
Imidazole	≥ 99 % purity	Sigma-Aldrich
Isopropanol	≥ 99.5 % purity	Thermo Fisher Scientific
Lactose	≥ 98 % purity	Sigma-Aldrich
LB Agar		Neogen
Magnesium sulphate	Analytical grade	Thermo Fisher Scientific
Methanol	≥ 99.8 % purity	Thermo Fisher Scientific
Sodium Chloride	≥ 99.5 % purity	Thermo Fisher Scientific
Sodium dodecyl sulphate (SDS)	Prepared as 10 % (w/v) stock	Melford
Sodium hydroxide	Analytical grade	Thermo Fisher Scientific
Sodium phosphate monobasic	Monohydrate; ≥ 98 % purity	Melford
Sodium phosphate dibasic	Anhydrous; ≥ 98 % purity	Melford
Tetramethylethylene-diamene (TEMED)	≥ 99 % purity	Melford
Tris(hydroxymethyl) aminomethane hydrochloride (Tris-HCL)	Molecular Biology Grade Ultrap Pure	Melford
Tris(hydroxymethyl) aminomethane base (Tris-base)	Molecular Biology Grade Ultra Pure	Melford

A range of media, buffers and solutions were prepared for the following methodologies. Their composition is detailed in Table 2.2 with respective pH and sterilisation requirements. All solutions were prepared with dH_2O unless specified otherwise.

Table 2.2: Composition of routinely used buffers and solutions. If no pH adjustment was required, '~' is denoted. Buffers purchased from external companies indicated by ('product source') confidential materials / quantities denoted by '*'.

Buffer / Solution	Composition	pH
Autoinduction bacterial growth media (1L) (Studier 2005)	1x 2xYT broth 0.05 % (w/v) L-arabinose 0.05 % (w/v) glucose 0.2 % (w/v) lactose 0.5 % (v/v) glycerol 50 mM ammonium chloride 25 mM sodium phosphate dibasic 25 mM potassium phosphate monobasic 5 mM sodium sulphate 2 mM magnesium sulphate	7
Coomassie destain buffer	40 % (v/v) methanol 10 % (v/v) acetic acid	~
Coomassie stain buffer	40 % (v/v) methanol 10 % (v/v) acetic acid 0.1 % (w/v) R250 Coomassie blue	~
CutSmart buffer 10X (New England Biolabs)	50 mM potassium acetate 20 mM Tris-acetate 10 mM magnesium acetate 100 µg mL^{-1} recombinant albumin	7.9
DC Protein Assay Reagent A	Alkaline copper tartrate solution	13.6
DC Protein Assay Reagent B	Dilute Folin reagent	<2.1

Gel Loading Dye no SDS 6x (New England Biolabs)	60 mM EDTA 19.8 mM Tris-HCl 15 % Ficoll®-400 0.12 % Dye 1 0.006 % Dye 2	8
Lysogeny broth (LB) agar (Neogen)	10 gL^{-1} tryptone 5 gL^{-1} yeast extract 10 gL^{-1} sodium chloride 15 gL^{-1} agar	~
NEBuffer™ r3.1		
Ni affinity elution buffer	500 mM iImidazole 50 mM Tris-HCl	8
Ni affinity elution buffer	10 mM iImidazole 50 mM Tris-HCl	8
N3 buffer (Qiagen)	4.2 M guanidium hydrochloride 0.9 M potassium acetate	4.8
Ni affinity elution buffer	500 mM Imidazole 50 mM Tris-HCl	8
Ni affinity elution buffer	10 mM Imidazole 50 mM Tris-HCl	8
P1 buffer (Qiagen)	50 mM Tris-HCl 10 mM EDTA 100 µg mL^{-1} RNase A 0.1 % (v/v) LyseBlue	8
P2 buffer (Qiagen)	200 mM sodium hydroxide 1 % (w/v) SDS	~
PB buffer (Qiagen)	5 M guanidium hydrochloride 30 % (v/v) isopropanol	~
PE buffer (Qiagen)	100 mM sodium chloride 10 mM Tris-HCl 80 % (v/v) ethanol	7.5

Q5 Reaction Buffer 5X (New England Biolabs)	Confidential*	8.5
Q5 High GC Enhancer 5X (New England Biolabs)	Confidential*	
Quick Ligase reaction buffer 2X (New England Biolabs)	66 mM Tris-HCl 10 mM magnesium chloride 1 mM dithiothreitol 1 mM adenosine triphosphate 7.5 % (v/v) polyethylene glycol; PEG600	7.6
SDS-PAGE loading buffer	0.2 % Tris-HCl pH 6.8 10 % (v/v) β-mercaptoethanol 8 % (w/v) glycerol 2 % (w/v) SDS 0.04 % (w/v) bromophenol blue	~
SDS-PAGE running buffer	192 mM glycine 25 mM Tris base 0.1 % (w/v) SDS	8.3
TAE buffer	40 mM Tris-acetate 1 mM EDTA	8.8
10x TB buffer	170 mM KH_2PO_4 720 mM K_2HPO_4	~
10x TBS Stock solution	200 mM Tris 1.5 M NaCl	7.4
TB Media	12 gL^{-1} tryptone 24 gL^{-1} yeast 4 mL glycerol 100 mL 10x TB buffer	~
Trace metals 5000X	50 mM iron (III) chloride 200 µM calcium chloride 100 µM manganese chloride 100 µM zinc sulphate	~

	20 µM cobalt chloride 20 µM copper chloride 20 µM nickel chloride 20 µM sodium selenite 20 µM sodium molybdate 20 µM boric acid	
Tris-HCl buffer	50 mM Tris-HCl/ Tris-base	8
2x YT broth (Thermo Fisher Scientific)	16 gL^{-1} tryptone 10 gL^{-1} yeast extract 5 gL^{-1} sodium chloride	7

Antibiotics were prepared as 1000x stock solutions, sterilised by filtration through 0.22 µM cellulose acetate syringe filters (Sartorius) and stored at -20°C. Detailed description of the antibiotics and working concentrations used can be found in Table 2.3.

Table 2.3: Antibiotics

Antibiotic	Working Concentration	Solvent	Supplier
Ampicillin	100 µg/mL	dH_2O	Melford
Kanamycin	50 µg/mL	dH_2O	Melford
Tetracycline	10 µg/mL	EtOH (70%)	Melford
Spectinomycin	50 µg/mL	dH_2O	Melford

2.1.2 Nucleic acid modifying enzymes

Commercially available enzymes were used for as part of various DNA manipulation methods. Table 2.4 details the specific activity of these enzymes and the company they were supplied by.

Table 2.4: Enzymes' specifications

Enzymes	Activity	Supplier
DpnI	20000 units/mL	New England Biolabs
Q5 High-Fidelity DNA Polymerase	2000 units/mL	New England Biolabs
Quick Ligase	Confidential	New England Biolabs
RNase A	5000 units/mL	Generon
T4 Polynucleotide Kinase (PNK)	10000 units/mL	New England Biolabs

2.1.3 Bacterial strains and growth media

Bacterial liquid growth media 2xYT (2x Yeast Extract Tryptone) medium (Melford), solid media lysogeny broth (LB) Agar (Melford) (bought as powder) were prepared as per the manufacturer's instructions, and sterilised by autoclaving at 121°C for 15 minutes. LB Agar was left to cool down before the required antibiotic was supplemented, poured into plates and stored at 4°C degrees once setuntil their use. Autoinduction media prepared as detailed in Table 2.2 was used to produce proteins encoded in the pBAD plasmid.

2.1.4 Bacterial strains

Transformation and protein expression were performed using two *E. coli* strains: Top10™ (Invitrogen, ThermoFisher) and BL21 (DE3) (NEB). *E. coli* Top10™ cells were used for plasmids with the arabinose-inducible promoter pBAD plasmid system, whereas *E. coli* BL21 (DE3) cells for were used for expression of genes in pET-based plasmids with a T7 promoter. The genotypes and the proteins expressed in each bacterial strain are summarised in Table 2.5.

Table 2.5: *E. coli* strains

***E. coli* Strain**	**Genotype**	**Proteins Expressed**	**Supplier**
One Shot™ TOP10 *E. coli*	F⁻*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ⁻*rps*L(Str^R) *end*A1 *nup*G	sfGFP mCherry mNeptune	ThermoFisher Scientific
BL21 (D3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	NB56 Nano14 Nano85	New England Biolabs

2.2 Molecular Biology

2.2.1 Side-directed mutagenesis (SDM)

Mutants were generated using the primers detailed in Table 2.6. Primers were designed using Amplifx and SnapGene Viewer and later synthesized by Integrated DNA Technologies. The construction of mutants was performed using inverse (also known as whole plasmid) PCR-based site directed mutagenesis. Annealing temperatures were determined using NEBTmCalculator (https://tmcalculator.neb.com). Primers used to enable sequencing of gene constructs are also detailed in Table 2.6.

Site-directed mutagenesis was performed via PCR amplification to introduce specific mutations at target residues detailed on Table 2.6. The Q5 High-Fidelity DNA polymerase kit (New England Biolabs) was used. The reaction protocol and setup are summarised in Table 2.7.

Prior to protein expression, selected bacterial cultures were used to purify DNA (Method 2.2.1.3) to be sent for sequencing to ensure the presence of the desired mutation. DNA samples were sent to Eurofins Genomics through the TubeSeq service for Sanger Sequencing. Samples for sequencing were prepared by mixing 18 µL of plasmid DNA with 2 µL of sequencing primer. Plasmid DNA was used at a final concentration of 50 – 100 ng/µL.

2.2.1.1 Vectors for protein expression

Various genetic constructs (Figure 2.1) were used to host the genes for recombinant protein expression in *E. coli* and protein expression in mammalian cells (HeLa cells). Fluorescent proteins including mCherry, mNeptune and sfGFP were expressed using a pBAD vector. Expression was induced under the arabinose-inducible araBAD promoter (P_{BAD}) system, which also contained a selective ampicillin resistance gene. Wild type mNeptune [Addgene: #54714] was acquired from Addgene while mCherry and sfGFP constructs were kindly donated by the Jones lab at Cardiff University.

A secondary plasmid vector (pDULE) is required for the incorporation of non-canonical amino acids (ncAA). The pDULE [Addgene: #85495] system contains an orthogonal tRNA/aminoacyl tRNA synthetase pair that was engineered to incorporate the required ncAA in response to the amber stop codon (UAG). The pDULE system was co-transformed to incorporate 4-cyano-phenylalanine (pCNPhe) and ethyl-phenylalanine (pCCPhe). The pDULE plasmid contained a selective spectinomycin resistance gene.

Figure 2.1: Plasmid maps of the genetic constructs used for the expression of Fluorescent Proteins in bacteria. a) mCherry WT (donated by DDJ lab). b) mNeptune [Addgene: #54714]. c) pDule2 [Addgene: #85495]. Plasmid maps exported from SnapGeneViewer.

Table 2.6: Primer sequences and annealing temperatures (Tm) used in this thesis. Mutations are highlighted in red. The pBAD SeqP pair was designed by Samuel Reddington.

Primer	Sequence (5' – 3')	Tm (°C)
F_mCherry_M66C	GCTCCAAGGCCTACGTGAAG	69
R_mCherry_M66C	CGTAGCAGAACTGAGGGGACA	
F_mCherry_Y67TAG	GTCCCCTCAGTTCATGTAGGGCTC	69
R_mCherry_Y67TAG	AGGATGTCCCAGGCGAAGG	
F_mCherry_W58TAG	GCCCTTCGCCTAGGACATCCTGT	66
R_mCherry_W58TAG	AGGGGGCCACCCTTGG	
F_mCherry_Y120TAG	GCGAGTTCATCTAGAAGGTGAAGCT	69
R_mCherry_Y120TAG	CGTCCTGCAGGGAGGAG	
F_mCherry_W143TAG	CCATGGGCTAGGAGGCCTCCTC	69
R_mCherry_W143TAG	TCTTCTTCTGCATTACGGGGCC	
F_mCherry_E144TAG	GGGCTGGTAGGCCTCCTCCG	69
R_mCherry_E144TAG	ATGGTCTTCTTCTGCATTACGGGG	
F_mCherry_Y181TAG	CAAGACCACCTAGAAGGCCAAGAAGC	69
R_mCherry_Y181TAG	ACCTCAGCGTCGTAGTGGC	
F_mNeptune_Y67TAG	CTGCTTCATGTAGGGCAGCA	66
R_mNeptune_Y67TAG	GTAGCCAGGATGTCGAAGG	
F_sfGFP-CMV-F-Lifeact_H148S	CAACTTCAACAGCAGCAACGTCTATATCAC	69
R_sfGFP-CMV-F-Lifeact_H148S	TACTCCAGCTTGTGCCCCAGGATGT	
F_pBAD_SeqP	ATGCCATAGCATTTTTATCC	
R_pBAD_SeqP	GATTTAATCTGTATCAGG	
F_sfGFP-CMV-F-CytERM	CGGTAGGCGTGTACGGTGGGAGGTC	
F_sfGFP-CMV-F-Lifeact	CGCAAATGGGCGGTAGGCGTG	

Table 2.7: Polymerase chain reaction conditions. Adapted from the NEB #M0491 protocol.

Material				Volume in reaction (100 µL)	
5x Q5 Reaction Buffer				20	
DNTPs (10 mM)				2	
Forward Primer (10 µM)				5	
Reverse Primer (10 µM)				5	
Template Plasmid DNA (1 ng/µl)				1	
Nuclease Free Water				62	
5x Q5 High GC Enhancer				4	
Q5 DNA polymerase (2000 units/ml)				1	
PCR Conditions					
Initial Denaturation	30 Reaction Cycles			Final Extension	Hold
	Denaturation	Annealing	Extension		
98°C 30 seconds	98°C 10 seconds	Tm (°C)[a] 30 seconds	72°C 30 seconds/ Kb	72°C 2 minutes	4°C Indefinite

a, Annealing Tm calculated using NEB's online Tm calculator tool for each set of primers as outlined in Table 2.6.

2.2.1.2 PCR product purification

The QIAquick® PCR purification kit (QIAGEN) was used to remove residual impurities. The reaction mixturess were mixed with 5x PB buffer and then applied to the QIAquick spin column. The spin column was placed in a collection tube and centrifuged for 1 minute at 13500 rpm. This allows the DNA to bind to the silica membrane and the impurities to be discarded as part of the flow-through. The column was washed, with the addition of 750 µl of PE buffer. The sample was centrifuged twice, for 1 minute at 13500 rpm, discarding the flow-through each time. The DNA was finally eluted in nuclease free water (12 µL) by transferring the spin column to a sterile 1.5 ml microcentrifuge tube. The purified PCR product was then subject to phosphorylation and ligation steps to be prepared for cloning.

2.2.1.37 Plasmid DNA purification

The QIAprep Miniprep Kit (QIAGEN) was used to isolate and purify plasmid DNA from bacterial cell cultures. The cultures were grown overnight at 37°C, then centrifuged at 4000 rpm for 20 minutes using a Thermo Fisher Fiberlite F9-6 x 1000 LEX Fixed Angle Rotor. Cell pellets were resuspended in 300 µL of P1 buffer and transferred to a 1.5 mL centrifuge tube where 300 µL of P2 buffer was added. The solution was mixed by inverting the tube until the solution turned blue (LyseBlue reagent). Then 350 µL of N3 buffer was added, and the solution was mixed by inverting until the solution turned colourless. The solution was then centrifuged in a microfuge for 10 minutes at 13000 rpm. The supernatant was added to a QIAprep spin column and centrifuged for 1 minute at 13000 rpm. The spin column was washed with 750 µL of PE buffer and centrifuged twice at 13000 rpm for 1 minute, discarding the flow-through each time. The spin column was then transferred to a clean 1.5 mL microcentrifuge tube and 30 µL of nuclease-free water was added to the centre of the column and left to stand for 10 minutes. The samples were finally centrifuged for 1 minute at 13000 rpm to elute the pure plasmids. DNA was quantified using the NanoDrop ND 1000 Spectrophotometer (ThermoFisher).

2.2.1.4 DNA phosphorylation and ligation

Linear DNA produced by PCR was functionalised prior to cloning. Purified PCR product (12 µL) was mixed with 12 µL of Quick Ligase reaction buffer, 2 µL DpnI, 2 µL CutSmart buffer and 1 µL T4 PNK (Table 2.4), and left to incubate for 30 minutes at 37°C. Quick Ligase (1 µL) was added to the solution and left to incubate at room temperature for up to 8 minutes. The reaction was then stopped by another round of DNA purification (Methods 2.2.7) to remove the enzyme and buffer salts. DNA was quantified using the NanoDrop ND 1000 Spectrophotometer and purity checked via gel electrophoresis.

2.2.1.5 Non-canonical amino acids preparation

Non-natural amino acids were prepared by dissolving in minimal volume of 1M sodium hydroxide. Once the growth media reached optical density (OD_{600}) of 0.5, the required ncAA was added to a final concentration of 1 mM.

2.2.2 Agarose gel electrophoresis

For the analysis of DNA fragments, 1 % (w/v) agarose gels were prepared by dissolving 0.5 g agarose powder (Melford, Ipswich, UK) in 50 mL 1xTris-acetate-EDTA (TAE) buffer supplemented with ethidium bromide (Sigma Aldrich) to a final concentration of 0.3 mg/ml. Once the gel was set, it was submerged in a tank containing 1x TAE buffer. DNA products from site-directed mutagenesis and DNA purification were prepared by mixing with 1 µL of 6x purple gel loading dye (NEB) on Parafilm before loading. Samples were loaded in 5 µL volumes and DNA ladder (1 Kb plus; NEB) in a 2.5 µL volume. Electrophoresis was performed at 100 V for 45 minutes. PCR products were visualised under UV light using a GelDoc-It UV transilluminator (UV products Ltd.)

2.2.3 Gel excision

The desired DNA fragment was excised from the agarose gel using a clean, sharp scalpel. Buffer QG was added to the gel in a 3:1 ratio and incubated at 50°C until the gel completely dissolved. Once dissolved, 1 gel volume of isopropanol was added to the sample and was mixed. The mixture was then transferred to a QIaquick (QIAGEN) spin column and centrifuged for 1 minute to allow the DNA to bind onto the column. Another 500 µL of QG buffer were added and the sample was centrifuged again for 1 minute discarding the flow through. The sample was washed with 750 µL of PE buffer, while left to incubate for 5 minutes before centrifuging twice to make sure all buffer is removed. DNA was then eluted in 30 µL of nuclease-free water in a clean 1.5 mL tube.

2.2.4 *E. coli* Transformation

Fluorescent protein constructs in the pBAD plasmid were transformed into chemically competent *E. coli* TOP10™. Nanobody constructs in the pET plasmid were transformed into chemically competent *E. coli* BL21(D3). Transformations were carried out using a heat shock protocol adjusted according to the type of competent cells used. Competent cells were thawed on ice for 5 minutes, and 1 ng of purified plasmid DNA was then added to 50 µL of competent cells and incubated on ice for 30 minutes. The cells were then incubated in a 42°C water bath for 10/30 seconds (BL21(D3)/TOP10™), and then subsequently placed on ice for another 5/2 minutes (BL21(D3)/TOP10™). Super Optimal Broth (SOB) media (950 µL) was added, and the cells were allowed to recover at 37°C for 1 hour. The recovery media was then microfuged, and the cell pellet was resuspended in 50 µl of the remaining media. To select for cellular uptake of plasmid DNAs, the cells were spread onto a LB agar plate supplemented with the appropriate antibiotic. The plates were incubated overnight at 37°C. Single colonies were selected, picked and incubated in separate 5 mL 2xYT media aliquots overnight at 37°C supplemented again with the appropriate antibiotic. The overnight cultures were then used to inoculate large scale cultures (0.5 L) to be used for protein expression once mutation was confirmed.

2.3 Recombinant protein expression and purification

2.3.1 Recombinant protein expressionproduction

Wild type mCherry, mCherry^{M66C} and WT mNeptune were produced using a pBAD vector in transformed *E. coli* TOP10™ cells (Methods 2.2.10). A 5 mL overnight culture was used to inoculate a 0.5 L of 2xYT media supplemented with 1x ampicillin (Table 2.3). Cultures were left to grow at 37°C until they reached an OD_{600} of 0.6 where when 0.2 % (v/v) arabinose was added to induce expression. The culture was left to incubate overnight at 37°C, shaking at 200 rpm.

For the incorporation of pCNPhe/pCCPhe in mNeptune, co-transformation of the mNeptune:pBAD construct with the pDULE2 plasmid was employed. Colonies carrying both plasmids were grown in 5 mL overnight cultures and then used to inoculate 0.5 L 2xYT

media as mentioned above supplemented with 50 mg/mL kanamycin and 50 mg/mL spectinomycin (Table 2.3). The ncAA was added subsequently to arabinose. The culture was left to incubate overnight at 25°C while shaking at 200 rpm.

2.3.2 Fluorescent protein purification

For fluorescent proteins, cells were harvested via centrifugation at 5000 xg for 20 minutes at 4°C ThermoFisher Fiberlite F9-6 x 1000 LEX Fixed Angle Rotor. The cell pellet was resuspended in 20 mL of 50 mM Tris-HCl buffer (pH 8) and cells were lysed using a French press. The resulting lysate was clarified by centrifugation at 25000 xg for an hour at 4°C using a ThermoFisher Fiberlite F21-8 x 50 Fixed Angle Rotor where and the resulting supernatant was collected. Clarified cell lysates were loaded onto a 5 ml His-trap HP nickel affinity column (Cytiva) equilibrated in wash buffer (50 mM Tris-HCl, 10 mM iImidazole pH8.0) for Nickel-affinity chromatography. The column was washed with wash buffer allowing only for the target protein to bind. Bound protein was then eluted by washing the column in with 500 mM imidazole (pH8.0). Pooled protein samples were then subjected to Size Exclusion Chromatography (SEC) using a HiLoad™ 16/600 Superdex™ S75 pg column that was equilibrated with 50 mM Tris-HCl buffer (pH8.0). For proteins not carrying the photosensitive AzF, specific absorbance wavelengths based on the protein λ_{max} were used to detect protein elution. Purity of the proteins was then checked by SDS-PAGE analysis (Section 2.4.1).

2.4 Protein Analysis:

2.4.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse all the protein samples generated after purification steps. A Mini-PROTEAN^{R} Tetra cell tank was used to carry out electrophoresis. A 12.5 % resolving gel was prepared for fluorescent proteins. The 12.5 % resolving gel consisted of 0.1 % w/v SDS, 0.375 mM Tris-HCl (pH 8.8), 15% bis acrylamide (37.5.1 40 % solution), 0.05 % w/v ammonium persulphate (APS) and 0.02 % v/v tetramethylethylenediamine (TEMED). The 5% resolving

gel consisted of 0.2 % w/v SDS, 65 mM Tris -HCl (pH 6.8), 13 % bis acrylamide solution, 0.1 % w/v APS, and 0.02% v/v TEMED. The gels were prepared and setup according to BIORAD protocol (2011).

Samples were prepared using a 5xSDS-PAGE reducing buffer (Table 2.2) in a 2:1 ratio protein:buffer. Samples were then denatured for 10 minutes at 98°C. A protein molecular weight standard ladder was used as a reference (Colour Prestained Protein Standard, Broad Range, NEB). The gel was run at 180 V for 50 minutes.

Gels were then stained in the stain solution (Table 2.2) for at least thirty minutes. They were then destained in the destain solution (Table 2.2) and visualised using a GelDoc-It system (Bio-Imaging Systems).

2.4.2 Protein concentration

The concentration of purified proteins was determined using the DC Protein Assay (BioRAD) using the manufacturers guidelines. Either BSA or wt mCherry (0.2-1.5 mg/mL) was used as the concentration standard.

2.4.3 Fluorescence Spectroscopy

Absorbance and fluorescence excitation measurements were carried performed out using Carry 60 UV-Vis spectrophotometer (Agilent technologies) and Varian Carry Eclipse Fluorescence spectrophotometer (Agilent Technologies) respectively. Absorbance spectra were recorded using , a 1 cm path-length quartz cuvette. Protein fluorescence spectra were recorded using a 5x5 mm quartz cuvette and data was collected with a 5 nm slit width at a rate of 600 nm/min. Each protein was excited at their respective excitation maximum. For analysis of mCherry and mCherry^{M66C}, spectral scans were recorded at concentrations of 2.5 µM and 5 µM while excitation and emission spectra were adjusted accordingly.

2.4.4 Biochemical Assays

2.4.4.1 pH profile:

Wild type mCherry and mCherry^{M66C} were used at a 5 µM final concentration and transferred to buffers of different pH values (Table 2.8). Absorbance and emission were recorded as described in Section 2.4.3 for all samples.

Table 2.8: Buffers used for pH studies

Buffer	pH value
100 mM acetate buffer	4.5
100 mM acetate buffer	5.0
100 mM acetate buffer	5.5
100 mM KH_2PO_4-NaOH	6.0
50 mM Tris-HCl	8.0
100 mM HEPES	7.0
100 mM glycine-NaOH	9.0
100 mM glycine-NaOH	10.0

2.4.4.2 Oxidation potential:

Wild type mCherry and mCherry^{M66C} were used at a 2.5 µM final concentration and mixed with increasing concentrations of (hydrogen peroxide) H_2O_2 ranging from 0.01 %-5 %. Protein samples at both pH 7 and pH 8 were used. Absorbance and Fluorescence emission were recorded at 0 minutes of incubations and at 1 hour of incubation. Baseline correction was performed on absorbance spectra using Origin24 (OriginLab Corporation).

2.4.5 Mass Spectrometry

Purified mCherry and mCherry^{M66C} were diluted to 10µM and analysed at the Mass Spectrometry facility in the School of Chemistry, Cardiff University. The sample was subject to liquid chromatography-mass spectrometry (LCMS) where 5 µL of sample was injected into a Waters Acquity UPLC/Synapt G2-Si QTOF mass spectrometer. The column used was Waters Acquity UPLC protein BEH C4 300 Å, 1.7 µm (2.1 x 100 mm) kept at 60 °C.

The flow rate was set to 0.2 mL/min, with an increasing gradient concentration of eluent B (0.1 % formic acid in acetonitrile); 5 % for three minutes, followed by an increase to 65 % for fifty minutes, an increase to 97 % for fifty-two minutes with a final hold for two minutes. The data were then analysed using Waters Masslynx 4.1.

2.5 Protein Crystallography

2.5.1 mCherry^{M66C} Crystal formationcrystallisation.

Purified protein samples in 50 mM Tris-HCl were concentrated by via centrifugation at 3500 rpm in using 30 kDa concentrating columns (Merck) until a final concentration of ~10 mg/ml was reached. Crystal trials were set up at Diamond Light Source (DLS) under the supervision of Dr. Halina Mikolajek . Crystal formation was screened using sitting drop vapour diffusion across a wide variety of conditions as described by the BSC, PACT *premier*™*,* JCSG-*plus*™, SG1™ and Morpheus® crystallisation screens (Molecular Dimensions, Suffolk, UK). A multichannel pipette was used to aliquot (20 µl) of the crystallisation conditions into the 96-well plate. Drops were set up with various protein:crystallisation buffer ratios including 1:1, 1:2 and 2:1 and left to incubate at room temperature until crystals formed. The process of incubation and data acquisition is completely automated at DLS. Trays have a barcode that is scanned and all the information regarding which protein and the specific crystallisation conditions are automatically entered onto the DLS database. Trays are monitored for crystal formation and pictures are regularly taken to track crystal formation. Once crystals are formed, trays can be requested to be x-rayed by the beamline.

2.5.2 X-ray Crystallography diffractionsof mCherry^{M66C}

Crystals observed from the trays mentioned in 2.5.1 were later passed through the I03 beamline for x-ray scattering and diffraction. The I03 is a macromolecular crystallography (MX) beamline at DLS; it is a tuneable beamline with a working wavelength range of 0.6 – 2.48 Å. The standard working wavelength is 0.976 Å (12.7 keV) with a focused beam size of 90 x 30 um (FWHM). Following data acquisition (diffraction patterns) from

DLS, the structure of mCherry^{M66C} was solved with molecular replacement using PHASER. Molecular replacement was performed using the structure of wt mCherry (PDB: 2HQ5) as a model, while incorporating the appropriate chromophore including the cysteine mutation at residue 66. The PDB entry 2H5Q was the search object for this step. PHASER confirmed there was only one copy in the asymmetric unit producing a large Log likelihood score of 6303 and TFZ (Twin Fraction Zero refinement) of 26, indicating high confidence in the result. The associated map clearly indicated where there were differences with the model, when displayed in COOT. The model was adjusted to match the correct sequences, and to fit it into the electron density map. Subsequent iterative refinement with REFMAC5 and model adjustment with COOT allowed a better fit to the map, after several cycles. Residues 4 to 223 (C-terminus) were visible in the map, including a SO_4 molecule from the crystallisation medium, and 87 water molecules. TLS (Translation/Libration/Screw-rotation) refinement was employed during REFMAC5 runs to allow improved accounting for thermal motion in the model. The final refinement residuals were R/R-free of 0.17/0.21. The final model and electron density map representations were made using PYMOL.

2.6 Computational modelling and analysis

2.6.1 *In silico* modelling

In silico modelling and analysis of all the protein structures used for this thesis were carried out using the molecular visualisation graphics software PyMOL (Schrödinger and Delano 2015). All input files were in the .pdb format, and they were either obtained from the Protein Data Bank (Berman 2000; Berman et al. 2003) or generated after structure determination.

2.6.2 Computational requirements

Molecular Dynamics (MD) simulations required high performance computation resources, provided by the Supercomputing Wales HAWK server. The HAWK server contains 280 nodes, 12736 cores and 69 TB of total memory. Both CPU (AMD(R) Epyc 7502 2.5 GHz with 32 cores each) and GPU (Nvidia V100 GPUs with 16 GB of RAM on 15 nodes)

was utilised for this research. The GROMACS (Groningen MAchine for Chemical Simulations) software package was used for MD simulations (Van Der Spoel et al. 2005; Pronk et al. 2013; Abraham et al. 2015). MD simulations were performed with the help of Prof Dafydd Jones and Dr Georgina Menzies. For the MD simulations of mCherry, input structures for the wt protein (PDB:2H5Q) were retrieved from the Protein Data Bank repository (Berman 2000; Berman et al. 2003)

2.6.3 Parametrisation of mCherry

The CHARMM27 force field was employed to parametrise the chromophore of mCherry and mCherry^{M66C} (Mackerell et al. 2004; Bjelkmar et al. 2010). For parametrisation, the force field parameters need to be generated, including atomic changes bond lengths, angles, dihedrals, improper and non-bonded interactions, that accurately describe each compounds' chemical properties. Initial geometries were obtained from x-ray crystallographic data, with atomic charges derived using Atomic Charge Calculator 2 (Raček et al. 2020). The CHARMM27 forcefield function (Equation 1) is used to describe the molecular mechanics for each parametrisation.

Equation 1

$$V(r) = \sum_{bonds} k_b(b - b_o)^2 + \sum_{angles} k_\theta(\theta - \theta_o)^2 + \sum_{dihedrals} k_\chi[1 + cos(n\chi - \delta)] + \sum_{impropers} k_\omega(\omega - \omega_o)^2 + \sum_{Urey-Bradley} k_\upsilon(\upsilon - \upsilon_o)^2 + \sum_{nonbonded} \varepsilon_{LJ}\left[\left(\frac{R_{min,ij}}{r_{ij}}\right)^{12} - \left(\frac{R_{min,ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\varepsilon r_{ij}}$$

The first part of the equation describes the potential energy (v) of a molecular system which is the summation of internal and non-bonded interactions. Bonds (b), valence angles (θ), dihedral angles (χ), improper dihedral angles (ω) and Urey-Bradley (υ) together they define the internal interactions and referred to as the internal terms. Each

internal interaction is associated with a force constant ($k_b, k_\theta, k_\chi, k_\omega$ and k_u) which defines the energy contribution (kcal/mol) per unit deviation (Å^2, radian^2) from equilibrium values. The potential energy contributions from bond, angles, dihedral angles, improper dihedral angles and Urey-Bradley are all modelled as quadratic functions. They are modelled in that way to reflect the deviation from equilibrium values (b_O, θ_O, υ_O, *and* ω_O).

The distance between two bonded atoms is compared against its equilibrium distance to give the bond lengths, while the angle term describes the angle between two bonds that connect three atoms, measured against its equilibrium value. The Urey-Bradley function accounts for the distance between the outer atoms in a three-atom sequence and it's equilibrium distance. Improper dihedrals maintain planarity by measuring the angle between the planes formed by four atoms (l, j, k and i), where the three peripheral atoms (l, j and k) are bonded to a central atom (i), compared to its equilibrium value. In the dihedral term, multiplicity and the phase are represented by n and δ respectively and dictate the location of the minima and maxima. In CHARMM the phase can be set to any value, and when combined with the expansion of the dihedral term into a Fourier series through the multiplicity, it allows for extended flexibility around bonds (Blondel and Karplus 1996). Rotational freedom is expressed by the dihedral energy term that depends on the angle between planes which are defined by four consecutively bonded atoms (Smith and Karplus 1992).

Non-bonded interactions are comprised by the Coulomb and Lennard-Jones (LJ) terms. The Lennard Jones potential describes van der Waals interactions where LJ well-depth is represented by ε and the distance at the LJ minimum interaction energy is expressed by R_{min}. The Coulomb's law, where q_i and q_j are the partial atomic charges, r_{ij} is the distance between atoms i and j respectively and ε is the dielectric constant, is used to calculate electrostatic interactions.

Parameter values, namely force constant and bonded parameters, were calculated using the AnteChamber Python Parser interfacE (ACPYPE). ACPYPE was used to automatically generate the topology and parameter files to be used with CHARMM27 (Sousa da Silva and Vranken 2012). Chromophore parametrisation for wt mCherry and

mCherry^{M66C} were done by Dr Ozan Aksakal (Aksakal (2025)) and Prof Dafydd Jones respectively. The chromophore structure was parametrised to include residue F65 to include the additional extended double bond network. Residue F65 is not part of the chromophore forming triad.

2.6.4 Input structure preparation.

Input models were converted from the PDB format to the GROMACS-compatible Gromos87 (.gro) files, where hydrogens were added. All crystallographic waters were retained. CHARMM27 forcefield parameters that were modified as described in Section 2.6.3 were used to inform GROMACS on the topology of the new amino acid within the mCherry chromophore. The parametrisation ensured accurate representation of the molecular structures within the simulations. Each model was placed in a cubic box with dimensions set to 7.56898 for mCherry and 7.47605 for mCherry^{M66C} between atoms in the structure in order to prevent boundary effects. Then the simulation box was solvated with a CHARMM-implemented TIP3P water model. The overall charge of each system was calculated, and Na^+ ions were added to neutralise the system.

2.6.5 Energy minimisation

To achieve the proteins most stable conformation, energy minimisation was performed. Steric clashes were relieved and potential energy was lowered using the steepest descent algorithm with a step size of 0.01 nm and a maximum force threshold of 1000 KJ/mol/nm to optimise the potential energy. Using a Verlet cut-off scheme the interactions between neighbouring atoms were calculated during minimisation, with a neighbour list updated at every step. The Particle-Mesh Ewald (PME) algorithm was used to determine total electrostatic interactions with a 1 nm cut-off for Coulomb interactions and up to 1nm cut-off for van der Waal interactions. Finally, Periodic Boundary Conditions (PBC) were applied in all three dimensions (*xyz*) to simulate an infinite system.

2.6.6 Two-step equilibration

Following energy minimisation, temperature, pressure and density were controlled via a two-step equilibration using a leap-frog integrator. Each equilibration step was conducted for 100 ps with a 2 fs time step. Further bond parameters were included where holonomic constrains were applied using the LINCS algorithm. The LINCS algorithm ensures accurate correction of bond length during rotational motions by resetting bonds to their correct length (Hess et al. 1997). Constraints were applied to all bonds involving hydrogen atoms. Atom neighbour finding was carried out using the Verlet cut-off scheme, and PME was used for electrostatics with a 0.16 nm grid spacing for Fourier transformations. Velocity rescaling for temperature coupling is used with a stochastic distribution for both protein and non-protein groups. To maintain the temperature at 300 K, mimicking biological conditions, temperature coupling was applied with a time constant of 0.1 ps.

For the second equilibrium step, isotropic pressure coupling was included using the Parrinello-Rahman barostat, allowing uniform scaling of the simulation box dimensions. Pressure coupling was employed with a 2 ps time constant and a reference pressure of 1 bar. Finally, density values were verified before proceeding to the production phase.

2.6.7 MD production run

The equilibrated systems were used as inputs for the final production run, maintaining the same parameters as used during equilibration. Each production simulation was conducted for 100 ns with a 2 fs time step. A total of three independent production runs were performed for each protein variant. Trajectory files for each simulation were converted so that the proteins were centred in the water box for the following analysis. MD simulations were performed with the help of Dr Georgina Menzies (mCherry) and Prof Dafydd Jones (mCherry^{M66C}).

2.6.8 Calculating Root Mean Square Deviations (RMSD)

RMSD provide a measure of how much a structure deviates from a reference structure over time by using the 'gmx rms' command which is calculated using Equation 2. The first frame of the trajectory is used as the reference structure. The deviation for the whole protein RMSD calculations, is measured using the Cα atoms, while all atoms were used for the chromophore ones. At each given time point (t), every atom has a position of r_i, in respect to the reference coordinate of the same atom r_i^{ref} and N is equal to the number measured atoms, while m_i is the mass of the atom measured.

Equation 2

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} m_i \left| r_i(t) - r_i^{ref} \right|^2}$$

2.6.9 Calculating Root Mean Square Fluctuations (RMSF)

The RMSF is a measure of the fluctuations of individual atoms or residues with respect to their reference structure using the 'gmx rmsf' command and is calculated with Equation 3. In a similar manner to RMSD calculation, for whole protein RMSF calculations, the Cα was selected to measure fluctuations of the residues while all heavy atoms (no hydrogens), were selected for the chromophore analysis. At any given time point (t), each atom (i) has a position of r_i that is relative to the reference position of the measured atom r_i^{ref}. The total number of time points (T) is used to average and normalise all given time points through the entire simulation.

Equation 3

$$RMSF(i) = \sqrt{\frac{1}{T}\sum_{t=1}^{T} \left| r_i(t) - r_i^{ref} \right|^2}$$

The RMSF values where averaged across the 3 simulations and the standard deviation calculated.

2.6.10 Hydrogen bonding and pairwise distance analysis.

The number and frequency of hydrogen bonds was calculated on specified cut-offs for both distance and angle using the 'gmx hbond' command. Hydrogen bonds were defined by a cut-off distance of 0.35 nm between the donor and acceptor atoms and an angle cut-off of 30° for the hydrogen-donor-acceptor configuration. The hydrogen atoms of OH and NH groups were treated as hydrogen bond donors, and oxygen and nitrogen atoms were regarded as acceptors.

2.7 Cell Imaging

2.7.1 Mammalian cell constructs

Two mammalian cell genetic constructs were acquired from Addgene in order to fuse sfGFP to cellular proteins for production in HeLa cells. The fusion partner was LifeAct (F-actin binding peptide), that binds specifically to filamentous actin. and CytERM-N-17 (Endoplasmic Reticulum). Partial sequences were acquired from Addgene and are presented in Figure 2.2.

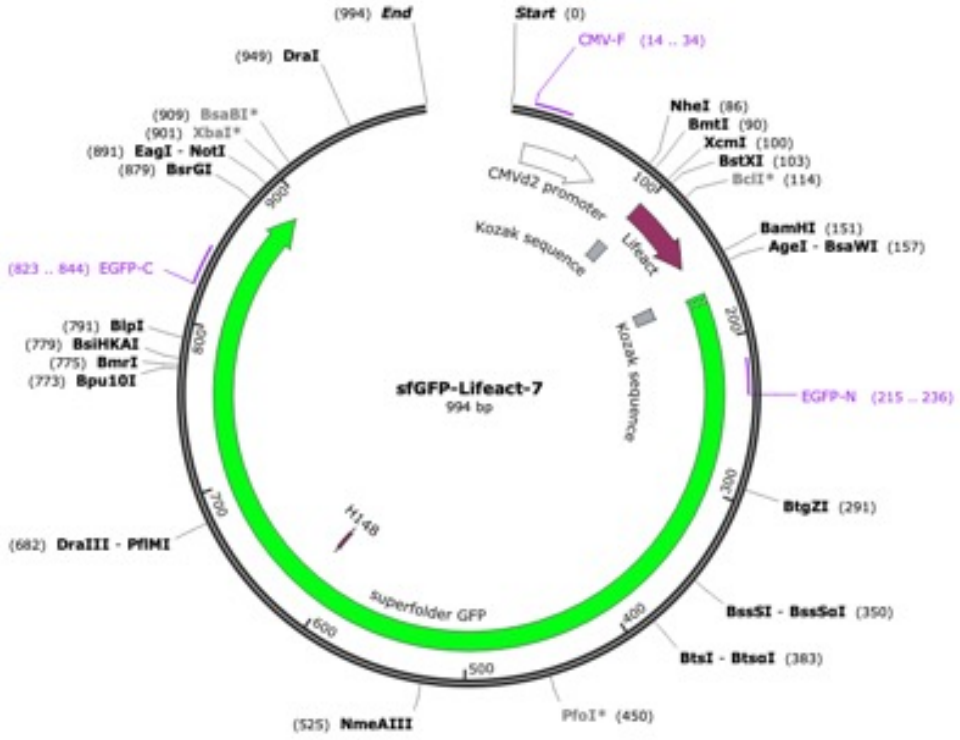

Figure 2.2: Plasmid map of the sfGFP-LifeAct fusion. Plasmid was designed and produced by Twist Biosciences. Plasmid map exported from SnapGeneViewer.

2.7.1 Transfection of mammalian constructs and Cell imaging

For the production of high yields of DNA required for transfection, the QIAGEN Maxiprep Plasmid Purification Kit was used. Overnight cultures of 100 ml *E. coli* TOP10™ cells transformed with a single mammalian construct, were harvested by centrifugation at 5000 rpm for 20 minutes at 4°C using a Thermo Fisher Fiberlite F9-6 x 1000 LEX Fixed Angle

Rotor. The supernatant was discarded, and the bacterial pellet was resuspended in 10 ml of Buffer P1. Subsequently, 10 ml of Buffer P2 was added, and the solution was gently mixed and left to incubate at room temperature for 5 minutes. This was followed by the addition of 10 ml of pre-chilled Buffer P3, mixed thoroughly and left to incubate on ice for 20 minutes. The lysate was centrifuged at 20000 rpm for 30 minutes at 4°C using a ThermoFisher Fiberlite F14-6 x 250y Fixed Angle Rotor. Before applying the supernatant to the QIAGEN-tip column, 10 ml of Buffer QBT were was used to equilibrate the column. Once the supernatant was applied it was left to enter the resin by gravity flow. The column was then washed twice with 30 ml of Buffer QC. The DNA was eluted in 15 ml nuclease-free water and precipitated by adding 10.5 ml of room-temperature isopropanol. The DNA was centrifuged at 20000 rpm for 30 minutes at 4°C. The DNA pellet was washed with 5 ml of 70% ethanol and centrifuged for a further 10 minutes at 15000 rpm, and then carefully decanted the supernatant. Finally, the DNA pellet was air-dried for 10 minutes, and the DNA was redissolved in 500 µl of nuclease-free water. The concentration and quality of the DNA was determined using a NanoDrop ND 1000 Spectrophotometer (ThermoFisher).

The following transfection and cell culture methodologies were performed by Prof Pete Watson. Human cervical carcinoma (HeLa) cells (ATCC, UK) were cultured directly onto live cell dishes (Mattek, USA), and allowed to adhere overnight before transfecting the DNA using Fugene (Promega) according to manufacturer's instruction using 1 µg of DNA per 3 µl Fugene in a volume of 100 µl of OptiMEM (Life Technologies, UK). Cells were then washed, and media was replaced with Fluorobrite DMEM containing 10 % FBS (Life Technologies, UK). Wide-field epi fluorescence measurements were conducted on an inverted Olympus IX73 microscope and a Prior Lumen200Pro light source using filter set (89000, Chroma, Vermont, USA) selecting the ET 490 nm/20 nm as the excitation filter and the ET 525 nm/36 nm as the emission filter. The fluorescence emission was detected with a Hamatsu Orca-flash 4.2 V2 sCMOS Camera operated utilizing the HCImage software package (Hamamatsu). A 100x oil immersion objective with an NA of 1.4 was used to collect sequential images of 1 s exposure over 121 timepoints.

Cells were imaged live at 37°C 5 % CO_2 in a humidifier chamber. For intensity and photobleaching quantification, a 10 µm diameter circular region of interest was measured for average intensity within individual transfected cells utilizing the multi measure function within the ROI manager of FIJI (cite 57). All regions of interest were checked to ensure that all pixels were within the dynamic range of the microscope setup. All images were captured and processed using the same methodology. Over 20 ROIs were assessed per condition. Fluorescence intensity measurements were normalized to their brightest value (at timepoint 0), background was adjusted to fit to the single-phase decay model in GraphPad Prism.

2.8 Imaging techniques

2.8.1 Single molecule analysis

Single molecule imaging and analysis was carried out by Dr David Jamieson. Single molecule imaging was performed using a custom built total internal reflection fluorescence (TIRF) microscope based on a Nikon Ti-U inverted microscope and Andor iXon ultra 897 EMCCD camera. Laser coupling in the microscope was achieved through a custom-built optical circuit (components were sourced from Thorlabs, Chroma and Semrock) followed by a single mode fibre-optic launch. The total internal reflection illumination angle was generated using a combination of fibre-optic micro-positioning and a high numerical aperture TIRF objective (Nikon, CFI Apochromat TIRF 60X oil, NA1.49). Excitation was achieved using a fibre-coupled Venus 473 nm DPSS laser with a power output of 100 mW, achieving ~212.5 mW/cm^2 at the same focal plane (as measured using a photdiode and powermeter (S121C and PM130D, Thorlabs). A 488 nm dichroic mirror coupled with a 500nm long pass filter, and a 525/50 nm bandpass filler were used to isolate the fluorescence signal. Acquisitions were made in areas without prior laser exposure to minimise the effects of photobleaching prior to image capture. An exposure time of 0.06 seconds and an acquisition period of 96 seconds (1600 frames) was used in each experiment. Acquired image stacks were processed using the FIJI distribution of ImageJ to normalise for laser power fluctuation and spatial variations in laser. The first acquisition

frame was removed from all image sequences to account for latency of shutter opening by the camera TTL trigger. All images were processed to normalise for spatial variation in intensity profile of the laser illumination using a reference image look-up of relative spatial illumination intensity, mapping the laser illumination created from the Gaussian blurred (20-pixel radius) median z-projection of a fluorescent image stack. The resulting image stack was then corrected for temporal laser intensity fluctuations to minimise the noise in extracted traces. This was achieved by quantifying fluctuations in the global image background and scaling the corresponding frame accordingly, relative to the mean. Practically this was achieved by removing bright fluorescent spots, defined as any pixel with an intensity greater than 0.05 standard deviations above the median pixel intensity of that frame. Identified pixels were assigned a value equal to the median pixel intensity, effectively erasing them to give a background only image stack. Each frame was scaled relative to the mean intensity of all frames (all pixels) and used to a create a temporal lookup table of relative frame to frame laser power fluctuations. This allowed correction of the main image stack. Background counts were subtracted by the pixel-wise subtraction of time averaged median pixel intensity of a background region of interest. Detection and extraction of single molecule time series was also carried using the ImageJ plugin TrackMate (Tinevez et al. 2017), integrated in the FIJI (Schindelin et al. 2012) distribution of ImageJ. detection was used as means of automatically identifying spots and removing distinct “off” states which due to their abundance can mask peaks within intensity distributions. TrackMate detects spots occurring above a background threshold thus spots which are either photobleached or existing in a dark state for the total duration of any given frame are not included in the detection process. An estimated spot diameter of 4 pixes was applied with a difference of Gaussian (DoG) detection routine. This applies differently sized Gaussian blurs (greater or lesser than the estimated spot diameter) to two copies of each frame which are then subtracted from one another. This process acts as a special bandpass filter enhancing features in the range of the estimated spot diameter enabling detection. Extracted time series were exported to MATLAB where a forward backwards moving window Chung Kennedy filter (Chung and Kennedy 1991) was applied to aid in data

set observations. An algorithmic approach was then developed to identify “on” (bright) and “off” (dark) states in individual data sets by first identifying the photobleaching time and then looking for instances prior to this event where values surpassed a given intensity threshold for a period longer than a given temporal window. An intensity threshold for each time series was determined by first segmenting using a minimum sum log-likelihood deviation from segmented means method (Gnanasambandam et al. 2017). Values from the lowest intensity segments, taken to be photobleached, background state, determined by this method were averaged. An intensity threshold was determined by exploring different multiples of standard deviation from the mean to identify a case where misdetection of events in a given lowest intensity segment was >0.1 % and where mis-detections across all data sets occurred in >2 % of the population. As a result of this process the identified threshold was ten standard deviations from the mean. A second round of temporal thresholding was applied to values meeting the intensity threshold criteria. A window size of 15 frames (0.9 seconds) was used as a lower limit to separate on-state events from background fluctuations in noise. Any intensity values showing sequential temporal separation below this threshold were grouped into single “on” blinking events. Photobleaching lifetime was determined as the end point of the last of these events in a given time series. Cumulative on-time was taken as the sum of all on-events within a given molecule’s photobleaching lifetime. Mean on-time of the two species was determined by fitting a log-normal distribution to a probability density function generated from all on-times from each sfGFP variant using GraphPad Prism. Similarly, the survival half-life was determined by fitting a single component exponential decay to the empirical cumulative distribution function of photobleaching times for each protein. Datasets for the proteins were made up of three experimental repeats each and a total of 3766 and 1283 single molecule spots and corresponding time series were extracted from the YuzuFP and sfGFP datasets respectively.

2.8.3 SRS Imaging

All stimulated Raman spectroscopy (SRS) measurements were performed by Dr David Regan. For the SRS set up, a purpose-built inverted microscope at Cardiff University was used (Figure 2.3). The microscope was first aligned using a polystyrene reference sample to ensure that both the Pump and Stokes beam path are suitable for measurements.

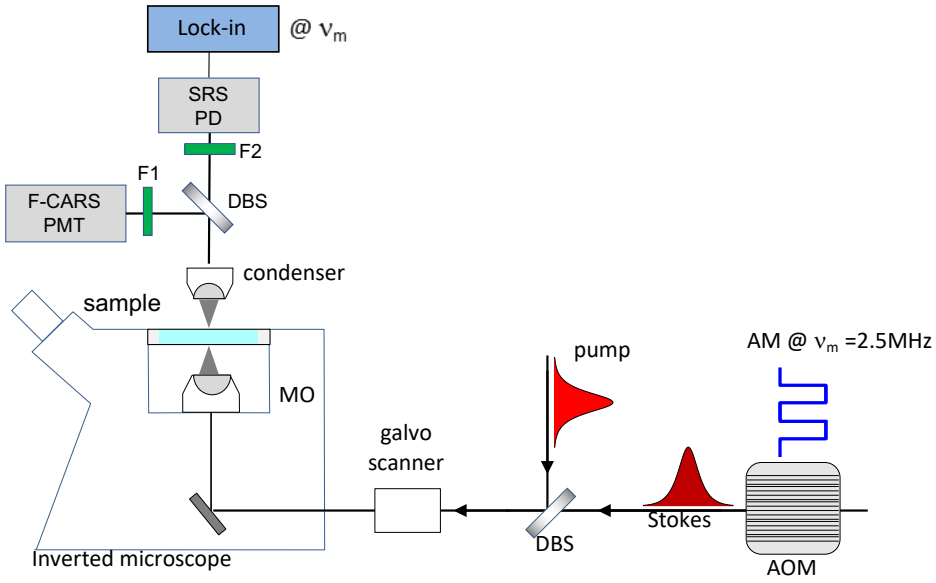

Figure 2.3: Schematic representation of SRS microscope. SRS Photodiode (PD) detects the SRS signal and converts the modulated light to electrical signals. A lock-in amplifier takes these electrical signals at specific modulation frequencies (V_m) and filters out noise. Dichroic Beam Splitter (DBS) reflects specific wavelengths while transmitting other frequencies for the separation of the SRS signal. The inverted microscope contains the microscope objective (MO) to focus light onto the sample. The galvo scanner moves the Pump and Stokes beams across the sample to scan for during imaging. The Acousto-Optic Modulator (AOM) modulate the intensity of the Stokes beam at a frequency of ν_m = 2.5MHz. The amplitude modulation (AM) is critical for the lock-in detection in SRS measurements.

The lock-in amplifier time constant was set to auto, and the phase of the SRS signal was verified using the lock-in reference phase, adjusted to zero degrees (initially 63 degrees). The SRS signal and SRS_DC channels were used for signal acquisition. Prior to the scan, the Pump and Stokes beam shutters were closed, and offsets were applied to all channels. The scan was initiated, and the SRS signal was monitored at dim 0 of the lock-in amplifier. All SRS images were taken using 40 mW Stokes power (100 %) and 13.3 mW Pump power (33 %).

Scan centres were set to 1740 cm^{-1} for the fingerprint region scans and to 2100 cm^{-1} for the biological silent window scans.

3 Incorporation of Raman active bonds in Far – Red Fluorescent proteins

3.1 Introduction

Fluorescence microscopy has revolutionised cell biology in the 20th century, with wide applications such as identification and labelling of cell structures, biochemical reactions and protein-protein interactions to name a few. Although GFP was discovered in the 1960's, it was first sequenced in 1992 and used as a tag in 1994. Although the array of available fluorescent probes constantly increases, there are still limitations due to the nature of the proteins.

Some of the most common problems associated with fluorescence microscopy are photobleaching and autofluorescence. Photobleaching tends to be a problem when studies require monitoring over a long time. As the fluorescent probe is excited over prolonged periods, they eventually undergo chemical damage which changes the chemical nature of the fluorophore and thus loss of fluorescence (Demchenko 2020; Zheng and Blanchard 2013).

Fluorescent protein (FP) chromophores are no exception and are prone to chemical damage by, for example, oxidation due to the presence of molecular oxygen close the chromophore (Duan et al. 2013). This damage does not only affect time-course experiments but also data acquisition in experiments looking to quantify gene expression, monitor dynamic processes and biochemical reactions as fluorescence intensity will not be proportional to the target gene or protein once the reporter FP is damaged. Additionally, it has been previously reported that extended excitation of fluorophores can also lead to the production of reactive oxygen species which in turn can cause oxidative stress within cells leading to cytotoxicity (Icha et al. 2017).

Another limitation of fluorescent microscopy, is the level of multiplexing possible before the signal from the different fluorescent probes is unable to be individually characterised, making it a challenge to have more than 5 separate probes at a time (Gao et al. 2010) (Favreau et al. 2013). This is because fluorescent proteins are excited and emit over a broad wavelength spanning around 50-100 nm meaning that there is already some spectral mixing between the different probes. Overlapping emission and excitation spectra

between the different probes can result in FRËT further disrupting the distinction between individual fluorescent signals. Even though multiplexing to enhance fluorescent signals using FRËT is a widely used method (Holzapfel et al. 2018) it is not always a desired experimental approach when looking at multiple distinct targets.

An emerging way to tackle this, is the use of probes with fluorescence-based microscopy and the development of new optical microscopy techniques that are non-perturbing, are photostable and quantitative. Raman Scattering vibrational microscopy (described in Section 1.2.1) is an emerging analytical tool for imaging as it provides a label-free approach.

In the attempt to generate the chemically unique protein probes for Raman spectroscopy, red fluorescent proteins, mCherry and mNeptune, were engineered for the incorporation of non-natural amino acids. The incorporation of the nnAAs was done to introduce Raman active bonds in the target proteins with the aim to develop genetically encoded Raman scattering tags for live-cell imaging. Incorporating these specific chemical bonds would allow a new method of monitoring different targets using multiple protein probes while removing the limitations posed by fluorescent imaging. The target proteins were chosen due to their spectral properties being appropriate for pre-resonant coherent Raman scattering excitation. Both mCherry (λ_{max} 587 nm) and mNeptune (λ_{max} 600 nm) excite on the far-red range of the UV-visible spectrum and were selected as an alternative to the near infrared fluorescent protein mRhubarb720 (λ_{max} 700 nm) that was engineered by Dr Ozan Aksakal as part of this collaborative project. Unlike mRhubarb720, both mCherry and mNeptune do not required a co-factor to be fluorescent making them an ideal molecular probe. As with most fluorescent proteins, mCherry and mNeptune have a β-barrel structure with a chromophore positioned inside. The chromophore forms through the covalent rearrangement of 3 contiguous amino acids (66-X-Y-G-68) by an autocatalytic process using molecular oxygen (Miyawaki et al. 2012).

In this chapter, the effect of the non-natural amino acid (nnAA) p-ethynyl-phenylalanine (pCCPhe) incorporation within the chromophore of mNeptune as well as the incorporation of the para-cyano-phenylalanine (pCNPhe) in surface residues of mCherry

was explored. *In silico* analysis was initially carried out using PyMol to identify appropriate target residues for mutagenesis and model the incorporation of pCCPhe/pCNPhe. Primers for site-directed mutagenesis were designed for each target residue and once mutants were generated, their spectral properties and Raman spectroscopy potential was analysed. The primers were designed to introduce an amber stop codon (TAG) required for the incorporation of the nnAA and its incorporation took place during protein expression using a reprogrammed genetic code approach. The mCherry:pBAD plasmid was kindly donated by the Jones lab whereas the mNeptune:pBAD plasmid was acquired from Addgene (ID: #54714). Para-cyano-phenylalanine (pCNPhe) and para-ethynyl-phenylalanine (pCCPhe) were selected as the nnAAs of interest due to their Raman spectroscopic properties (Bazewicz et al. 2012). These specific amino acids have been engineered to contain a nitrile (C≡N) or an ethynyl (C≡C) vibrational reporters detected in the Raman spectrum around 2200-2260 cm^{-1} and 2100-2150 cm^{-1} respectively making them excellent bond types since FPs have their strong vibrational resonance in the biologically silent window.

Therefore, this chapter will first focus on the incorporation of the non-natural amino acids at selected residues in mCherry and mNeptune followed by analysis of any changes in their overall spectral properties and finally identify if there is any Raman signal present.

3.2 Results and Discussion

3.2.1 *In silico structural* analysis of mCherry and mNeptune and target residues selection

To enable the incorporation of non-natural amino acids, structural analysis of mCherry and mNeptune was carried *in silico* to identify suitable residues for site-directed mutagenesis. Both proteins share a β-barrel structure typical of fluorescent proteins, with an internal chromophore responsible for fluorescence (see Section 1.1.2.2.1). For the purpose of this thesis, the focus was residues located in close proximity to the chromophore or within the chromophore, as the aim was to create a coupling effect between the vibrational reporter group of the nnAA and the chromophore/side residues, potentially enhancing Raman signals.

For mCherry, target residues were selected based on previous study by Reddington et al. (2015), where they successfully introduced azF within the mCherry core without it affecting the protein's functionality. Corresponding homologous positions were chosen in mNeptune, based on structural alignment and conservation of chromophore architecture. These residues include both those immediately adjacent to the chromophore (W143, E144) as we as other more distant ones (W58, Y120, E144, Q163 and Y181), to test for possible π-π- stacking interactions or indirect coupling effects with the chromophore (Figure 3.1). This is key for generating the Raman probe as the coupling effect will ideally create an enhanced Raman signal.

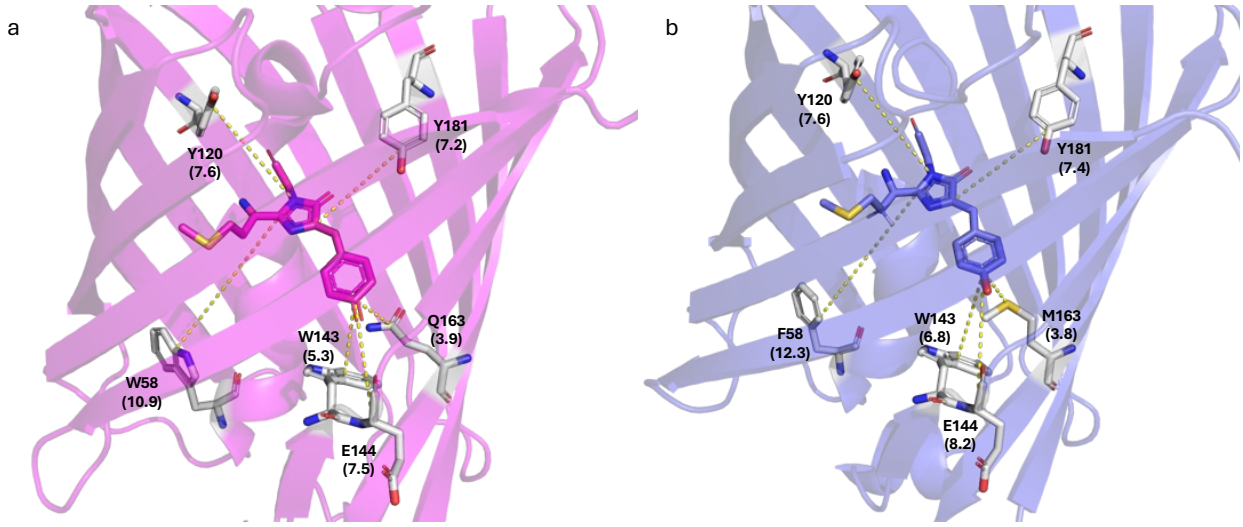

Figure 3.1: Target Residues in mCherry and mNeptune. Cartoon and stick configuration of (a) mCherry and (b) mNeptune with labelled target residues (grey sticks) and the distance (yellow dashed lines) to chromophore (pink and blue sticks respectively) in angstroms (Å). Distance between residues and chromophore were measured using the Distance function in PyMol.

The nnAA was introduced to the target sites via the mutagenesis function on PyMol using the PySwissSideChain package which includes a variety of published non-natural amino acids (Figure 3.2). As mentioned before, our target residue selection was informed from previous studies where the incorporation of the non-natural (azF) in mCherry showed to functionally tolerated within the chromophore area. Although pCNPhe and pCCPhe were selected primarily for their Raman-active chemical bonds (C≡N and C≡C), similar to azF they are ideal for incorporation as they are similar in size to natural amino acids therefore potentially making them easier to be tolerated within the protein. Based on that

information as well as the aim to create a coupling effect between the chromophore and the target residues we came up with a rational for each residue summarised in Table 3.1.

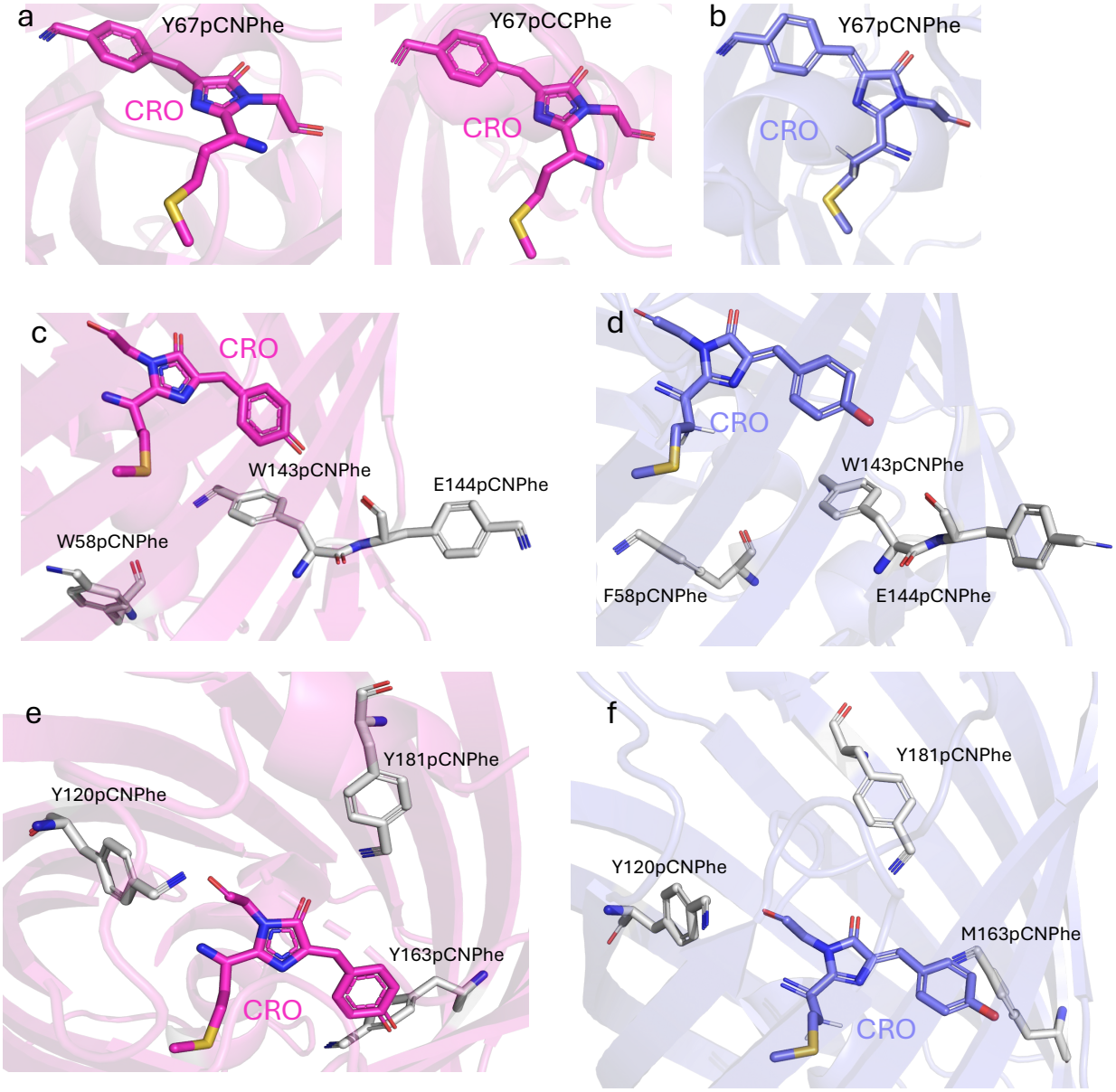

Figure 3.2: Mutagenesis of target residues. a) Incorporation of nnAAs in the chromophore of mCherry (pink) and b) mNeptune (blue). Incorporation of pCNPhe in residues 58, 143 and 144 in c) mCherry and d) mNeptune. Incorporation of pCNPhe in residues 120, 163 and 181 in e) mCherry and f) mNeptune.

Table 3.1: Rational for residue selection for the incorporation of the non-naturals in mCherry and mNeptune

Residue	Rationale
W/F58	Tryptophan and phenylalanine have a similar size to pCNPhe and pCCPhe; to
Y67	Within the chromophore
Y120	Close to the chromophore; has similar size to pCNPhe and pCCPhe
W143	Close to the chromophore; has similar size to pCNPhe and pCCPhe; potential π-stacking between the rings
E144	Side chain faces chromophore Tyr-moiety of the chromophore
Q/M163	Side chain close to Tyr-moiety of the chromophore
Y181	Close to the Tyr-moiety of the chromophore; Close to K70 which directly influences chromophore chemistry

3.2.2 Mutant construction and production

The first step to incorporating the non-naturals was introducing the amber stop codon TAG at the target sites. Site-directed mutagenesis (SDM) via whole vector PCR amplification (Section 2.2.5) was performed with mCherry and mNeptune as the templates using the designed primer pairs (Table 2.6). Following PCR amplification, agarose gel electrophoresis was used to confirm the presence of DNA with a length size of ~5 Kb (Figure 3.3). Both genes are expected to have a similar band size as they have similar gene lengths (4758 bp for mCherry and 4806 bp for mNeptune) and are resident in the same pBAD parental plasmid (see Figure 2.1 for plasmid maps).

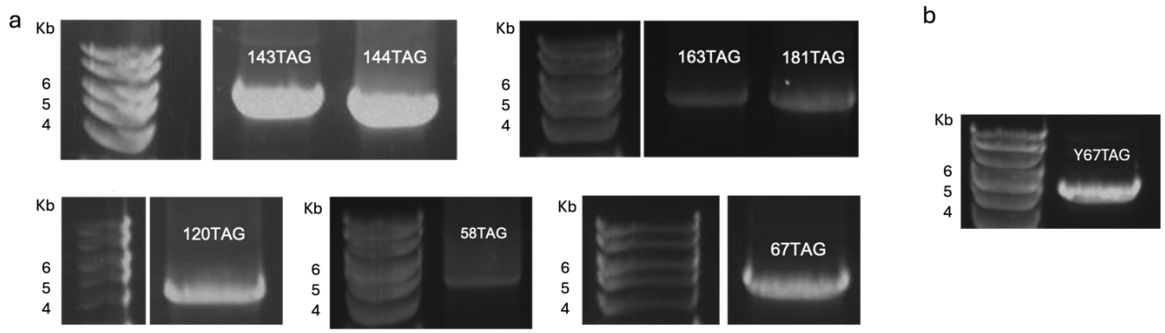

Figure 3.3: DNA analysis after whole plasmid SDM PCR. Gel electrophoresis following SDM via PCR amplification of (a) mCherry and (b) mNeptune TAG codon replacement variants.

The amplified DNA was then purified, phosphorylated and ligated to re-circularise the vector before transforming it into *E.* coli Top10™ electrocompetent cells that were subsequently plated onto LB agar plates supplemented with 1x ampicillin. Colonies were selected and purified plasmid DNA for each mutant was sent for Sanger Sequencing at Eurofins Genomics, to confirm the presence of the TAG codon. Following confirmation of successful introduction of the amber stop codon, plasmid DNA encoding the mCherry and mNeptune mutants was co-transformed with the pDule 2 plasmid (see Figure 2.1 for plasmid map), which facilitates site-specific incorporation of non-natural amino acids such as pCCPhe and pCNPhe. The pDule 2 system function via amber suppression, wherein the UAG codon is reassigned to encode the desired nnAA. As mentioned in Section 1.2.3, this is achieved through an orthogonal tRNA/aminoacyl-tRNA synthetase pair encoded by the pDule 2 plasmid, which specifically recognizes the UAG codon and selectively charges the corresponding tRNA with the supplied nnAA. The orthogonal system operates independently of the host's endogenous translational machinery, enabling efficient and site-specific incorporation of the nnAA at the intended position within protein. Co-expression of the mutant constructs with pDule2 in the presence of the appropriate nnAA resulted in full-length protein expression, indicating successful suppression of the stop codon and incorporation of the non-natural residue.

On successful expression of the pCNPhe or pCCPhe containing variants (with a mature chromophore), a coloured protein product would be expected that should be clearly observed in the cell cultures. For example, mCherry turns cells a purple colour. On

expression of each chromophore variant, the cell cultures had no purple colour suggesting that incorporation of the nnAA affects chromophore maturation. For those mutants where no cell colour was observed, the cells where still lysed and subjected to initial metal-affinity purification to confirm that protein had been produced, despite the lack of colour.

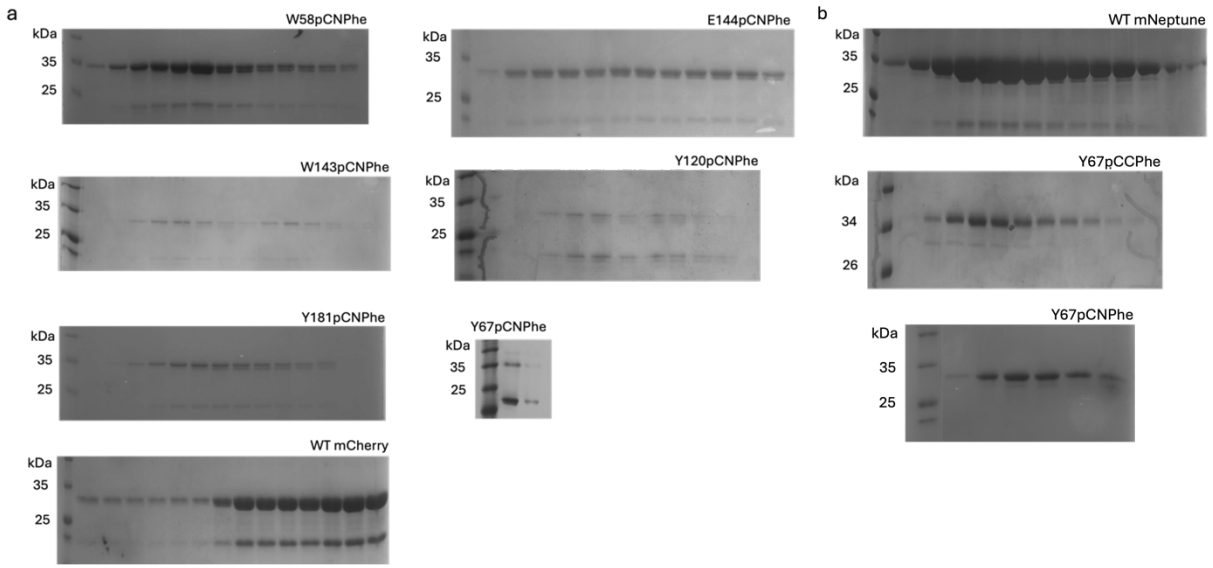

Figure 3.4: SDS-PAGE Analysis of purified proteins. a) Purified wt and mCherry variants with incorporated pCNPhe at target residues following size exclusion. b) Purified wt and mNeptune Y67pCCPhe and Y67pCNPhe after size exclusion. Each lane corresponds to a fraction collected during elution of the purified protein except the first one which corresponds to the molecular weight ladder.

After purification by nickel-affinity followed by size exclusion, SDS-PAGE was performed to confirm the purity of the proteins before downstream analyses. As can be seen in Figure 3.4, each protein is largely purified from waste proteins, eluting only the protein of interest. Although both mCherry and mNeptune are ~28 kDa, both wt and their variants run at around ~34 kDa. They run approximately 6 kDa slower on an SDS-PAGE than the native protein due to the presence of additional His-tags and incomplete denaturation of the β-barrel structure (Baird et al. 1999). The presence of the lower band around 20 kDa corresponds to a cleavage at the chromophore taking place during denaturation, a common occurrence in beta-barrel FPs (Gross et al. 2000). It can also be observed, that upon incorporation of the nnAAs, protein yield and subsequently the concentration of the

protein is reduced compared to that of their wild-type form. Fractions of eluted protein collected measured up to 1 ml of protein each. The first indication that incorporation of the nnAA affects the final amount of protein produced, is the decrease in fractions of pure protein collected which reflects a lower volume of concentrated protein being eluted. In addition to that, the width and intensity of the bands on the gel has decreased illustrating a decrease in concentration per fraction hinting to an overall decrease in protein yield.

3.2.3 Spectral analysis of incorporating non naturals in mCherry and mNeptune

The successfully produced mCherry/mNeptune variants were then analysed to identify any changes in their spectral properties. Spectral analysis of the mCherry variants is shown in Figure 3.4. Overall, λ_{max} blue-shifted slightly for all new variants from 587 nm to 582 nm for mCherry^{W143pCNPhe}, to 586 nm for mCherry^{W58pCNPhe} and mCherry^{E144pCNPhe} and to 585 nm for mCherry^{Y181pCNPhe}. Unlike the other three variants, the absorbance of mCherry^{Y181pCNPhe} shifted to a secondary peak at 500 nm. A reason for this spectral change can potentially be due to disruption within the chromophore environment affecting nearby residues like I197 and K70 which have previously been shown to play a role in chromophore maturation and reduce the green emission (~500 nm) (Baird et al. 2000; Strack et al. 2010). Absorbance was also reduced by ~31-33 % for the 143, 144 and 181 mutants while it increased by ~7 % for the 58 mutant compared to the wild type (72 $mM^{-1}cm^{-1}$) (Table 3.2). Upon excitation at 587 nm emission dropped for mCherry^{W58pCNPhe} by 14 %, for mCherry^{W143pCNPhe} by 77 %, for mCherry^{E144pCNPhe} by 35 % and for mCherry^{Y181pCNPhe} by 36 % and 94 % (when excited 500 nm). Along with the decrease in emission a slight blue shift of the λ_{EM} from 614 nm to 613 nm (W58pCNPhe), to 608 nm (W143/E144pCNPhe) and to 607 nm (Y181pCNPhe) (Table 3.2).

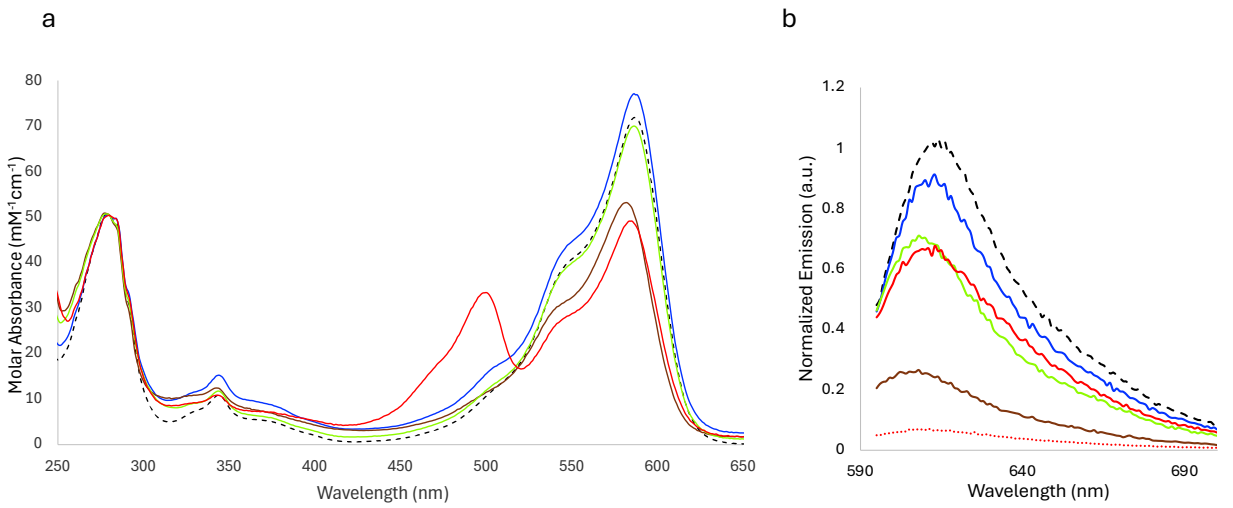

Figure 3.5: Spectral analysis of mCherry variants. Absorbance (a) and emission (b) spectra of mCherry^{WT} (black dashed), mCherry^{W58pCNPhe} (blue), mCherry^{W143pCNPhe} (brown), mCherry^{E144pCNPhe} (green) and mCherry^{Y181pCNPhe} (red). For emission spectra all proteins were excited at 587 nm, while mCherry^{Y181pCNPhe} was also excited at 500 nm (red dotted). Data were normalised to the λ_{max} at 280 nm.

Table 3.2: Spectral properties of mCherry mutants

Protein	λ_{max} (nm)	λ_{EM} (nm)	Absorbance at 587 nm ($mM^{-1}cm^{-1}$)	Emission on excitation at 587 nm
mCherry^{WT}	587	614	72	1
mCherry^{W58pCNPhe}	586	613	77	0.86
mCherry^{Y67pCNPhe}†	367	NR	0	NR
mCherry^{W143pCNphe}	582	608	50	0.23
mCherry^{E144pCNPhe}	586	608	70	0.65
mCherry^{Y181pCNPhe}	585	607	48	0.64
mCherry^{Y181pCNPhe*}	500	611	33**	0.063

*Secondary peak
** Absorbance at 500 nm
*** Emission on excitation at 500 nm
NR – Not recorded
† - Data recorded by Dr Aksakal

For mNeptune the para-cyano-phenylalanine was only incorporated within the chromophore replacing the tyrosine at residue 67. Both pCNPhe and pCCPhe were incorporated in the mNeptune chromophore and both variants showed a reduced absorbance at the expected λ_{max} at 600 nm (Figure 3.6a). The non naturals were only incorporated within the chromophore, due to a few reasons including the time required to successfully incorporate the amber stop codon into the other target residues; but also, upon incorporation the really low yield of protein after expression posed a huge limitation in acquiring enough protein for downstream analysis within the timeline of the project.

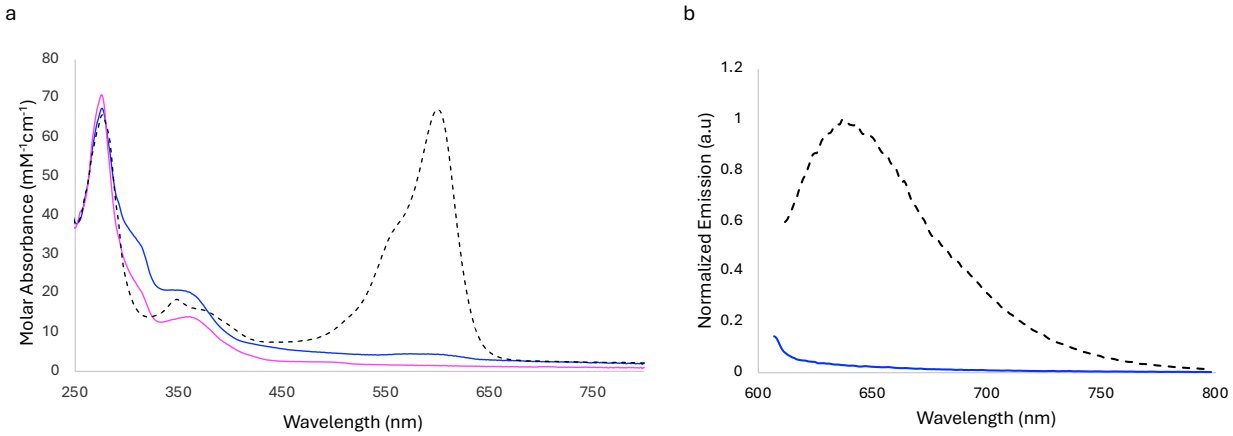

Figure 3.6: Spectral analysis of mNeptune variants. Absorbance (a) and emission (b) spectra of mNeptune^{WT} (black dashed), mNeptune^{Y67pCNPhe} (blue), mNeptune^{Y67pCCPhe} (pink). For emission spectra samples were excited at 600 nm. Data were normalised to the λmax at 280 nm.

As expected, no emission signal was detected when the mutants were excited at 600 nm (no data collected for mNeptune^{Y67pCCPhe}). Although no emission data were collected for mNeptune^{Y67pCCPhe} on excitation at 600 nm, a similar effect to that of mNeptune^{Y67pCNPhe} is expected as the protein shows no absorbance signal at that wavelength (600 nm). SDS-PAGE analysis confirmed the presence of protein; however, the absence of an absorbance spectra was expected, as upon production of the protein there was no observable colour. Therefore, incorporation of the nitrile and ethynyl has showed to disrupt the chromophore's ability to absorb in the red region of the spectrum reflected in the lack of an absorbance spectrum. The inclusion of these chemical moieties may also

interfere with fluorescence due to the chromophores inability to mature. It has previously been shown in many FPs that the hydroxyl group of the tyrosine participates in a hydrogen bond network with S146 known to be important in maintaining the chromophore in its deprotonated state (Yarbrough et al. 2001) and stabilising the intermediates during maturation (Strongin et al. 2007).

3.2.4 Epr-SRS analysis of wild type mNeptune and mNeptune^{Y67pCCPhe}

In this section, the SRS spectra of wild type mNeptune as well as mNeptune^{Y67pCCPhe} was analysed and compared with SRS analysis conducted by Dr Ozan Aksakal on wt mCherry, mCherry^{Y67pCNPhe} and mCherry^{Y67pCCPhe}. For this we primarily looked for any Raman enhancement within the silent (1800-2400 cm^{-1}) and fingerprint region (400-1800 cm^{-1}). The wild type mNeptune was used as a control compared to the mutant proteins with the incorporated nnAAs. This is because upon successful incorporation and functional protein expression, we would expect to see a change in Raman peaks i.e. a lack of a C-O (1180 cm^{-1}) enhancement in the mutants due to the replacement of the hydroxyl group by a nitrile (2250 cm^{-1}) or ethynyl (2125 cm^{-1}) group and vice versa.

It is evident that there is essentially little to no Raman enhancement visible at our regions of interest. There is a low Raman signal exhibited by wild type mNeptune in the fingerprint region (Figure 3.7a) with intensities of ~0.00005 mV at ~1610.5 cm^{-1} and ~0.00008 mV at 1686 cm^{-1}. The region of peaks marked with an asterisk correspond to the C=C, C=N and C=O stretchings (1500-1900, 1610-1680 cm^{-1} and 1680-1800 cm^{-1}). Unlike the mutant, we do get slight signal enhancement in that peak region (*), where those bonds bonds are attributed to. Even though it is a minimal Raman enhancement, it seems to be completely gone once the para-ethynyl-phenylalanine is incorporated within the mNeptune chromophore. On the other hand, there seems to be no Raman enhancement in the silent region in either of the two proteins. This was expected for wild type as it does not have the ethynyl bond that would have a Raman signal (2100-2150 cm^{-1}) within the silent region; but not for the mNeptune^{Y67pCCPhe} as pCCPhe contains the vibrational modes that should be detected within the biological region. Lack of any Raman signal enhancement

can potentially be attributed to the low concentration of the protein sample, due to distortion of the conjugated bond system (no C=C/C=N peak present in mNeptune^{pCNPhe} as well the fact that the system the measurements were recorded at was not optimal for mNeptune. This is because the in-house system we possess at Cardiff University has a pump wavelength of 820 nm. In order for optimal pre-resonance enhancement, the frequency of the vibrational modes needs to be close to the absorption maxima of the protein of interest. Since mNeptune is more blue shifted (λ_{max} = 600 nm) than the pump pulse of the current system, minimal pre-resonant enhancement is what could be achieved.

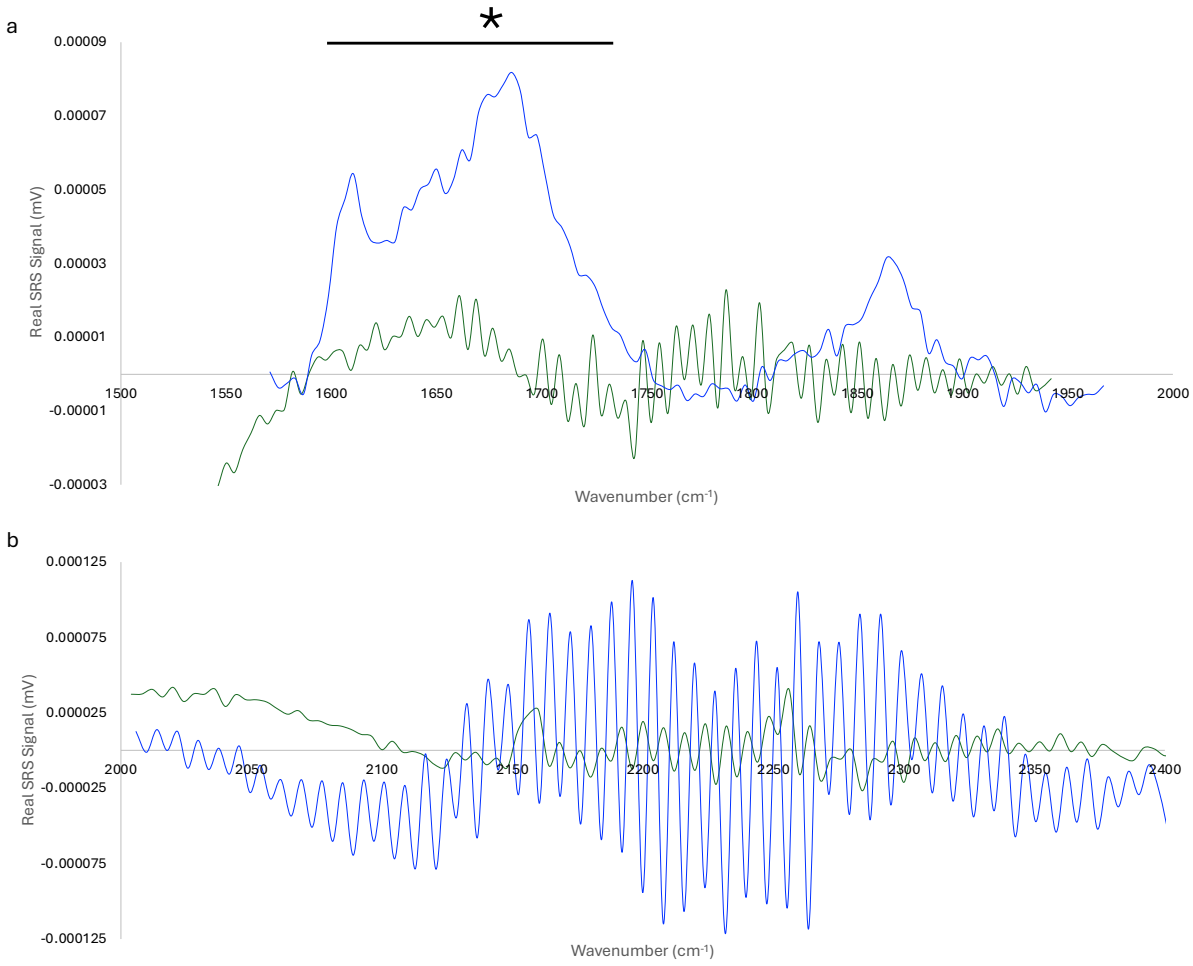

Figure 3.7: SRS analysis of mNeptune^{WT} and mNeptune^{Y67pCCPhe}. Real SRS measurements (after background subtraction) of fingerprint region (a) and the silent region (b) of mNeptune^{WT} (blue) and mNeptune^{Y67pCCPhe} (green). Measurements were taken at 1 mM sample concentration for the wild-type sample and 300 μM for the non-natural variant. Both samples were in 50 mM Tris buffer. Data were recorded by Dr David Regan.

3.3 Conclusion

In conclusion, the generation of proteins with incorporated non-natural amino acids has proved to be a success and a challenge at the same time. Despite successfully generating a number of the target variants originally planned, several limitations arose in the process. Upon the incorporation of the non-natural amino acids the overall protein yield after expression decreased for the majority of the variants. Although enough protein was made to allow for spectral analysis of the different variants, there was not sufficient

protein to allow for further analysis of the properties of the protein or downstream experiments including SRS measurements.

Through the experimental analysis of the SRS data of mNeptune^{Y67pCCPhe}, we could see no significant enhancement in the fingerprint region as we expected since the C≡C present in the ethynyl group of pCCPhe contains vibrational modes that should be detected within the biologically silent region. On the other hand, even, the slight enhancement observed initially in the fingerprint region of the wild-type mNeptune was lost once pCCPhe was introduced. However, solid conclusions cannot be made based on the current data available. First of all, high concentrations of 1 mM could not be achieved for mNeptune^{Y67pCCPhe} where the sample used for SRS measurements was about 300 μM unlike that of the wild-type equivalent. In addition to that, pre-resonance enhancement could not be achieved with the microscope set up at Cardiff University. In order to achieve pre-resonance enhancement for SRS, the Pump wavelength needs to be closer to the λmax of the protein of interest. Since λ_{max} for mNeptune is 600 nm, low to no enhancement was expected as the Pump wavelength is set at 820 nm.

Moving forward, optimising protein production to aid in the final yield but also fine tuning the microscope set-up are key improvements that can potentially allow to further explore and assess the potential of red-fluorescent proteins as Raman spectroscopy probes.

4 Functional and structural characterisations of mCherry^{M66C}

Zitti, A., Aksakal, O., Vitsupakorn, D., Rizkallah P. J., Mikolajek, H., Platts, J., Menzies, G. E., & Jones, D. D*. (2025). Structure, Function and Dynamics of mCoral, a pH-Responsive Engineered Variant of the mCherry Fluorescent Protein with Improved Hydrogen Tolerance. International Journal of Molecular Sciences (IJMS), 27(1), 154. doi: 10.3390/ijms27010154

*Corresponding author

Statement of work

Work in this chapter was performed with the aid of Cardiff School of Biosciences Protein Technology Hub for production and analysis of proteins, Advanced Research Computing Cardiff (ARCCA) facility for supporting the molecular dynamics work, and the VMXi facility and its staff at Diamond Light Source. I contributed to the conception of the project, generated the mCoral (mCherry M66C) mutant, produced protein, undertook the spectral analysis and contributed to structure determination. Dr Ozan Aksakal contributed towards the molecular dynamics analysis and some aspects of mCherry characterisation. Danoo Vitsupakorn contributed to structure determination. Dr Pierre Rizkallah contributed to structure determination. Dr Halina Mikolajek contributed to protein crystallisation and X-ray diffraction measurements. Dr James Platts undertook DFT analysis. Dr Georgina Menzies contributed towards molecular dynamics data generation and analysis helped direct the project. Prof Dafydd Jones contributed to project conception and directed the project, contributed to general data analysis and molecular dynamics analysis.

4.1 Introduction

In the previous chapter, the focus was on engineering red-fluorescent proteins for the incorporation of nnAA for subsequent enhancement of Raman spectroscopy. In this chapter an alternative approach was explored, where the focus was on the incorporation a cysteine within the mCherry chromophore. Cysteine was chosen as the alternative, since it can be detected within the biologically silent region (1800-2850 cm^{-1}), due to the thiol (SH) group that gives a signal around 2560-2590 cm^{-1} (Jenkins et al. 2005) and because of that, it could be used as an alternative probe to that with incorporated non-natural amino acids.

Changing the composition of the chromophore is one of the ways to alter the properties of a fluorescent protein as it is positioned inside the β-barrel; its interaction with the rest of the protein in turn tunes the protein's spectral characteristics. As mentioned before in Section 1.1.2 in mCherry, the three chromophore forming residues are 66-MYG-68, with the conjugated double bond system expanding into the peptide bond with F65 (so giving it the red region spectral properties).

In chapter 3, the focus was incorporating the non-naturals at position 67, where the tyrosine is widely conserved among FPs but can be changed to other aromatic residues (including non-natural amino acids) (Heim et al. 1994; Goulding et al. 2008) . In this chapter the target was the M66 in mCherry. Originally a glutamine, it was mutated to a methionine (mRFP1 to mRFP1.1) which improved chromophore maturation and slightly red-shifted the excitation and emission spectra relative to mRFP1 (Shaner et al. 2004). A cysteine has previously been incorporated at residue 66 along with additional mutations that gave rise to mTangerine and mBanana. In this project cysteine was introduced into the chromophore via SDM and purified mCherry^{M66C} was characterised through a series of experiments, including spectral analyses, pH and oxidation potential studies, high-performance liquid chromatography - mass spectrometry (LCMS), protein crystallisation and structure determination via x-ray crystallography. Molecular dynamic simulations were also conducted to provide a dynamic overview of what is happening in the chromophore environment upon the incorporation of cysteine.

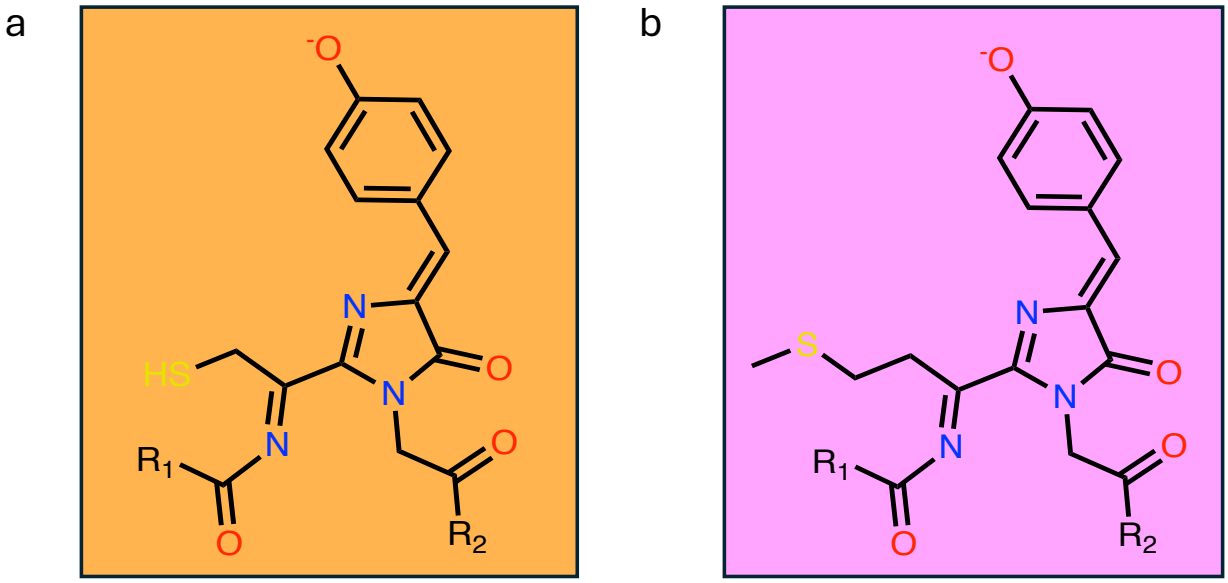

Figure 4.1: Chromophore structure of mCherry^{WT} vs mCherry^{M66C}. a) mCherry^{M66C} chromophore with a cysteine incorporated. b) mCherry chromophore with a methionine.

4.2 Results and Discussion

In silico structural analysis informed the decision to incorporate a cysteine within mCherry as it was identified that there was no other cysteine residue present within the amino acid sequence of mCherry. Since the initial purpose was to potentially enhance the Raman signal of cysteine within the chromophore, the lack of another cysteine residue eliminated the potential formation of disulphide bonds between the chromophore and the rest of the protein which would potentially affect the coupling effect as well as the structure and function of the protein.

Moreover, despite the fact that mTangerine and our engineered variant of mCherry have similar spectral characteristics in respect to their emission/excitation wavelengths (blue-shifted with respect to mRFP1.1 and mCherry respectively), there is a lack of structural and functional information on mTangerine, aside from its extinction coefficient value (38000 $M^{-1}cm^{-1}$) and a quantum yield of 0.3. In addition to that, residue 66 has previously been the target for mutagenesis with the aim to generate proteins with improved photostability and maturation. Along with cysteine, serine and threonine were previously used to replace glutamine in mRFP1, which gave rise to proteins with faster maturation

compared to DsRED but not mRFP1. Out of the three mRFP1 mutants, the one with threonine showed to have improved brightness as well as higher photostability compared to the other two mutants (Jach et al. 2006). It is crucial to understand that residue 66 clearly plays a vital role in determining the spectral properties, maturation and photostability on red fluorescent proteins which in turn made it an ideal target for this study.

4.2.1 Incorporation of Cysteine in residue 66

To replace methionine with cysteine, whole plasmid, inverse PCR side-directed mutagenesis was employed using primers designed with the GCA codon on the reverse primer at the target side (Table 2.6) (Section 2.2.5). Agarose gel electrophoresis was carried out to test the quality and presence of the desired vector DNA before purifying it for transformation into *E.* coli TOP10™ cloning cells. The presence of the full-length gene encoding mCherry^{M66C} produced a band of ~5 Kb, as observed in Figure 4.1a. Plasmid DNA was then extracted and purified and was subsequently sent for Sanger sequencing to confirm the presence of the M66C mutation. Once the mutation was confirmed the protein was successfully produced in significant quantities by expressing in *E.* coli.

Expression of mCherry^{M66C} was confirmed by the presence of pink coloured cultures. Coloured cultures are a primary indication of chromophore maturation and thus a functional fluorescent protein. Fractions with absorbance in the red region that exhibited a purple/pink colour were collected during elution and analysed via SDS-PAGE to confirm their purity. Bands of the expected size of the mCherry monomer were observed in both gels (Figure 4.1c, d) confirming the presence of the protein and purity of the samples. As mentioned in chapter 3, mCherry is ~27 kDa in size, but it runs around ~34 kDa due to the presence of additional His-tags and incomplete denaturation of the β-barrel structure thus slowing down its migration (Baird et al. 1999). The ~20 kDa band corresponds to covalent cleavage at the chromophore on denaturation (Gross et al. 2000).

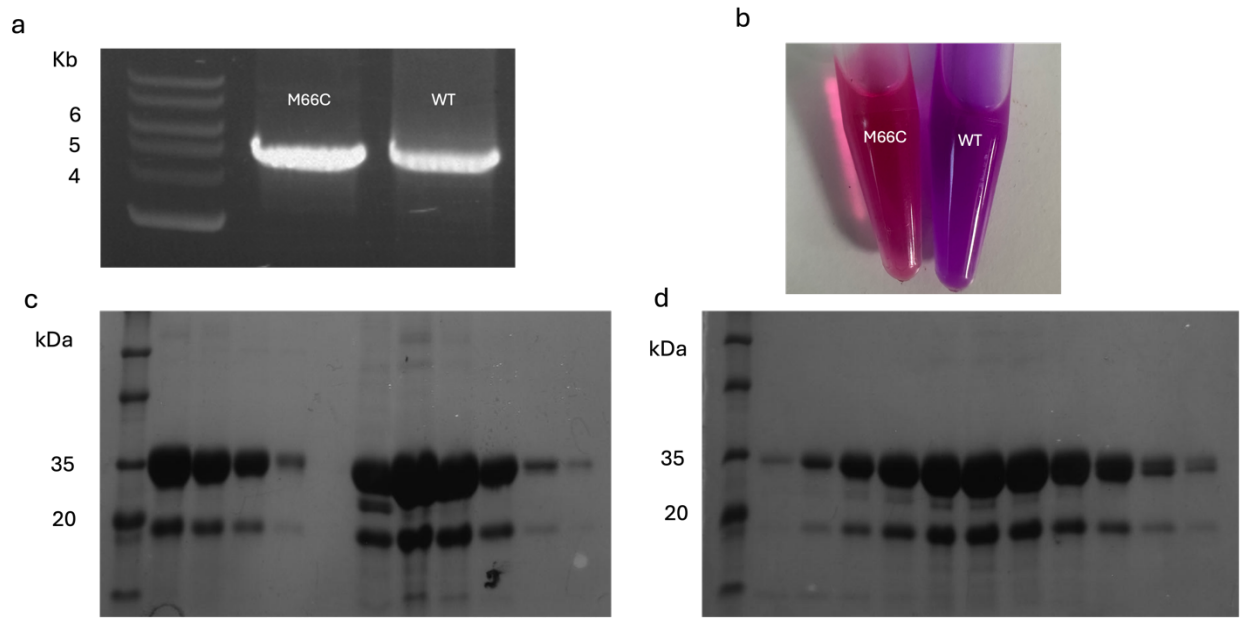

Figure 4.2: Generation and production of mCherry^{M66C}. a) Site directed mutagenesis of mCherry^{M66C} vs wild type mCherry. b) Purified mCherry^{M66C} compared to wild type mCherry. c) Fractions of protein (with pink colour) from Nickel affinity chromatography of mCherry^{M66C} and d) protein fractions (with pink colour) from SEC of mCherry^{M66C}.

4.2.2 Spectral characterisation of mCherry^{M66C}

Absorbance and fluorescence of mCherry^{M66C} were recorded to determine the spectral characteristics of the protein. The mCherry^{M66C} exhibits an absorption peak at 566 nm, with a molar absorbance of 49.9 $mM^{-1}cm^{-1}$. Upon excitation at 566 nm, mCherry^{M66C} shows a λ_{EM} of 585 nm. When compared to mCherry, a 21 nm blue shift in absorbance is observed, with a corresponding blue shift of 25 nm in the wavelength of fluorescence emission as well. This shift in spectral properties is also reflected in the colour of the protein as we see a change from a cherry purple to a strawberry pink (Figure 4.1b).

Using mCherry as the reference (quantum yield 22%), the quantum yield of mCherry^{M66C} is estimated to be 27%, thus higher than mCherry. Higher quantum yields are down to chromophore rigidity, stability and how changes in the pH and electrostatic environment affect it (Campbell et al. 2002; Shaner et al. 2005; El Khatib et al. 2016). Thus, despite the lower molar absorbance coefficient observed for mCherry^{M66}, the overall brightness is similar to mCherry (Table 4.1).

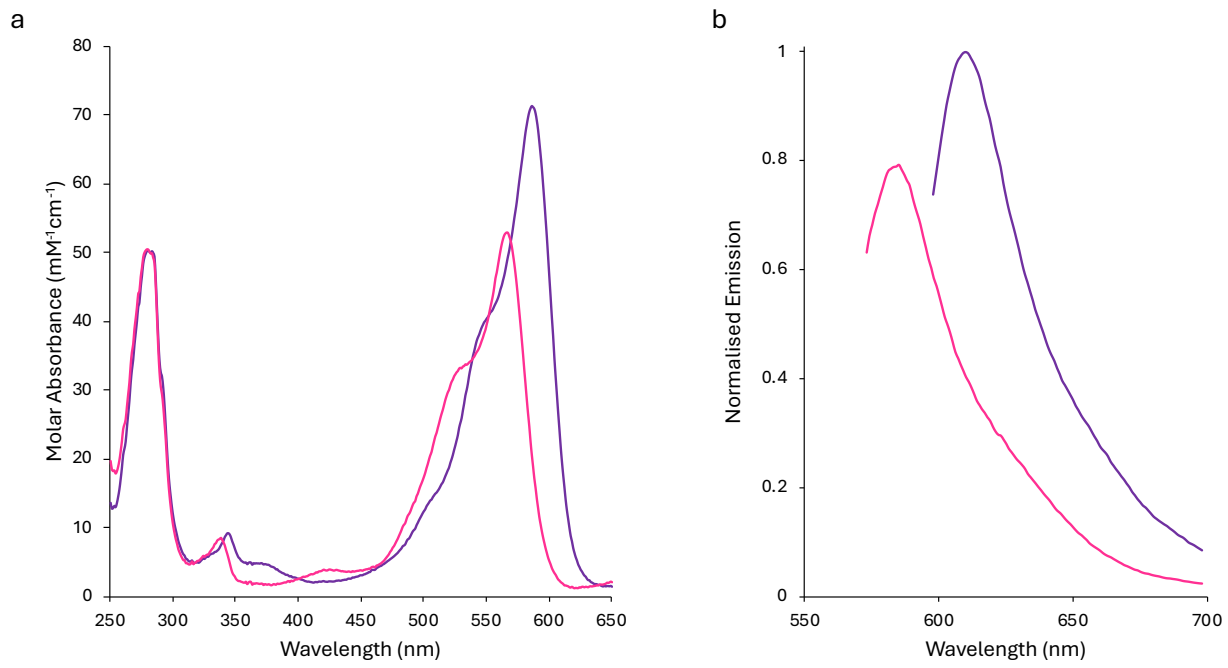

Figure 4.3: Spectral analysis of wild type vs mCherry^{M66C}. a) Absorbance spectra of mCherry (purple) against mCherry^{M66C} (pink) b) Emission spectra of mCherry (purple) against mCherry^{M66C} (pink) upon excitation at the λ_{max} of 587 nm and 564 nm respectively. Emission spectra is normalised to peak emission for mCherry.

Table 4.1: Spectral properties of mCherry and mCherry^{M66C}

Protein	λ_{max} (nm)	λ_{EM} (nm)	ε ($mM^{-1}cm^{-1}$)	QY	Brightness ($mM^{-1}cm^{-1}$)
mCherry	587	610	72	0.22	15.8
mCherry^{M66C}	566	585	49.9	0.27	13.5
mRFP1	584	607	50	0.25	12.5
mRFP1-Q66C	559	580	31.8	0.33	10.49
mTangerine	568	585	38	0.3	11.4

When it comes to the incorporation of cysteine at residue 66, aside from our new generated variant, others have been engineered before such as mRFP1-Q66C and mTangerine. When comparing the three proteins to mRFP1, mRFP1-Q66C doesn't carry any other mutations aside from Q66C (Jach et al. 2006), mTangerine has two additional mutations T147L and Q213L, while our mCherry^{M66C} variant has the cysteine incorporated at 66 in concert with the mutations that make up the mCherry background. Upon

comparing the spectrochemical properties of the three proteins, it is evident that mCherry^{M66C} has similar properties to that of mTangerine with a 2 nm difference between their excitation and emission peaks, 566/587 nm and 568/585 nm respectively (Table 4.1). Compared to mTangerine, mCherry^{M66C} is slightly brighter (13.5 $mM^{-1}cm^{-1}$ compared to 11.4 $mM^{-1}cm^{-1}$). To my knowledge, there has been no assessment of the Raman properties of these C66 mutants nor detailed analysis of their pH and ROS sensitivity.

Using this information, a series of functional experiments were conducted to further understand the protein's oxidation potential, ability for small molecule binding and the effect of changes in pH reflected by changes in excitation/emission. Molecular dynamics simulations conducted by Prof Dafydd Jones and Dr Georgina Menzies, provided further information on the transport of water around the chromophore and through the β-barrel of the mCherries of interest. LCMS was the last method employed and was carried out in the presence and absence of hydrogen peroxide to further aid in the understanding how oxidation affects the chromophore, its environment and in extend its spectral properties.

4.2.3 Epr-SRS analysis of mCherry and mCherry^{M66C}

As mentioned previously in Chapter 3, our current in-house Raman microscope setup at Cardiff University, is not ideal for epr-SRS measurements of mCherry as it is optimised for IR absorbing probes (pump power = 820 nm). Because of that, mCherry and mCherry^{M66C} were sent to the University of Konstanz in Germany, where a more blue-shifted laser set-up is available. Thanks to Dr Andreas Zumbusch and Dr Andrea Pruccoli, the Raman profiles were measured for each protein.

The following epr-SRS measurements were conducted using a pump excitation of 640 nm, which is much closer to the excitation maximum of mCherry (587 nm). The fingerprint region was initially examined to observe if enough Raman signal could be recorded for mCherry to compare with that of mCherry^{M66C} (Figure 4.4). Both proteins have a clear enhancement throughout this spectral range with various peaks from 1000-1700 cm^{-1}. The Raman peaks characterised in Figure 4.4 were sourced from 'The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules' (Griffiths 1992). The

following peaks in Figure 4.4 are attributed to Raman enhancement of the bonds within the chromophores of each mCherry.

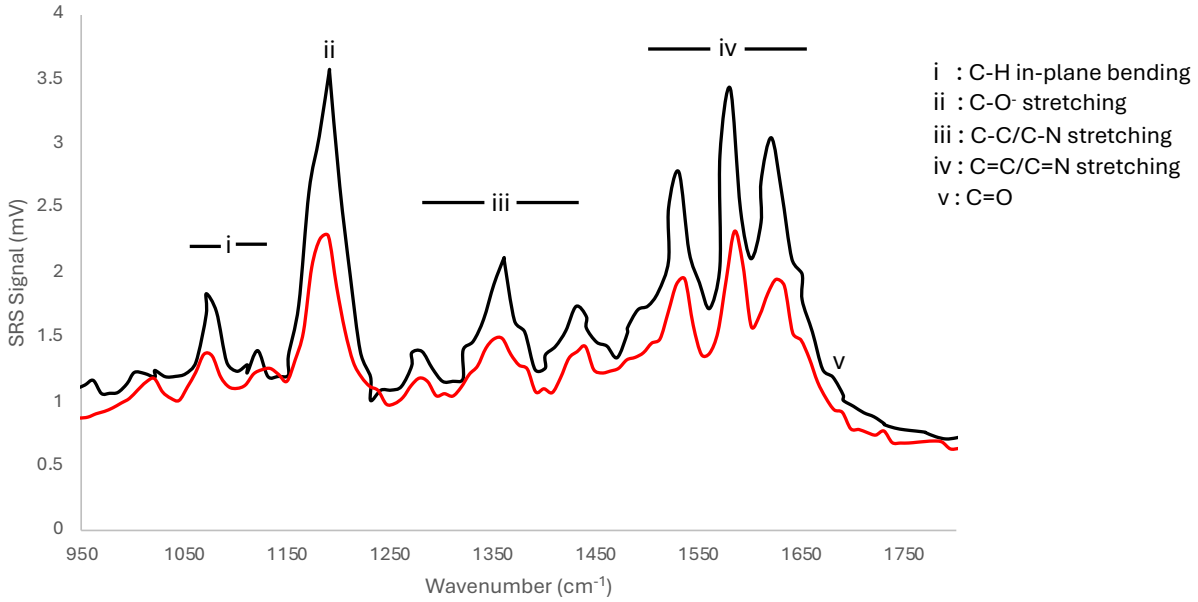

Figure 4.4: SRS analysis of mCherry and mCherry^{M66C}. Real SRS measurements (after background subtraction) of fingerprint region of mCherry (black) and mCherry^{M66C} (red). Measurements were taken at 1mM sample concentration. Both samples were in 50mM Tris buffer.

Each peak corresponds to a particular bond, with the first peak (i) being attributed to the C-H in plane bending most likely from the aromatic phenolic ring of the chromophore that contains four C-H bonds as well as the conjugated bond connecting the aromatic rings as they also contain a C-H bond in-plane. The second major peak (ii) at 1186 cm^{-1} is associated with the C-O^{-} stretching of the phenolic ring in the chromophore. The following range between 1250-1450 cm^{-1} (iii), is associated with the C-C and C-N stretching. The particular vibration originates from the C-C and C-N bonds found within the conjugated environment of the chromophore and strongly contribute due to the polar nature of the Nitrogens. The peaks around 1500-1660 cm^{-1} (iv) correspond to the C=C, predominantly the ones within the aromatic ring of the tyrosine within the chromophore and C=N stretching (present in the amide group of the imidazolinone ring of the chromophore). This vibrational mode is characteristic of conjugated double bonds in aromatic systems and is further

enhanced by the resonance within the ring structure, leading to a strong Raman signal. The last peak (v) at 1686 cm^{-1} corresponds to the C=O stretching and is associated with the carbonyl groups of the chromophore.

It is evident from the peaks on Figure 4.4 that mCherry has a higher Raman enhancement (~1.5 to 1.8- fold higher) than $mCherry^{M66C}$. The difference between the Raman signal of the two proteins is summarised in Table 4.2 This decrease in enhancement can be attributed to the fact that the pump excitation (640 nm) used in the German setup which is more appropriate for mCherry might be too red-shifted to get a high pre-resonant enhancement for $mCherry^{M66C}$ as its λ_{max} is blue-shifted (566 nm) compared to the λ_{max} of mCherry (587 nm). This posed another limitation and thus another reason as to why the focus for this specific mCherry variant shifted to a different direction than pursuing the Raman probe approach.

Table 4.2: SRS Signal of mCherry and $mCherry^{M66C}$ in the fingerprint region.

Wavenumber (cm^{-1})	SRS Signal (mV) of mCherry	SRS Signal (mV) of $mCherry^{M66C}$
1073	1.85	1.21
1186	3.59	1.90
1282	1.40	1.06
1355	2.13	1.38
1436	1.67	1.26
1532	2.79	1.60
1581	3.44	2.07
1621	3.06	1.91
1686	1.06	0.80

4.2.4 Response of mCherry^{M66C} to changes in environment

The replacement of methionine with cysteine changes the inherent chemistry of the chromophore. While sulphur moiety is retained, the chemistry is different. Both methionine and cysteine are sensitive to reactive oxygen species (ROS) with the sulphur atoms capable of being oxidised to sulfoxides . This in turn will increase the local bulk due to addition of 1-2 oxygen atoms (Figure 4.5). The thiol group of cysteine is also pH sensitive (pKa = 8-8.5). Both these aspects flux in biological systems and play an important role in various biological processes. Here, it was investigated how ROS in the form of H_2O_2 and pH change the spectral properties of mCherry^{M66C}.

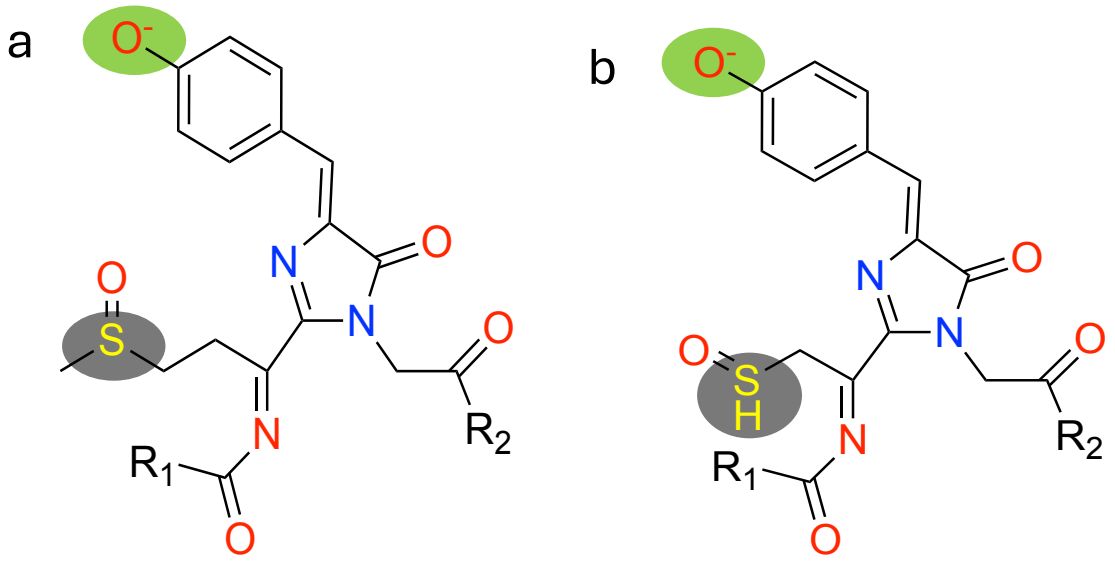

Figure 4.5: Oxidation of sulphur atoms in mCherry and mCherry^{M66C} chromophore. a) mCherry and b) mCherry^{M66C} chromophore with oxidised sulphur atoms.

4.2.4.1 The effect of pH on spectral properties.

Cells and their environment experience changes in pH. While pH values of around 7 are sampled by many cellular compartments, organelles such as lysosome endosomes and Golgi apparatus experience acidic pH values while the mitochondria matrix has a pH closer to 8 (Porcelli et al. 2005). Thus, the pH sensitivity and potentially sensing capability is an important feature of FPs.

Wild type mCherry and mCherry^{M66C} were subjected to different pH states and their absorbance and fluorescence spectra analysed. Fluorescence and absorbance spectra

were recorded for both proteins at each pH (pH 4.5-11) at a concentration of 5 μM for both proteins with pH 8 being the control pH state (Figure 4.6).

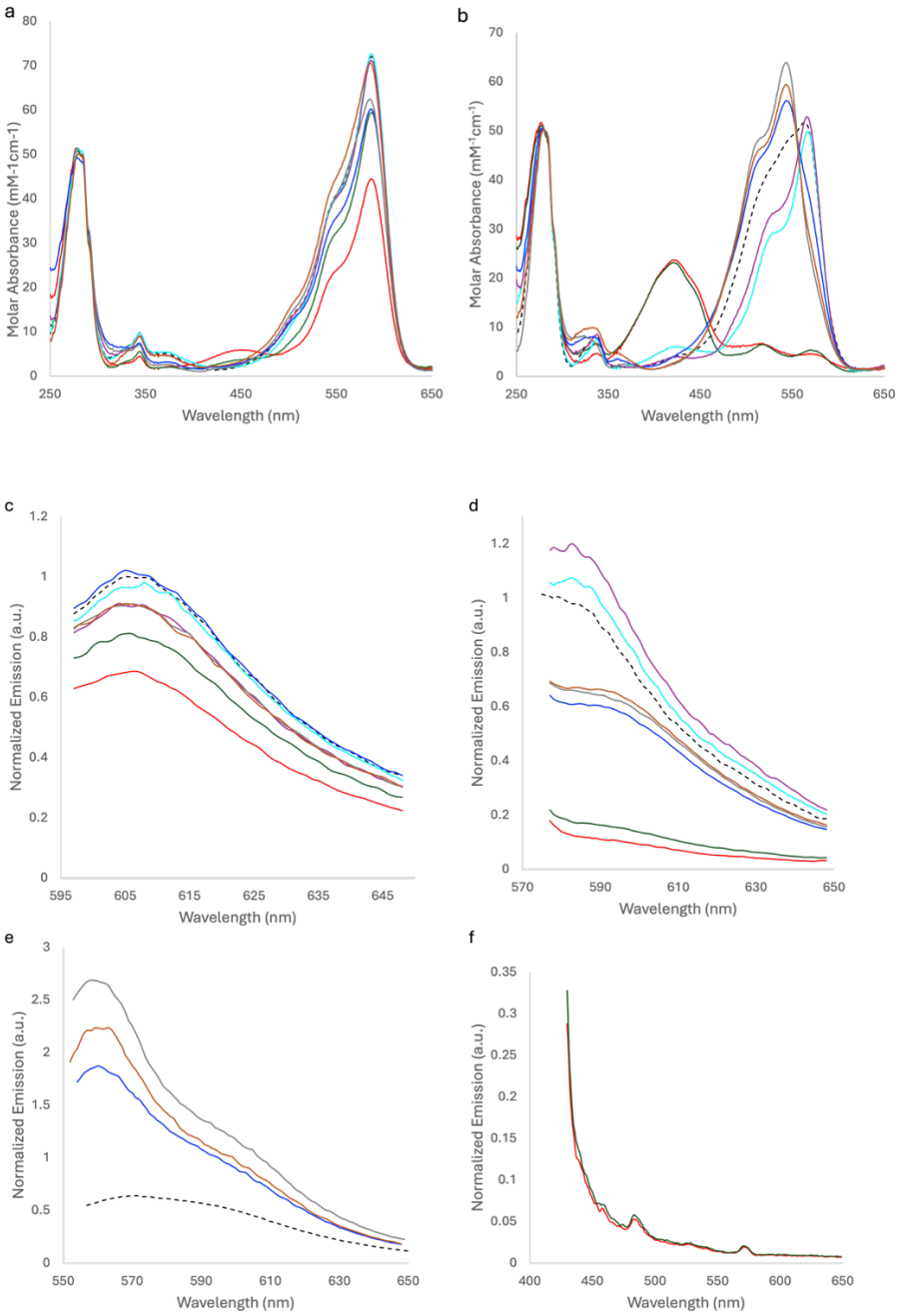

Figure 4.6: Effect of pH on spectral properties of wild type mCherry and mCherry^{M66C}. Absorbance spectra of a) wild type mCherry and b) mCherry^{M66C} at different pH states including, pH4.5 (red), pH5 (green), pH6 (turquoise), pH7 (purple), pH8 (black dashed), pH9 (blue), pH10 (grey) and pH11 (brown). c) Emission spectra of mCherry at the different pH states on excitation at 587nm. d) Emission spectra of mCherry^{M66C} at different pH states on excitation at 564 nm. e) Emission spectra of mCherry^{M66C} at pH9-11 compared to pH8 on excitation at 544 nm and f) emission spectra of mCherry^{M66C} at pH 4.5 and pH5 on excitation at 520 nm. Absorbance spectra were normalised to Molar Absorbance at 280 nm and Emission data to peak emission of control protein (pH8) for each variant.

The two proteins behave differently in terms of their response to different pHs. For mCherry, the absorbance spectra retain same general profile but below pH 5 the molar absorbance at 587 nm (λ_{max}) decreases 1.6-fold to ~44 $mM^{-1}cm^{-1}$. The same is true for fluorescence emission on excitation at 587 nm (the λ_{max}), with the major drop in fluorescence occurring at pH 4.5 (1.3-fold decrease). When molar absorbance values at each pH are fit to a sigmoidal curve , the mid-point is pH 4.4, which is equivalent to the pKa value (Figure 4.7). This is close to the reported value for mCherry (pKa 4.5 (Shaner et al. 2004)).

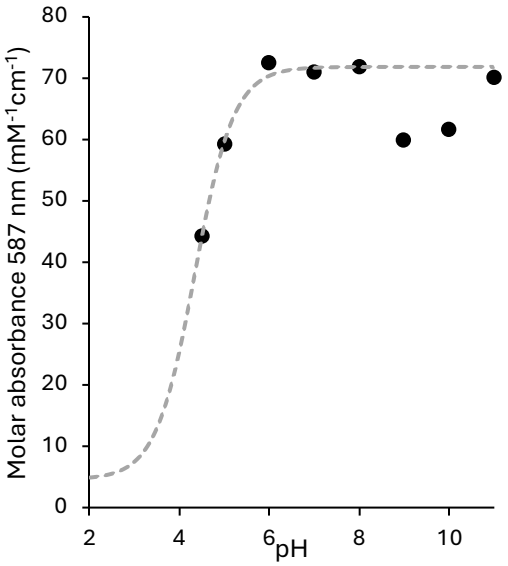

Figure 4.7: mCherry pKa. Sigmoidal curve of mCherry's Molar absorbance against increasing pH showing a pKa at 4.4.

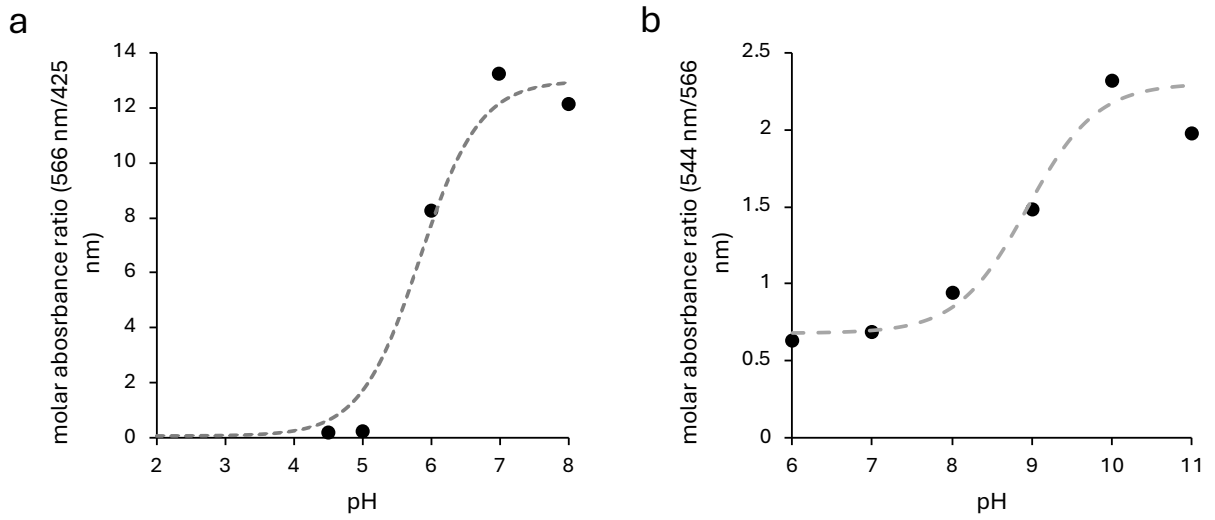

Figure 4.8: pH transition states of mCherry^{M66C}. a)Transition between acidic and neutral pH states with a pKa of 5.7. b) Transition between neutral and alkaline pH states with a pKa of 8.8.

On the other hand, spectral properties of mCherry^{M66C} show a very distinct pH profile with two clear transitions (Figure 4.6b). At pH 7, the absorbance spectra of mCherry^{M66C} shows a distinctive peak at 564 nm. On shifting to pH 8, the absorbance peak becomes broader with the blue shifted shoulder region become more prominent. Above pH 8, the major absorbance peak is now 544 nm, a blue shift of 20 nm compared to pH 7. The molar absorbance at the λ_{max} also increases 1-7 fold from pH 7 (36 $mM^{-1}cm^{-1}$) to pH 10 (64 $mM^{-1}cm^{-1}$). A sigmoidal plot of the relative absorbance values at the two major peaks (544 nm and 566 nm) shows that there is a transition pH 7 and 10 that equates with the mid-point being at pH 8.8, which is the pKa for this transition (Figure 4.8b).

In comparison, at low pH there is a large blue shift in the λ_{max} to 422 nm with a drop in the molar absorbance coefficient (Figure 4.6b). The abrupt change happens between pH 5 and 6. Excitation at 422 nm does not produce any fluorescence, with excitation at 566 nm having a 82-87% decrease in emission compared to that at pH 7. A sigmodal plot of the relative absorbance values at the two major peaks (425 nm and 566 nm) shows that there is a transition pH 5 and 7 that equates with the mid-point being at pH 5.7, which is the pKa for this transition (Figure 4.8a).

Between pH 6-8 peak absorbance remains at ~564 nm with a broad peak, while at pH 6 and 7 the peak becomes sharper. In alkaline conditions absorbance shows a ~20 nm blue shift and an increase in intensity at that specific λ_{max} (~544 nm) compared to that of the control at pH 8. At low pHs, absorbance is blue shifted from ~564 nm to ~420 nm and increases by 19.74 and 19.23 $mM^{-1}cm^{-1}$ at pH 4 and 5 respectively, when compared to the control sample at 420 nm. Absorbance at between pH 6-8 stays relatively the same with a decrease of 2.74 $mM^{-1}cm^{-1}$ at pH 6 and an increase of 0.67 $mM^{-1}cm^{-1}$ at pH 7 compared to control protein at pH 8. At the same time, fluorescence is higher by 6 % at pH 6 and 17 % at pH 7 compared to control emission at pH 8 (Table 4.3). Emission on excitation at 564 nm in higher pH states is reduced by 34-39 % as peak absorbance has shifted to ~544 nm (Table 4.3). This shift in absorbance is again reflected in emission on excitation at 544 nm. On excitation at 544 nm peak emission for control sample is observed at 570 nm whereas peak emission for samples in alkaline conditions shifts to a λmax of ~560 nm. Relative to the control peak emission at pH 8, a 3- to 4-fold enhancement in fluorescence intensity is observed at pH 9-11, where the emission maxima remain constant at 560 nm.

In addition, $mCherry^{M66C}$ lost its fluorescence in acidic conditions. This is demonstrated spectrally as on excitation at 564 nm an 82-87% decrease in emission is observed (Table 4.3). Further evidence is presented on Figure 4.6d where the protein was excited at 420 nm and emission intensity is less than 0.1 a.u. This was expected as most fluorescent proteins lose their fluorescence in acidic conditions (Shinoda et al. 2018).

Table 4.3: Emission of mCherry^{WT} and mCherry^{M66C} upon excitation at λ_{max}

pH	mCherry^{WT} Emission (a.u.) (excitation at 587 nm)	Fold Change (%)	pH	mCherry^{M66C} Emission (a.u.) (excitation at 564 nm)	Fold Change (%)
4.5	0.68	32	4.5	0.13	87
5	0.81	19	5	0.18	82
6	0.97	3	6	1.06	6
7	0.90	10	7	1.17	17
8	1.00	-	8	1.00	-
9	1.02	2	9	0.61	39
10	0.90	10	10	0.66	34
11	0.91	9	11	0.66	34

Two major shifts in λ_{max} can be observed from the absorbance data of the mCherry^{M66C}. Spectral analysis shows that the mCherry^{M66C} chromophore exhibits a three-state equilibrium model (Figure 4.9) governed by two pKa values, consistent with sequential protonation events. The 566 nm species dominates at intermediate pH states (6-8) and serves as the central intermediate in two pH-dependent transitions. At acidic pH (4.5-5), the mCherry^{M66C} chromophore transitions to a protonated form leading to a shift in absorbance from 566 nm to 420 nm with a pKa at 5.7 (Figure 4.8a). At alkaline pH (9-11) the intermediate species transitions to a deprotonated species leading to a shift in absorbance from 566 nm to 544 nm with a pKa of 8.8 (Figure 4.8b). This dual transition system suggests distinct proton-binding events affecting the chromophore environment especially since a similar change is not observed in wild type.

Upon the replacement of the methionine with the cysteine, there is now a thiol group available that can be deprotonated in addition to the phenol group. For fluorescent proteins in general, the main pH sensitive group is the phenol hydroxyl group. The pKa for this groups varies but is generally in the pH 4-6 region (Cranfill et al. 2016). This also holds true here with mCherry with the single transition with a pKa of 4.4 (Figure 4.7). Thus, it can be proposed that the major spectral changes in acidic pH is due to protonation of the phenol chromophore. In Figure 4.9 below, the neutral pH form is likely to be represented by top, deprotonated phenol form. The second transition has a pKa of 8.8, which is close to

the pKa of the thiol group (8-8.5). Thus, the alkaline form is likely to be the double deprotonated form.

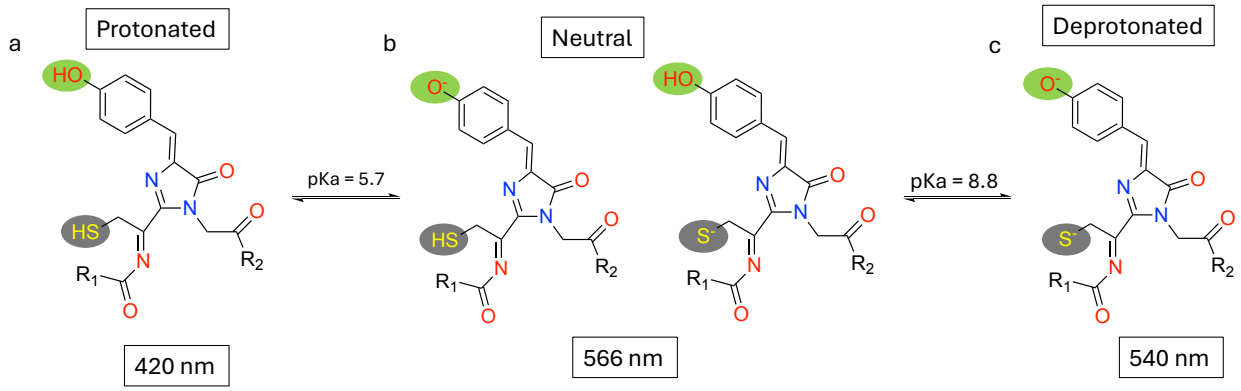

Figure 4.9: Three-state model of the mCherry^{M66C} chromophore. a) Protonated form of the chromophore in acidic conditions, b) neutral forms of the chromophore in pH6-8 and c) deprotonated chromophore form in alkaline conditions. Each form corresponds to a specific λmax. There are two transition states between the three chromophore species each associated with a specific pKa.

To investigate this further, quantum mechanics (QM) simulations were performed in collaboration with Prof Jamie Platts (School of Chemistry, Cardiff University), have confirmed this idea (Table 4.4). The aim was to predict the spectral properties of each of the forms shown above in Figure 4.9.

In its neutral state the chromophore can potentially exist in two forms as suggested by the QM data. One state where the oxygen is protonated and one where only the thiol group is protonated. In alkaline conditions both get deprotonated (Figure 4.9).

Table 4.4: Predicted Absorbance at different protonation states of wild type mCherry and mCherry^{M66C}. Predicted absorbances and their associated protonation state.

mCherry^{WT}			mCherry^{M66C}		
Protonation state	S0-S1/nm	Strength	Protonation state	λ/nm	Strength
OH	544	1.41	SH, Ph-OH	422	1.26
Ph-O-	584	1.72	S^-, Ph-OH	601 572 440	0.06 0.01 1.32
			SH, Ph-O^-	557	1.69
			S^-, Ph-O^-	541	0.96

The QM data has helped confirm the protonated and deprotonated states of mCherry^{M66C}. This is evident as the predicted absorbance λ_{max} at the fully protonated state (SH, Ph-OH) is 422 nm, which matches the observed λmax from the absorbance data collected in practice. In a similar manner the predicted major absorbance λ_{max} for the fully deprotonated chromophore (S^-, Ph-O^-) is 541 nm which almost matched the observed λ_{max} (Figure 4.6b). When it comes to the neutral form of the chromophore, there are two potential states it can take, one where only the thiol group is protonated and one where only the oxygen group is protonated. Based on the predicted absorbance values, one can argue that the two forms of neutral chromophore can exist at varying quantities that depend on smaller pH deviations. The dominant chromophore form is most likely the one where the thiol is protonated, and the oxygen deprotonated (anionic chromophore). QM data show a high oscillator strength at 557 nm for the thiol with the deprotonated oxygen, compared to the oscillator strength for any of the other wavelengths of the thiolate with the protonated oxygen. Although not an exact match to the experimental λ_{max} recorded it can still be attributed to the main excitation peak and the primary species of mCherry^{M66C} at pH 8. This is because compared to the other neutral pH states, at pH 8 the protein has a broader absorbance peak including 557 nm. The Ph-OH form (thiolate chromophore) has a strong oscillator strength for the 440 nm peak. On the other hand, wild type mCherry has two predicted protonation states, which would be expected as unlike mCherry^{M66C}, it only has one pKa (4.4).

Finally, another interesting attribute between the two proteins is the fact that upon deprotonation, absorbance of mCherry^{M66C} shift to a lower peak wavelength whereas in wild type the absorbance maxima remain the same. The fact that there is a distinct change in absorbance with changes in pH can potentially allow mCherry^{M66C} to be used as a biosensor to monitor changes in pH within cells. This distinct shifts between chromophore species allows for an easy identification and distinction between pH states. As observed from the recorded spectra (Figure 4.6b), upon the shifts in λ_{max} with the change in pH, the overall absorbance level changes as well. This can allow ratiometric quantification of the absorbance when imaging to help minimise error from sources other than a change in pH (Karsten et al. 2022).

4.2.4.2 The effect of H_2O_2 on spectral properties.

ROS play an important role in biology, with species such as hydrogen peroxide (H_2O_2) playing roles in signalling as a secondary messenger (Di Marzo et al. 2018) but more commonly as an oxidant that can damage biomolecules (Zhu et al. 2020). Thus, there is a need to detect ROS such as H_2O_2, with FPs being a potential source of such biosensors (Ermakova et al. 2014).

The two proteins were subjected to increasing concentrations of hydrogen peroxide and measurements before and after incubation for an hour. Proteins were tested both at pH8 (Figure 4.10) in which the protonation state of the thiol group [the most likely reactant group of the mCherry^{M66C} chromophore] is closer to the pH of the mitochondrion matrix, as well as at pH 7 (Figure 4.11) which is the most commonly sampled pH in biological systems (Xiong et al. 2008).

At pH 8 a steady decrease in the absorbance of both mCherry and mCherry^{M66C} can be observed in the presence of H_2O_2 after an hour of incubation. A 3 and 2-fold decrease is observed in the absorbance and emission of mCherry^{M66C} after an hour with 0.5 % H_2O_2 respectively, and a 13 and 33-fold decrease in absorbance and emission of mCherry^{M66C} with 5 % H_2O_2 respectively (Table 4.5). For mCherry an 18 and 14-fold decrease is observed in absorbance and emission after an hour of incubation with 0.5 % H_2O_2 and a 24 and 50-

fold decrease in absorbance and emission after an hour of incubation with 5% H_2O_2 respectively (Figure 4.10 and Table 4.5).

One noticeable trend is that mCherry is more susceptible to oxidation compared to mCherry^{M66C}. Despite absorbance and fluorescence decreasing for both over time and with increasing H_2O_2 concentration, mCherry spectral properties are affected more at lower hydrogen peroxide concentrations than mCherry^{M66C} (Table 4.5).

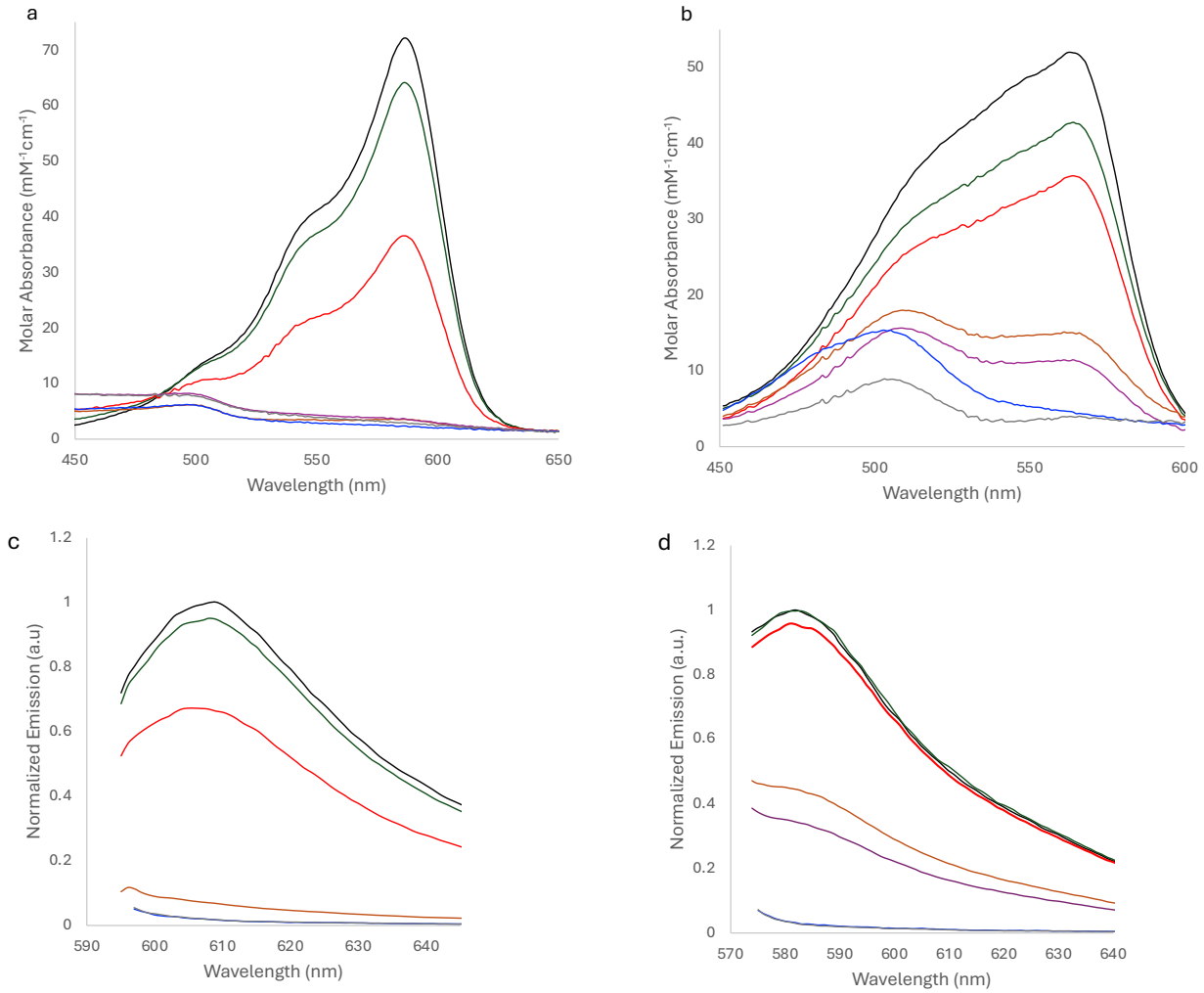

Figure 4.10: Effect of increasing H_2O_2 concentrations at pH 8. Absorbance spectra of a) mCherry and b) mCherry^{M66C} after an hour of incubation with different H_2O_2 concentrations. Emission spectra of c) mCherry and d) mCherry^{M66C} an hour after incubation with H_2O_2. Control sample without H_2O_2 (black), 0.01 % H_2O_2 (green), 0.1 % H_2O_2 (red), 0.5% H_2O_2 (brown), 1 % H_2O_2 (purple), 2.5 % H_2O_2 (blue) and 5 % H_2O_2 in grey.

A similar trend is observed for both proteins at pH 7 too. A 4 and 2-fold decrease in absorbance and emission after an hour incubation with 0.5 % H_2O_2 and a 105 and 5-fold decrease in absorbance and emission after being incubated for an hour with 5 % H_2O_2 respectively, is observed for mCherry^{M66C}. In a similar fashion, absorbance and emission at 0.5 % decreased by 9 and 5-fold and at 5 % it decreased by 24 and 11-fold for mCherry, respectively (Table 4.5). The reason those specific points were selected is because at 0.5 % is the midpoint where a significant decrease in spectral properties is detected and 5 % is the end point of the experiment where it shows the biggest change among the different samples and between the two pH states.

Table 4.5: Changes in spectral properties of mCherry and mCherry^{M66C} in the presence of H_2O_2 after an hour at pH 8 and pH 7.

H_2O_2 (%)	mCherry^{M66C} (pH 8)		mCherry (pH 8)	
	Abs ($mM^{-1}cm^{-1}$)	Em	Abs ($mM^{-1}cm^{-1}$)	Em
0	52	1	72	1
0.01	43	1	64	0.95
0.1	36	0.96	37	0.66
0.5	15	0.44	4	0.07
1	11	0.34	4	0.05
2.5	5	0.03	2	0.01
5	4	0.03	3	0.02
H_2O_2 (%)	mCherry^{M66C} (pH 7)		mCherry (pH 7)	
	Abs ($mM^{-1}cm^{-1}$)	Em	Abs ($mM^{-1}cm^{-1}$)	Em
0	42	1	72	1
0.01	42	0.98	68	0.91
0.1	34	0.83	38	0.53
0.5	11	0.43	8	0.14
1	9	0.38	5	0.1
2.5	2	0.24	4	0.09
5	0.37	0.21	3	0.09

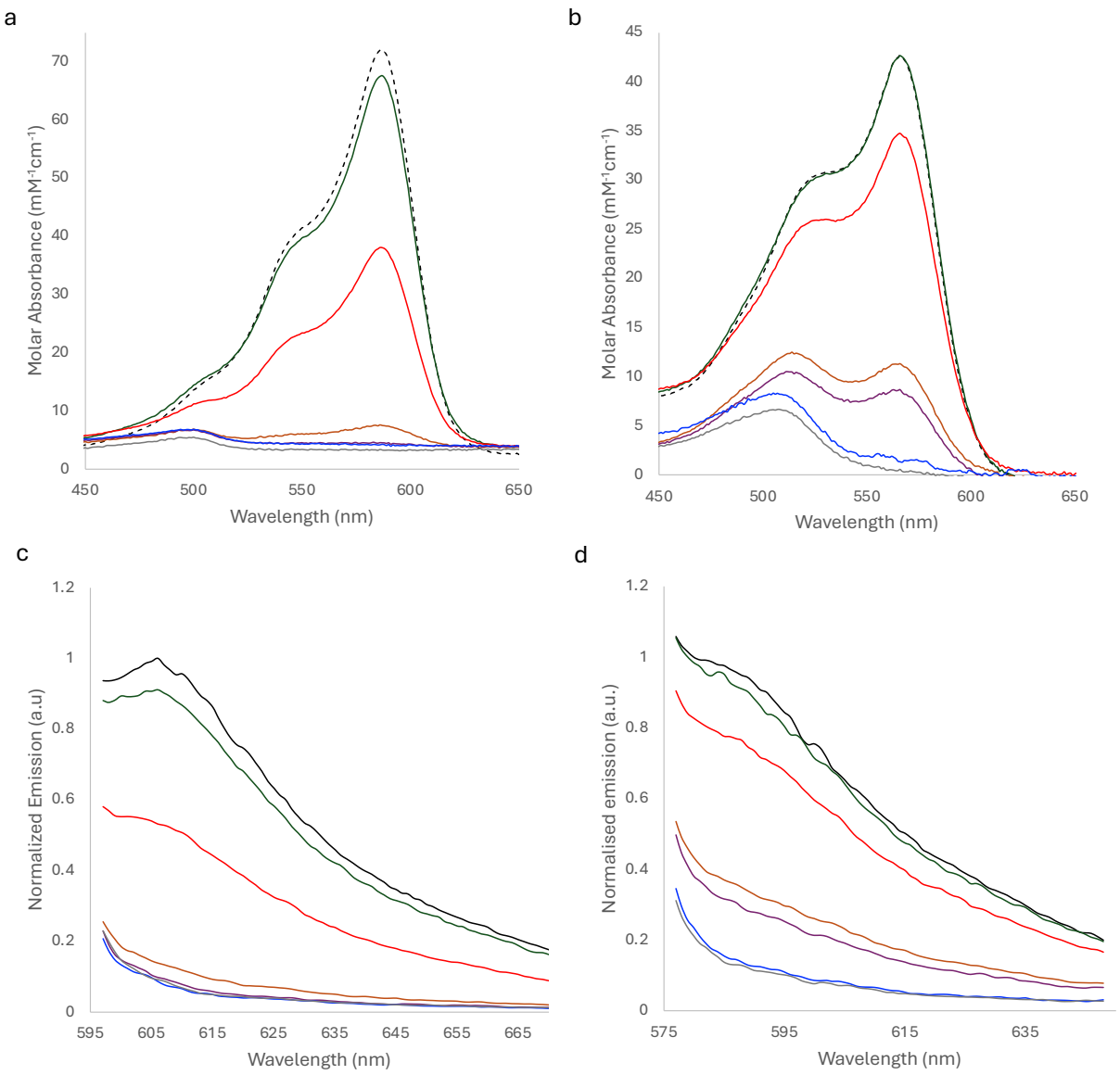

Figure 4.11: Effect of increasing H_2O_2 concentrations at pH 7. Absorbance spectra of a) mCherry and b) mCherry^{M66C} after an hour of incubation with different H_2O_2 concentrations. Emission spectra of c) mCherry and d) mCherry^{M66C} an hour after incubation with H_2O_2. Control sample without H_2O_2 (black), 0.01 % H_2O_2 (green), 0.1 % H_2O_2 (red), 0.5 % H_2O_2 (brown), 1 % H_2O_2 (purple), 2.5 % H_2O_2 (blue) and 5 % H_2O_2 in grey.

Upon plotting the λ_{EM} values against hydrogen peroxide concentration we can clearly see the trend of decrease for each sample at each pH (Figure 4.12). Emission for both mCherry and mCherry^{M66C} shows a steeper decrease at first that evens out after 0.5 % H_2O_2. We can see that at both pH states, emission of mCherry decreases much faster than

that of mCherry^{M66C}. All samples reach a short plateau between 0.5-1% and then massively drop to almost 0.1 at pH 7 and almost to 0 at pH8.

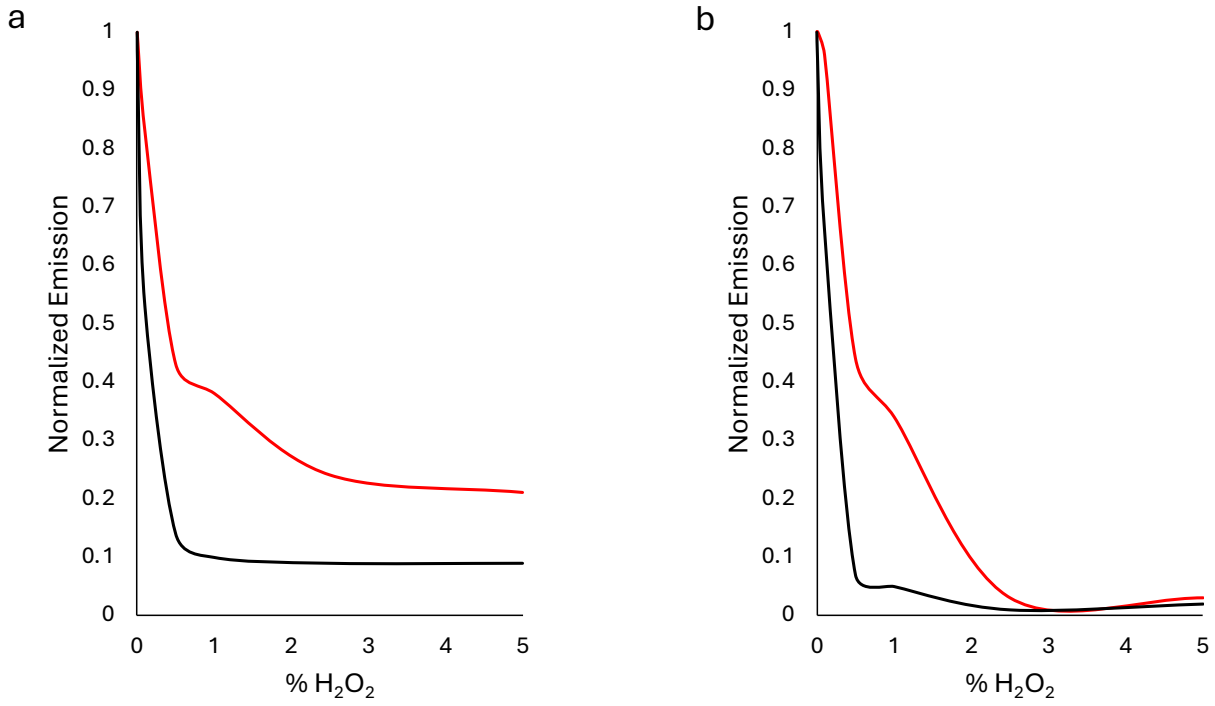

Figure 4.12: Decrease in emission of mCherry and mCherry^{M66C} with increasing H_2O_2 concentration. Loss of value plots for proteins at pH7 (a) and pH8 (b). Black and red lines represent the decrease in emission of mCherry and mCherry^{M66C} with increasing H_2O_2 concentrations respectively.

Although further information is required to draw better conclusions as to what might be happening at that point, it can be argued that the decrease at pH is not as high as expected in mCherry^{M66C} like we observed for the protein at pH 8. This can potentially be due to the presence of the different neutral chromophore species of at pH 7.

The overall loss in fluorescence is expected, as photobleaching by hydrogen peroxide has been previously observed in FPs (Vegh et al. 2014). However, since hydrogen peroxide plays a role in various cellular processes such as cell signalling and oxidative stress (Liguori et al. 2018; Heo et al. 2020) having FPs that can withstand chromophore disruption or possess unique spectral properties in the presence of H_2O_2 would be a great tool for bioimaging those events.

4.2.4.3 Mass Spectrometry of proteins in the presence of H_2O_2

To further analyse and understand the potential structural changes that give rise to the observed spectral properties (Figure 4.6) of the mCherry proteins discussed in this chapter, LCMS was carried out on both the wild type and mutant proteins at pH8. For this analysis, two samples were prepared for each protein one where 0.5% of H_2O_2 was added and one without hydrogen peroxide added. Hydrogen peroxide was added to the protein before loading on the column, therefore the mass spectrometry data will mostly relate to the spectral data recorded upon the addition of hydrogen peroxide. Predicted mass for each protein version was calculated using the Compute pI/Mw Expasy tool.

Table 4.6: Predicted mass of mCherry^{WT} and mCherry^{M66C}

Protein Version	Predicted mass (Da)
Full mCherry^{WT}	27689
Full mCherry^{M66C}	27661
mCherry^{WT} (No N-terminal methionine)	27537
mCherry^{M66C} (No N-terminal methionine)	27509

The measured mass for each protein was accurate to within a Da, when taking into consideration maturation (loss of 20 Da) and removal of the N-terminal methionine. The latter is commonplace in biology as part of a common post-translational modification where the methionine is recognized by methionine aminopeptidase (MAP) and cleaved off (Falb et al. 2006). This is ultimately done for protein maturity and stability. It is an evolutionary conserved process, as it does not affect the function of the protein especially fluorescent proteins as the chromophore is buried well within the chromophore far away from the N-terminus Met thus making it irrelevant for the function of the protein.

As expected, the predicted mass of the two proteins has a 28 Da difference which accounts for the change from methionine (149 Da) to cysteine (121 Da). Mass spectrometry data collected for mCherry^{WT} (Figure 4.13) and mCherry^{M66C} (Figure 4.14) in both the presence and absence of hydrogen peroxide confirmed the presence of the full protein. In the presence of hydrogen peroxide, a decrease of ~58 Da is observed in the mass of both proteins. This suggests that both proteins experience the same overall

modification in structure. This would be expected as the data for both proteins show a similar effect in their spectral data but at different rates, with mCherry^{M66C} being more resistant. Since the overall effect of hydrogen peroxide is the same for both mCherry and the mutant in terms of how the protein structure is affected, this suggests that the change is independed to the mutation and probably has to do with the whole of the chromophore environment. However, the exact mechanism is unclear. Normally, oxidation of sulphur containing amino acids occurs through the addition of 1-2 oxygens at the sulphur (DeLeon et al. 2016). This would generate an increase in mass (16 or 32 Da) rather than a decrease.

The loss of ~58 Da would equate to loss of a glycine residue from a terminal but this is unlikely as the terminal sequences are N-MVSKGEE.....KGSHHHHHH-C. The alternative is the loss of a side chain component. Oxidation can result in the modification of various side chains, including cleavage of certain portions. For example, oxidation of methionine side chains (9 in addition to the one resident in the mCherry chromophore) can result in the formation of formyl-glycine side chain (-CH2–CH2–S–CH3 → -CH(OH)–CHO or CHO–CH2OH). This equates to a loss of just 30 Da and there are no additional methionine residues close to the chromophore. An alternative is the side-chain modification of arginine (arginine → glutamyl semialdehyde (–CH_2–CH_2–CH_2–CH(NH_2)–CHO)) but again the mass loss is too low, being 28 Da. However, R95 is known to play a critical role in mCherry fluorescence (Pletnev et al. 2010) and the loss of the guanidino group (-CH_2-NH-(NH_2), results in a mass loss of 58. This does not account for any sulphur oxidation events that also likely to occur. Lysine can also be oxidized to a semialdehyde (–$(CH_2)_4$–NH_2 → –$(CH_2)_3$–CH(=O)–NH_2) with a mass loss of 28 Da. Loss of an Asn or Asp side chain equates to ~58 Da but this also does not consider any mass increase on oxidation of sulphur. Thus, while the mass spectra suggest the same molecular event is happening in both proteins, the exact nature of the chemical modification is currently unknown and is likely to comprise a mixture of modifications.

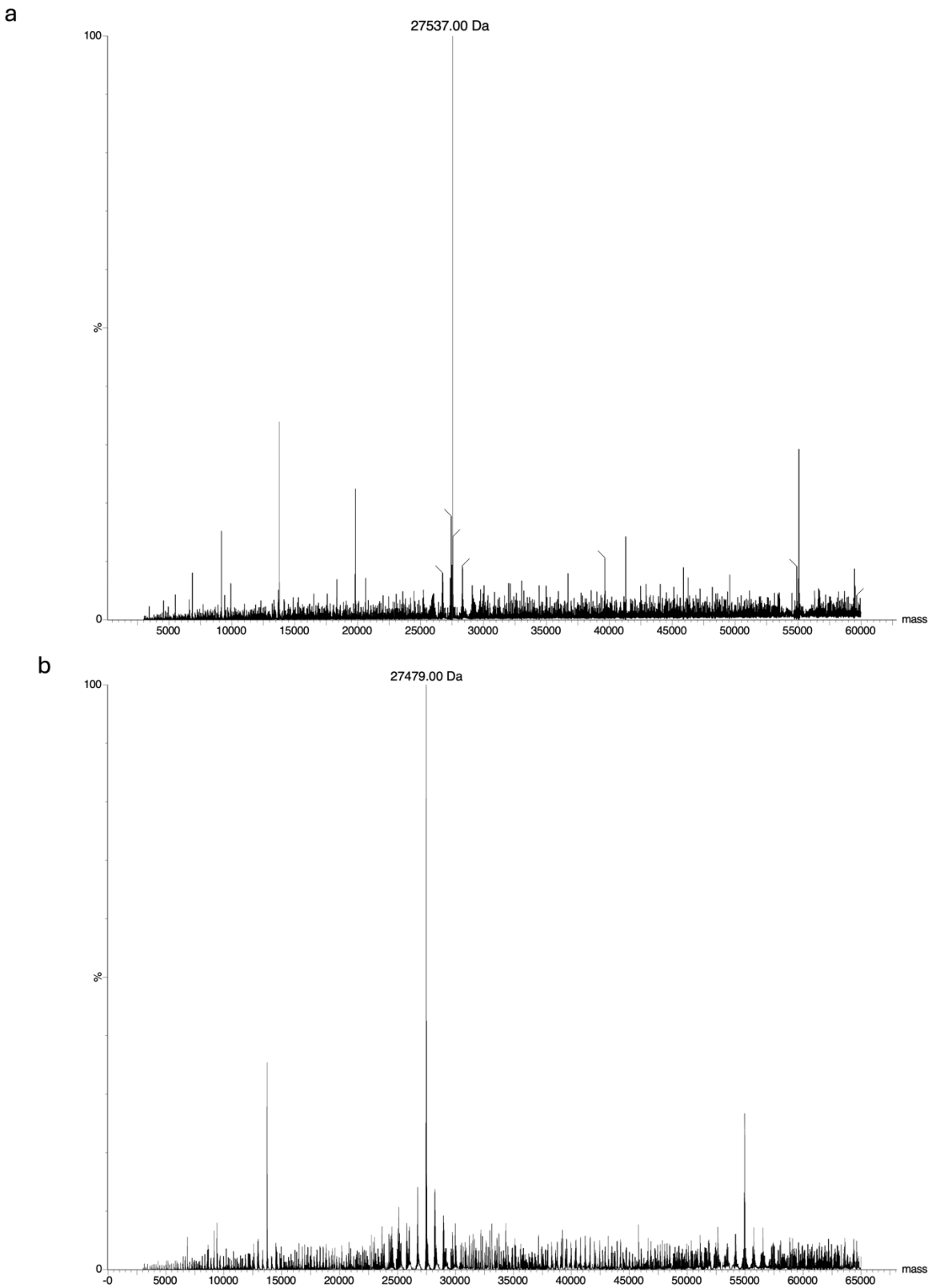

Figure 4.13: Mass spectrometry of wt mCherry. a) Mass corresponding to mCherry without hydrogen peroxide and b) mass corresponding to mCherry with added hydrogen peroxide.

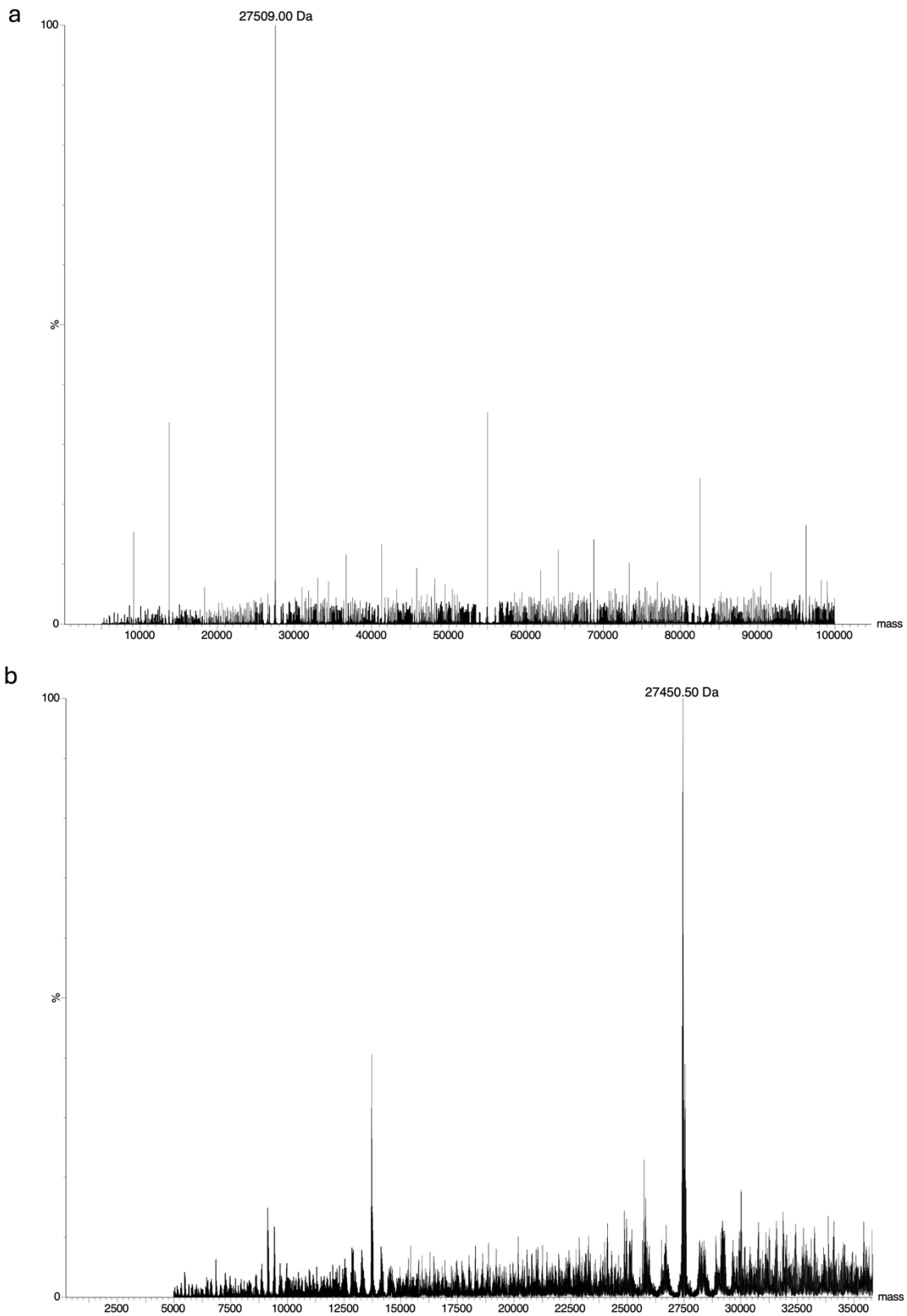

Figure 4.14: Mass spectrometry of mCherry^{M66C}. a) Mass corresponding to mCherry^{M66C} without hydrogen peroxide and b) mass corresponding to mCherry^{M66C} with added hydrogen peroxide.

4.2.5 Structure determination of mCherry^{M66C}

DsRed, was the first red fluorescent protein to be discovered and characterised in 1999. Despite being derived from a non-bioluminescent reef coral, DsRed has a homologous β-barrel structure to the one that was first observed for GFP. This new discovery along with five additional homologs gave new insight into the structure-function relationship of β-barrel proteins, as well as identified that FPs are not solely derived from bioluminescent organisms (Matz et al. 1999). The chromophore is always comprised of 3 contiguous residues with similar primary structure positions: 66-XYG-68 (DsRed numbering). Further comparative studies on the different variants showed the conservation of Gly67 across all β-barrel fluorescent proteins as it is crucial for the cyclization step for chromophore formation (Heim et al. 1995; Matz et al. 1999). The tyrosine is also conserved amongst the naturally occurring as it forms the phenol ring component of the chromophore. However, it can be mutated to other aromatics residues without total loss of chromophore maturation but impacts on spectral properties (e.g. mHoneydew (Shaner et al. 2004)). Out of the three residue 66 is the one that has been targeted the most across different RFP variants. Originally a glutamine in DsRed, residue 66 has previously been replaced with predominantly methionine as it was shown to provide a more complete chromophore maturation (Shaner et al. 2004). Other amino acids have been sampled including leucine, threonine, phenylalanine and cysteine.

Cysteine has previously been incorporated at residue 66 to generate proteins like mRFP1-Q66C and mTangerine. When compared to my engineered mutant their spectral properties are similar in terms of absorbance maxima. However, there is currently no published data on their protein structure or further analysis into the dynamics of the chromophore and how the cysteine affects the chromophore environment.

Here we solved the structure of the mCherry^{M66C} mutant followed by MD analysis that provided structural information which helped to relate to the overall function and spectral changes described in this chapter.

4.2.5.1 X-ray crystallography of mCherry^{M66C}

To get the crystal structure of mCherry^{M66C}, purified protein was concentrated down to 10mg/mL and taken to Diamond Light Source for crystallisation and x-ray diffraction. To determine the structural effects of incorporating a cysteine into the chromophore at position 66 of mCherry, the structure was solved by x-ray crystallography. Crystal trials were set up at Diamond light source, and crystals were harvested and subjected to x-ray scattering and diffraction using the I03 beamline. To solve the structure, molecular replacement was performed using the structure of wild type mCherry (PDB: 2H5Q) as a model, incorporating the altered chromophore as required. Structure refinement was done using REFMAC5 via CCP4 suite. Full diffraction data and refinement statistics can be found on Table 4.7. After refinement, data allowed for a final structure of a resolution at 2.04 (Å).

Table 4.7: Data collection and refinement statistics for mCherry^{M66C} mutant. One crystal was used for determining each structure. [1]Coordinate Estimate Standard Uncertainty in (Å), was calculated based on maximum likelihood statistics. Any figures in brackets correspond to outer resolution shell.

Crystal data collection	
Wavelength (Å)	0.81532
Beamline	I03
Space group	P 1 2_1 2
a, b, c (Å)	48.760, 43.182, 63.095
α, β, γ (^{O})	90.0, 90.114.91, 90.0
Resolution (Å)	2.04 – 45.46
Outer shell	2.04 – 2.10
R-merge (%)	9.1 (46.7)
R-pim (%)	9.1 (46.7)
R-meas (%)	12.8 (66.1)
CC1/2	0.980 (0.413)
I / σ(I)	20.5 (4.7)
Completeness (%)	97.3 (98.3)
Multiplicity	1.8 (1.8)
Total Measurements	27152 (2107)
Unique Reflections	14902 (1149)
Wilson B-factor (Å^2)	20.2
Refinement Statistics	
Refined atoms	1853
Protein atoms	1761
Non-protein atoms	5
Water molecules	87
R-work reflections	13309
R-free reflections	1376
	16.98 / 21.06
rms deviations (ML target in brackets)	
Bond lengths (Å)	0.012 (0.013)
Bond angles (^{O})	1.553 (1.661)
[1]Coordinate error	0.111
Mean B value (Å^2)	23.4
Ramachandran Statistics (PDB Validation)	
Favoured / allowed / Outliers	208 / 9 / 0
%	95.9 / 4.1 / 0.0

4.2.5.2 Structure analysis of mCherry^{M66C}

Analysis of the crystal structure revealed the expected monomeric β-barrel structure with the chromophore situated in the centre of it. At first look incorporation of cysteine in the place of methionine seems to free up space within the surrounding chromophore environment as cysteine has a shorter side chain to that of methionine. Residues within 4 Å of the chromophore were identified to see if they matched residues that have been engineered in other red proteins. The reason for this was to determine the potential environment of residues that can be responsible for the observed spectral changes of the protein. These include residues S62, P63, F65, W143, S146, Q163, I197, Q213 and E215 (Figure 4.15). This is important, as residues like F65, W143, S146 and Q163 play a key role in the hydrogen bond network of the chromophore as well as chromophore rigidity, and in extend the effect that has in the protein's brightness and photostability (Chica et al. 2010; Mukherjee et al. 2023). Residue R95 is also included to show its proximity to the chromophore, as it is known to have a role in fluorescence and as mentioned before loss of its guanidino group can potentially lead to the loss of 58 Da observed upon oxidation in the presence of H_2O_2.

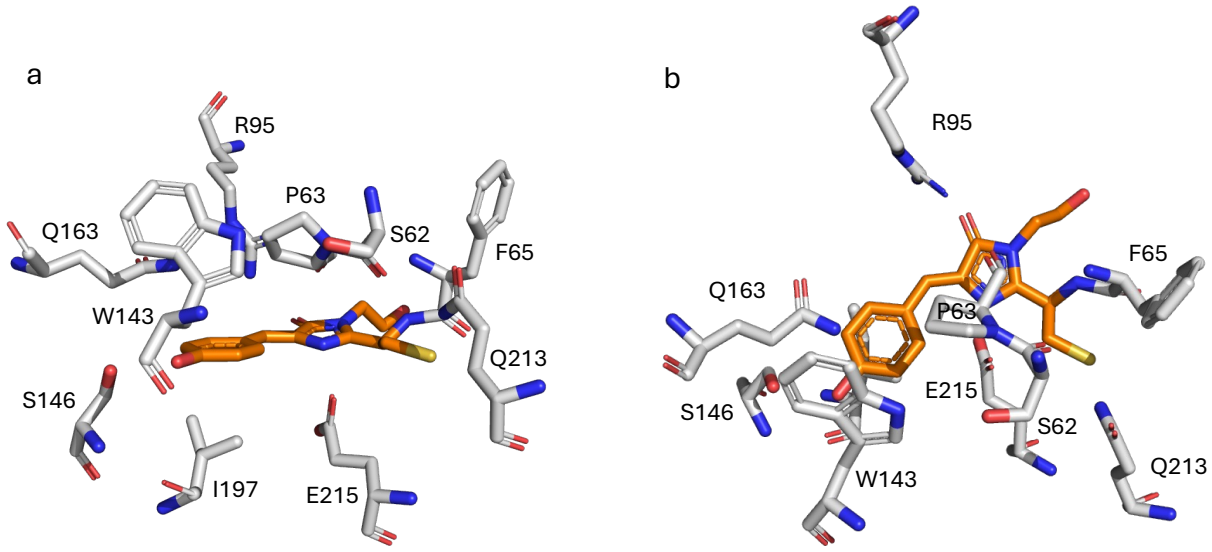

Figure 4.15: Key residues around the chromophore. a) Side and b) top view of the mCherry^{M66C} chromophore (orange) and key residues (white).

In addition to the residues surrounding the chromophore, there are also seven water molecules around the chromophore within 4 Å (Figure 4.16). We can see in the crystal structure that the extra space available because of the shorter side chain of cysteine (compared to methionine) is occupied by a water molecule (W7).

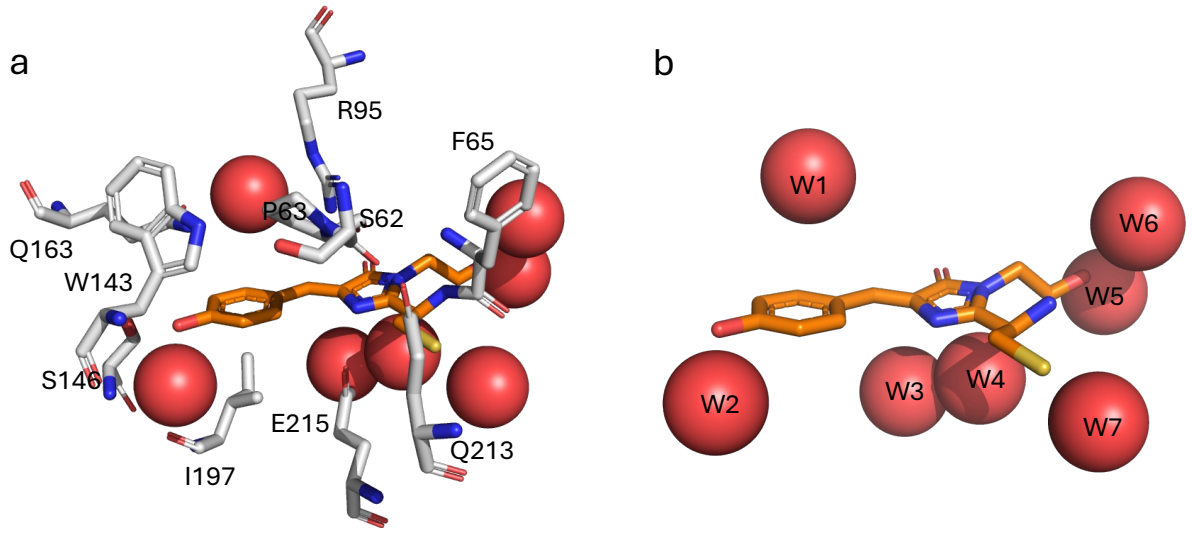

Figure 4.16: Water molecules around the mCherry^{M66C} chromophore. a) and b) show the mCherry^{M66C} chromophore with key residues and water molecules and just water molecules respectively. The chromophore is presented in orange sticks, water molecules as red spheres and key residues in white sticks.

4.2.5.3 Comparison of mCherry^{M66C} to mCherry

To understand the key structural differences upon the incorporation of cysteine in residue 66 of mCherry comparisons were made with the structure of mCherry (Figure 4.17). Initial alignment of the two structures in PyMOL showed they had similar overall structure with the Cα root mean square deviation (RMSD) of 0.152 Å. This is not surprising as there is only a single mutation within the chromophore of mCherry, hence it is unlikely to have a significant effect on the backbone of the protein. However, although an effect in the overall β-barrel is not observed, as mentioned before a bigger pocket is now available within the chromophore environment. Because of that increase in space, it was hypothesised that there is now the potential for small molecule transport through the protein including higher water molecule accessibility within the chromophore environment. Information from the

crystal structures indeed show that a water molecule now occupies that space (Figure 4.18a). However, it is not just that space being occupied by a water molecule. The crystal structure of mCherry^{M66C} reveals that there are seven water molecules occupying space around the chromophore, while there are only six water molecules occupying space around the mCherry chromophore (Figure 4.18). Upon comparison with the structure of wild type mCherry, a bigger pocket was identified within the chromophore where it was previously occupied by a bulkier thioether group as it is now occupied by a thiol group instead (Figure 4.18).

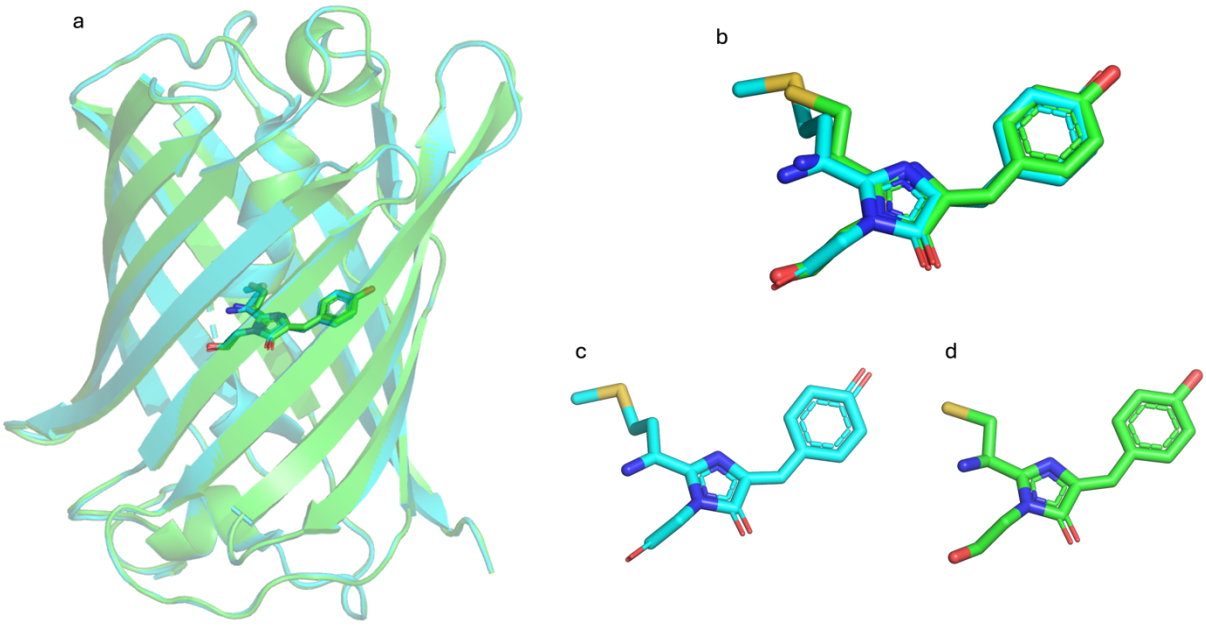

Figure 4.17: Structural differences between mCherry^{M66C} and mCherry^{WT}. mCherry^{M66C} is shown in green and mCherry^{WT} (PDB 2H5Q) is shown in cyan. a) Cartoon representation of the wild type against mCherry^{M66C}. b) Stick figure representation of the mCherry chromophore against the mCherry^{M66C} chromophore. c) and d) are stick figures of the mCherry and mCherry^{M66C} chromophore respectively.

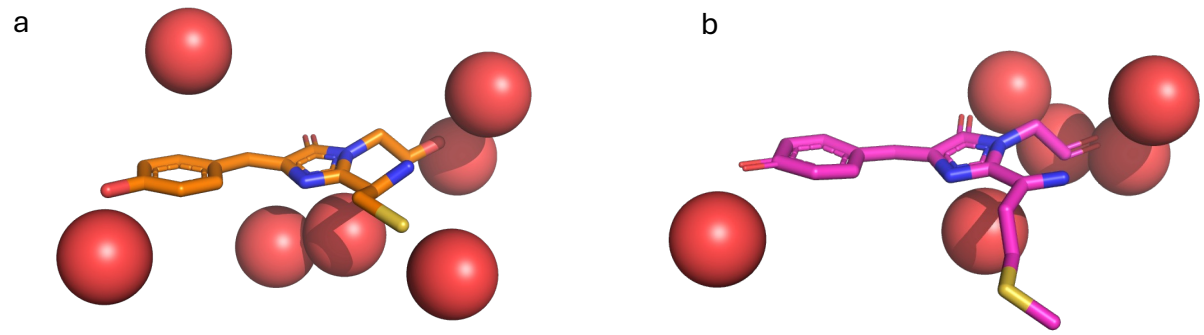

Figure 4.18: Water molecules around the chromophore. Water molecules occupying space around the a) mCherry^{M66C} (orange) and b) mCherry (pink) chromophore.

While the crystal structure of mCherry (Shu et al. 2006) and mCherry^{M66C} provide some insights into the functional differences and sensitivity to environmental factors between the two, proteins are inherently dynamic and thus should be taken into consideration when assessing the structure-function relationship. Further experimental and computational approaches were carried out to help pinpoint the nature of the molecule and help understand the dynamics of the chromophore environments.

4.2.6 Computational analysis of mCherry^{M66C} and mCherry^{WT}

Different computational approaches were undertaken to help identify H-bonding of the chromophore with its surrounding environment, as well as water transport across the protein and the chromophore, and determine how these interactions can potentially influence the spectral properties of our protein. These include molecular dynamic simulations carried out by Prof Jones and Dr Menzies.

4.2.6.1 Molecular Dynamics analysis

Both variants were subjected to molecular dynamics simulations (3 x 1000 ns at 300 K) to understand how their structural differences manifest as changes in dynamics, which in turn change the interactions between the chromophore, key residues and water molecules. Compared to the structural data obtained which give a “stationary” idea of the

changes in space within the chromophore environment and limited information as to the potential changes in interactions, this *in silico* analysis provided further information about how the dynamics of the chromophore changed and how it moves over time within its environment. To do that, different analyses including the root mean square deviation (RMSD), root mean square fluctuations (RMSF) of both the backbone and the chromophore were carried out. These provided information on the level of structure fluctuations of the whole β-barrel structure but also more specifically on the chromophore. These allowed for a comparison between mCherry^{WT} and mCherry^{M66C} and gave an inside into how the rigidity or flexibility of the structure and more specifically the chromophore can relate to the spectral properties observed in Section 4.2.3.

The backbone (Cα) RMSD analysis for both proteins revealed that their overall structure of each variant reaches a stable plateau without any gross fluctuations over the course of the simulation (Figure 4.19 a and b).

While the overall backbone Cα RMSD provides an overview of protein stability, the RMSD of specific components (e.g. the chromophore) allows us to dissect the contribution of individual element. Thus, RMSD analysis was performed on the chromophore to assess the effect of the M66C mutation on chromophore stability (Figure 4.19).

For mCherry the RMSD analysis of the chromophore showed fluctuations across all three simulations (Figure 4.19c). As can be seen in Figure 4.18c, there are clears steps between states with values ~0.15 nm to states with values of around 0.05 nm. Upon analysing the conformation of the chromophore at the different timepoints at which various discrete conformational states were sampled over the simulation, we could see that there was some fluxing around the C-N (Figure 4.19a) that leads to a change in orientation of the oxygen (Figure 4.19a). That change is potentially important as the change happens at a critical dihedral angle which involves the extra bond that gives RFP's their colour properties; essentially the final two double bonds in the conjugated network are taken out of the plane of the rest of the system (Figure 4.20). In contrast, the RMSD analysis of the mCherry^{M66C} chromophore (Figure 4.19d), suggests that chromophore reaches a largely stable state with smalls fluxes. By extracting structures at various time points, the

small-scale fluxes are down to positioning of the F65 side chain and the thiol group. It should be mentioned here that while F65 is not part of the chromophore, it has been included in the MD simulations as the conjugated double bond network extends to the peptide bond between F65 and M66/C66. So, for accurate parameterisation of the chromophore, F65 was included as part of the chromophore in simulations.

The slight flexibility observed comes from the phenylalanine residue at position 65 which is not part of the chromophore structure (Figure 4.19b).

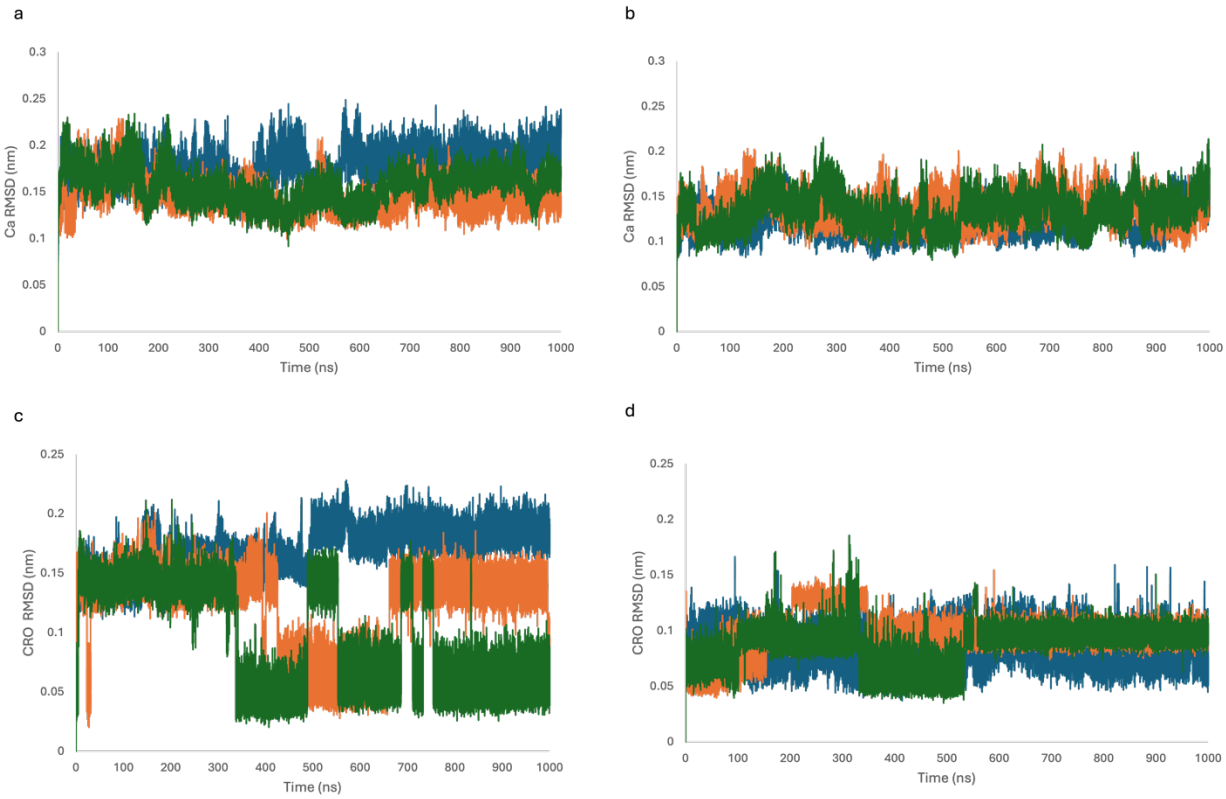

Figure 4.19: RMSD analysis of mCherry^{WT} vs mCherry^{M66C}. RMSD profile of the protein backbone of mCherry (a) and mCherry^{M66C} (b). RMSD profile of the chromophore of mCherry (c) and mCherry^{M66C} (d). Each colour corresponds to a simulation i.e. simulation 1 (blue), simulation 2 (orange), simulation 3 (green).

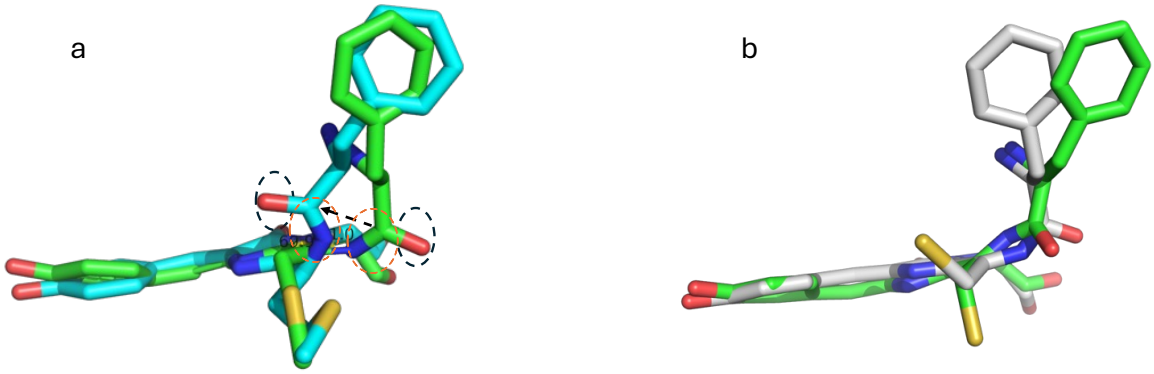

Figure 4.20: Stick representation of residue flexibility in the chromophore. Structure in green represents the initial chromophore state, while the ones in cyan and white represent the chromophore conformation at 300 ns. a) Chromophore conformations for mCherry^{WT}. b) Chromophore conformations for mCherry^{M66C}. The oxygens that change in orientation are marked by black dashed circles, while the C-N bond is marked by orange dashed circles. The shift from one orientation to the other is illustrated by a black arrow.

The RMSF profile is used to analyse the flexibility of each residue within the protein as well as individual atoms of the chromophore and highlights any regions of fluctuation within the different mCherry proteins (Figure 4.21 a and b). Most residues exhibit RMSF values under 1.5 Å throughout the simulations, which indicates very low fluctuation for the majority of the protein. A few regions and more specifically around residues 60-80 and 160-180 show an RMSF higher than that. The Cα RMSF profile of both the mCherry was used to compare with that of mCherry^{M66C} (Figure 4.21c). Upon comparison of the RMSF difference of the protein backbone between the two proteins, it is evident that the residues of interest (60-70) which include the chromophore forming residues and its immediate environment, shows a higher flexibility for mCherry. Once again confirming the idea that the mCherry^{M66C} chromophore is fluxing less than that of mCherry, hence showing more rigidity.

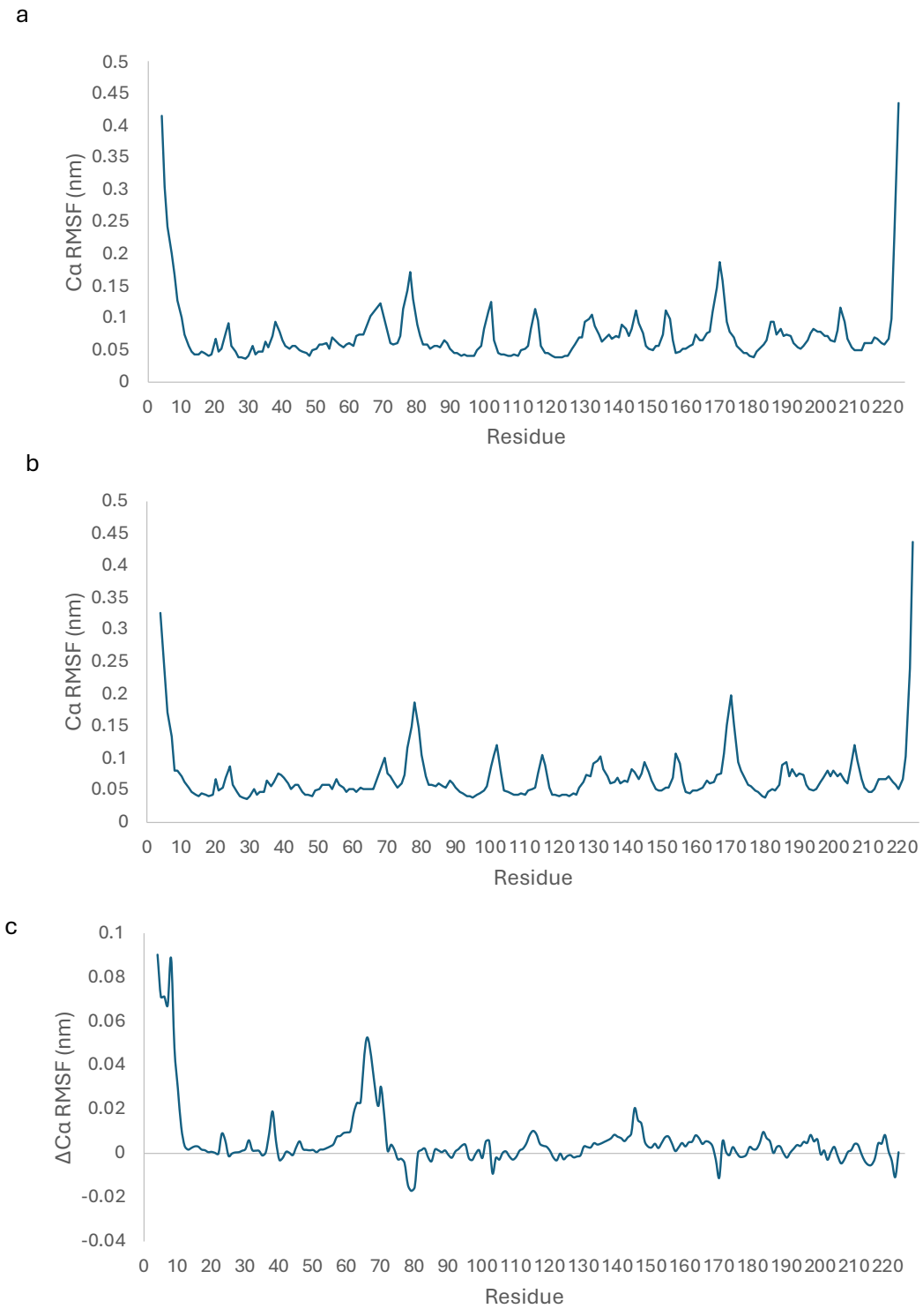

Figure 4.21: RMSF analysis of protein backbone. The average Cα RMSF analysis for mCherry (a) and mCherry^{M66C} (b) residues across the 3 simulations. The average Cα RMSF difference was calculated by subtracting the mCherry^{M66C} value from that of mCherry and was plotted to show the difference in residue flexibility between the two proteins (c) where any positive value means that the RMSF of mCherry^{WT} is higher than that of mCherry^{M66C} and thus more flexible.

When the flux of the individual atoms of the chromophore are assessed, RMSF analysis confirms that the main flux in the mCherry protein comes from the oxygen group (labelled O2) that changes orientation, while the rest of the chromophore remains fairly stable (Figure 4.22a). For mCherry^{M66C}, only the new thiol group shows some flexibility, but the rest of the chromophore is relatively stable with minimal movement (Figure 4.23a).

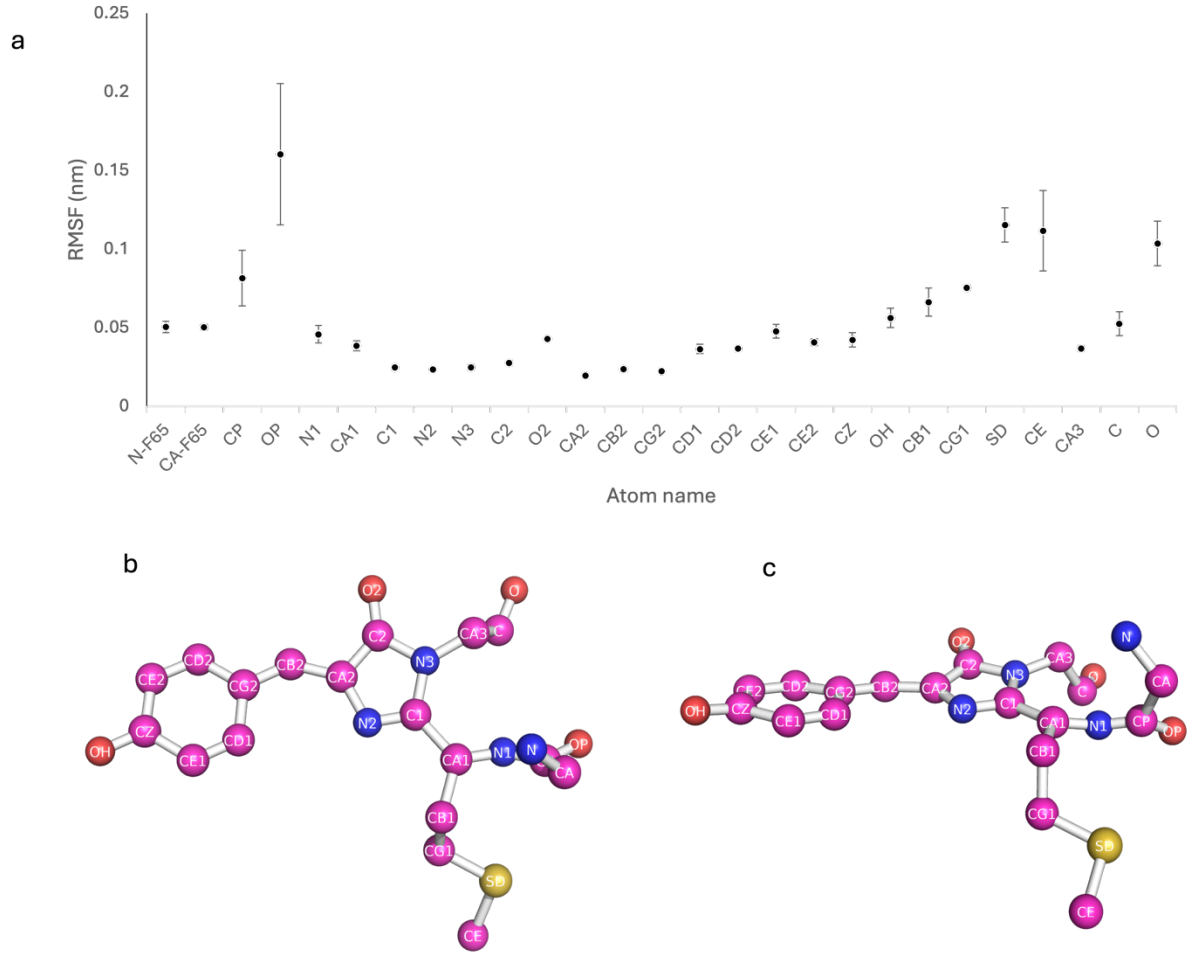

Figure 4.22: RMSF analysis of the mCherry^{WT} chromophore. a) Average RMSF values with standard deviation for each chromophore atom. b) Top and c) side view of the chromophore structure with individual atoms labelled. Shown is the average per atom RMSF over the 3 independent simulations with the error bar representing the standard deviation.

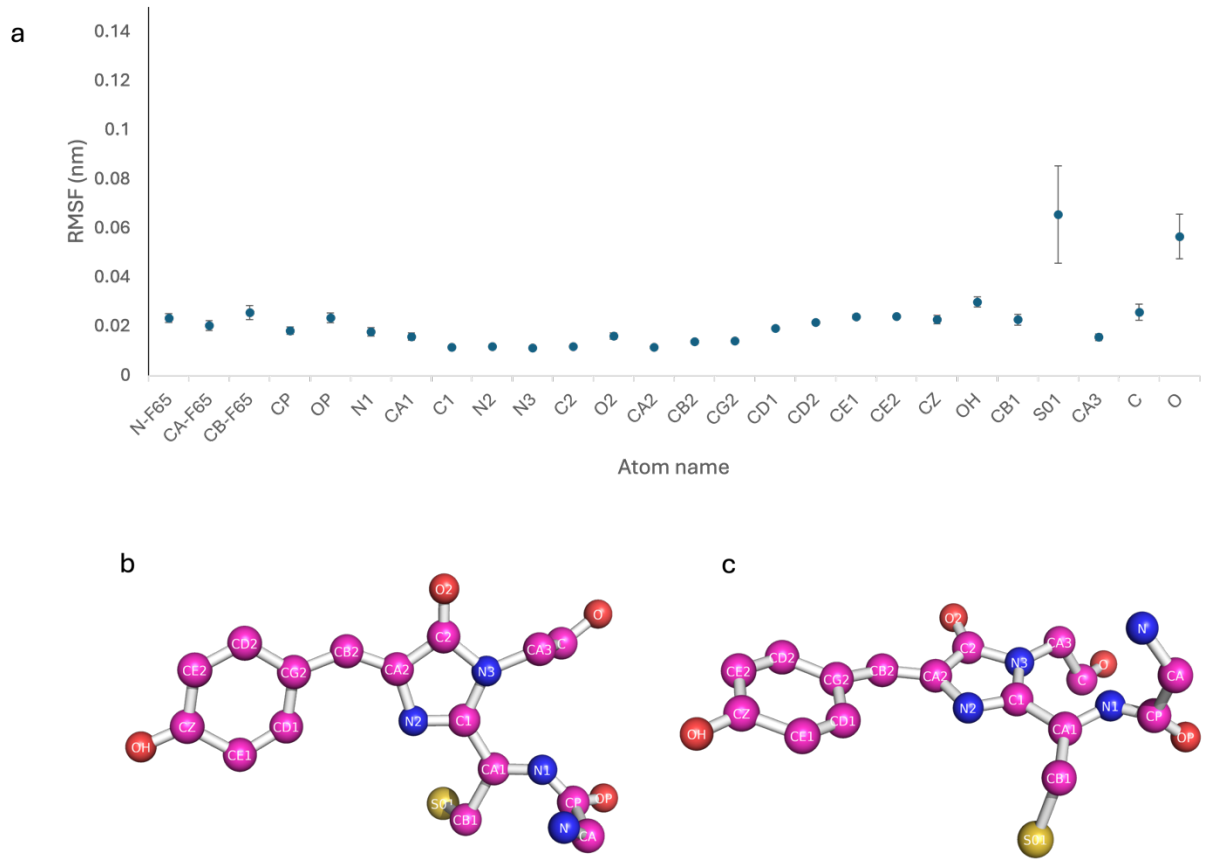

Figure 4.23: RMSF analysis of the mCherry^{M66C} chromophore. a) Average RMSF values with standard deviation for each chromophore atom. b) Top and c) side view of the chromophore structure with labelled atoms. Shown is the average per atom RMSF over the 3 independent simulations with the error bar representing the standard deviation.

This flexibility within the mCherry^{WT} chromophore environment can potentially explain why the protein is more prone to degradation in the presence of hydrogen peroxide compared to the mutant. We originally hypothesised, that because there is more space due to the shorter thiol group of the cysteine and therefore it would be more accessible and more reactive; however, we can now speculate that the reason oxidation does not affect mCherry^{M66C} the same way as mCherry is because oxidative modifications of the sulphur atom can be coped with better in the mutant because there is space for the extra bulk, while that would potentially cause steric clashes in mCherry due to less space.

MD analysis of hydrogen bonds and water movement across the two proteins, allowed the investigation of how the two chromophores interact with them and how they

are potentially giving rise to the new properties observed in the mutant mCherry characterised so far in this chapter.

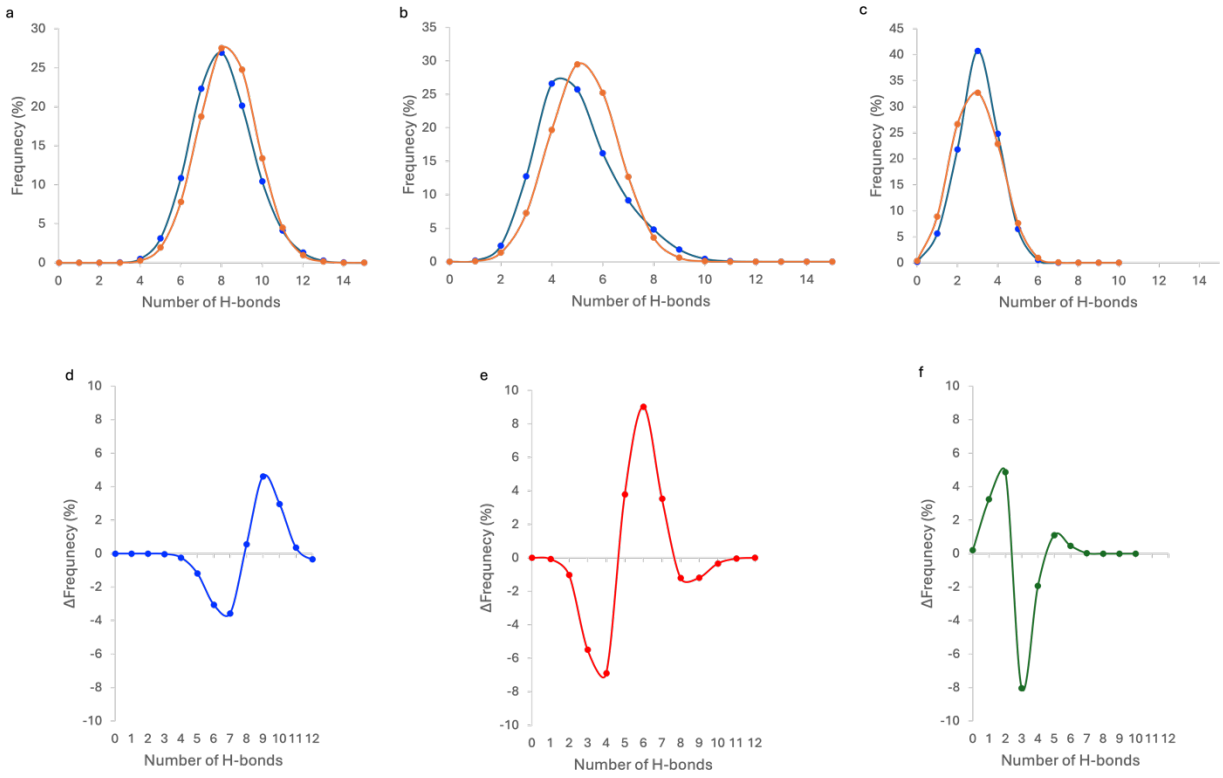

Figure 4.24: Hydrogen bond analysis. Percentage frequency of a) all hydrogen bonds, hydrogen bonds between chromophore b) to water molecules and c) the rest of the protein. Percentage frequency of mCherry^{WT} is represented in blue and of mCherry^{M66C} in orange. Percentage difference in hydrogen bond frequency between mCherry and mCherry^{M66C} for d) all hydrogens, e) hydrogen bonds between chromophore and water and f) hydrogen bonds between chromophore and protein. For d), e) and f) positive values correspond to higher H-bond frequency in mCherry^{M66C} and negative values correspond to a higher H-bond frequency in mCherry.

Analysing the difference in H-bond frequency between the proteins has revealed that the mCherry^{M66C} chromophore forms a large number of hydrogen bonds at a higher frequency to that of mCherry wt (Figure 4.24). A similar trend is observed with the amount and frequency of H-bonds between the mCherry^{M66C} chromophore and water molecules (Figure 4.24 b and c). This analysis shows that, mCherry^{M66C} makes a lot more contact with water molecules rather than with the protein itself (Figure 4.24 b and e), unlike mCherry^{WT} where the most H-bonds are between the chromophore and the rest of the protein (Figure 4.24 c and f). Overall, a higher H-bond frequency was observed in mCherry^{M66C} compared

to mCherry. Upon analysing the accessibility of the chromophore to water, the mCherry^{M66C} chromophore appeared to experience an increase in solvation, and particularly close to the M66C mutation site (Figure 4.25).

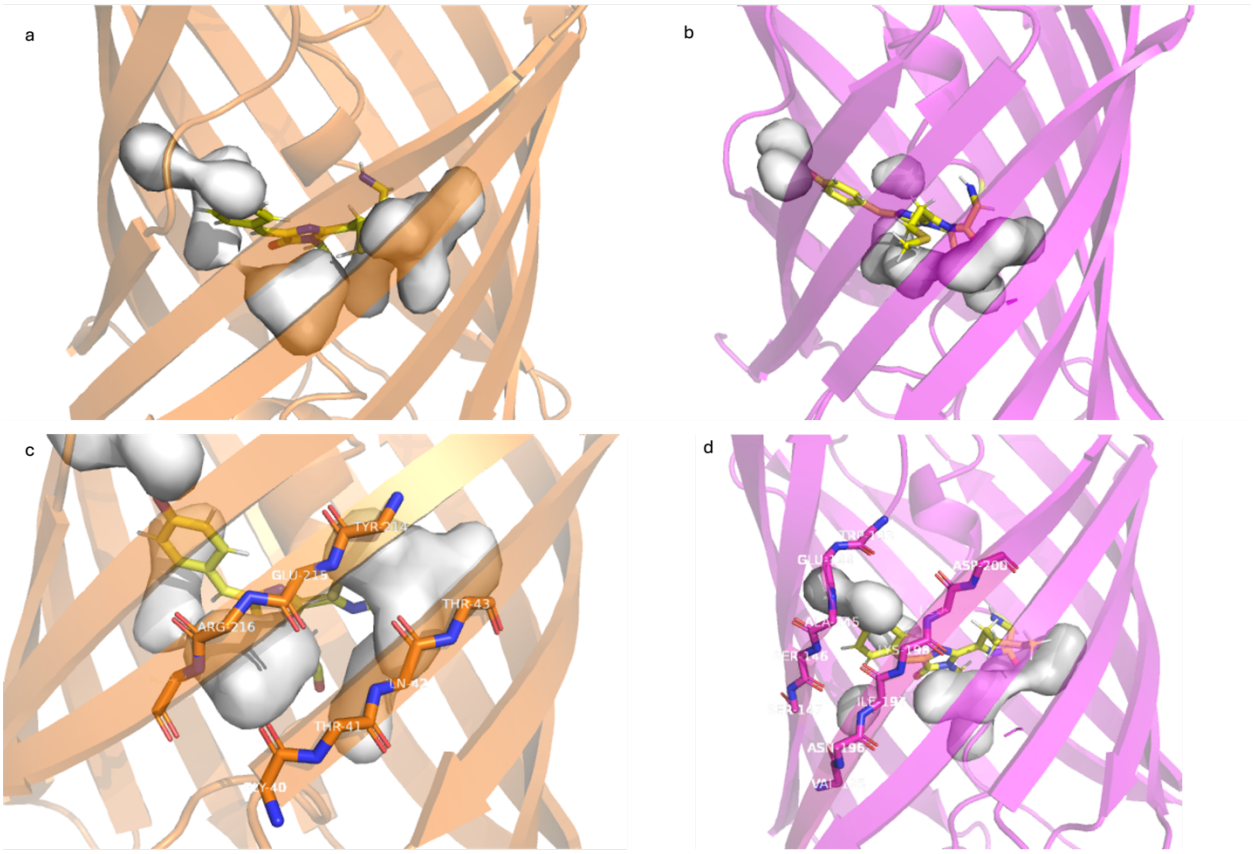

Figure 4.25: Water accessibility to chromophore. Cartoon and stick representation of the mCherry proteins where mCherry^{M66C} is presented in orange and mCherry^{WT} in magenta. a) and b) show the water molecules around the chromophores and c) and d) show the residues involved in the different entry/exit points.

This confirmed the hypothesis that the empty space is occupied by a molecule and that it is in fact filled with water. This is an important finding as the presence of this extra water interacting with the chromophore is possibly the reason for the novel properties of the protein observed. Analysis of the movement of the internalised water molecules during MD simulations revealed that on average water molecules, leave the mCherry^{M66C} chromophore environment quicker than in mCherry. Analysis of the exit points of water molecules also showed that water primarily exits at two regions, between the strands

comprised by residues 144-199 and 43-214 (Figure 4.25 c and d). More specifically, water exits mCherry^{M66C} between 43-214 which they are found opposite the newly introduced thiol group, while for mCherry the exit point is between the next two strands where a clear break point in the strand-to-strand interaction can be seen (Figure 4.25d). The RMSF value of the strand break point residues (143-146) is also higher in mCherry than mCherry^{M66C} but only by 0.007-0.02 nm. The difference in the distance between the exit point preferred by each protein was compared to see if there is a significant increase in the distance for each exit point to identify if the reason each exit point is preferred is due to size. In both cases the difference in distance was very small (Table 4.8), but the exit point region preferred by mCherry is marginally closer together in distance in mCherry^{M66C} than mCherry.

Table 4.8: Distance between water exit points.

mCherry					
Residues	43-214	40-217	S146-I197	A145-K198	E144-L199
Average Cα to Cα Distance (Å)	4.4	5.2	7.6	7.8	10.6
mCherry^{M66C}					
Residues	43-214	40-217	S146-I197	A145-K198	E144-L199
Average Cα to Cα Distance (Å)	4.5	5.3	7.5	7.5	10.5
Difference (mCherry^{M66C} - mCherry)					
Residues	43-214	40-217	S146-I197	A145-K198	E144-L199
Average Cα to Cα Distance (Å)	0.1	0.1	-0.1	-0.3	-0.1

Based on the analysis of the structural data acquired along with the MD simulations compiled, we could see that the reduction in the length of the side chain has potentially removed steric bulk in mCherry^{M66C}. In turn, this opened up space within the chromophore

environment allowing more water molecules to occupy that space. Because of that, water molecules entering mCherry^{M66C} chromophore environment are more likely to stay close and then leave through the nearest possible exit points which would be between the strands comprised by residues 40-43 and 214-217. One can therefore argue that a good exit point can also be a good entry point for other solutes. Therefore, the presence of these water molecules during the changes in pH, can potentially be why the three-state model observed in mCherry^{M66C} exists with the two intermediate chromophore species where we get alternate protonation between the phenol and the thiol group. On the other hand, in mCherry the extended side chain of the methionine occupies that empty space present in mCherry^{M66C} creating a steric bulk. As a result, this limits the exit of water molecules as well as the entry through residues 40-43 and 214-217 (main exit points in mCherry^{M66C}) as it can act as a molecular plug (Figure 4.26). This is why the main exit and thus entry point in mCherry is through the sheet break point formed via residues 144-146 and 197-199. This region is the one which is closer to the phenol group of the chromophore and thus provides direct access to it. As it is the most accessible way to the chromophore of mCherry it could be argued that this is the reason there is only one transition state in response to pH change as revealed from the QM analysis compared to mCherry^{M66C} where it can be accessed from both points.

Tunnel analysis performed using CAVER showed the ends of the tunnels are close to the potential entry/exit points for molecules for both mCherry and mCherry^{M66C} as revealed from the MD analysis. For mCherry the tunnel ends are found through the sheet break formed via residues 143-147 and 195-200 (Figure 4.26b). On the other hand, for mCherry^{M66C} we can see a more extended tunnel through the β-barrel structure with one end found near the entry/exit point comprised by residues 40-43 and 214-217 (Figure 4.26a). Figures 4.26c and d show the tunnels in mCherry^{M66C} and mCherry along with the water molecules that surround the chromophore (within 4 Å) respectively. When the waters are present, we can see that the environment around the chromophore is quite crowded in mCherry (Figure 4.26d) and aligns with the hypothesis that the presence of the extended side chain of methionine limits movement of water and thus acts as a molecular plug.

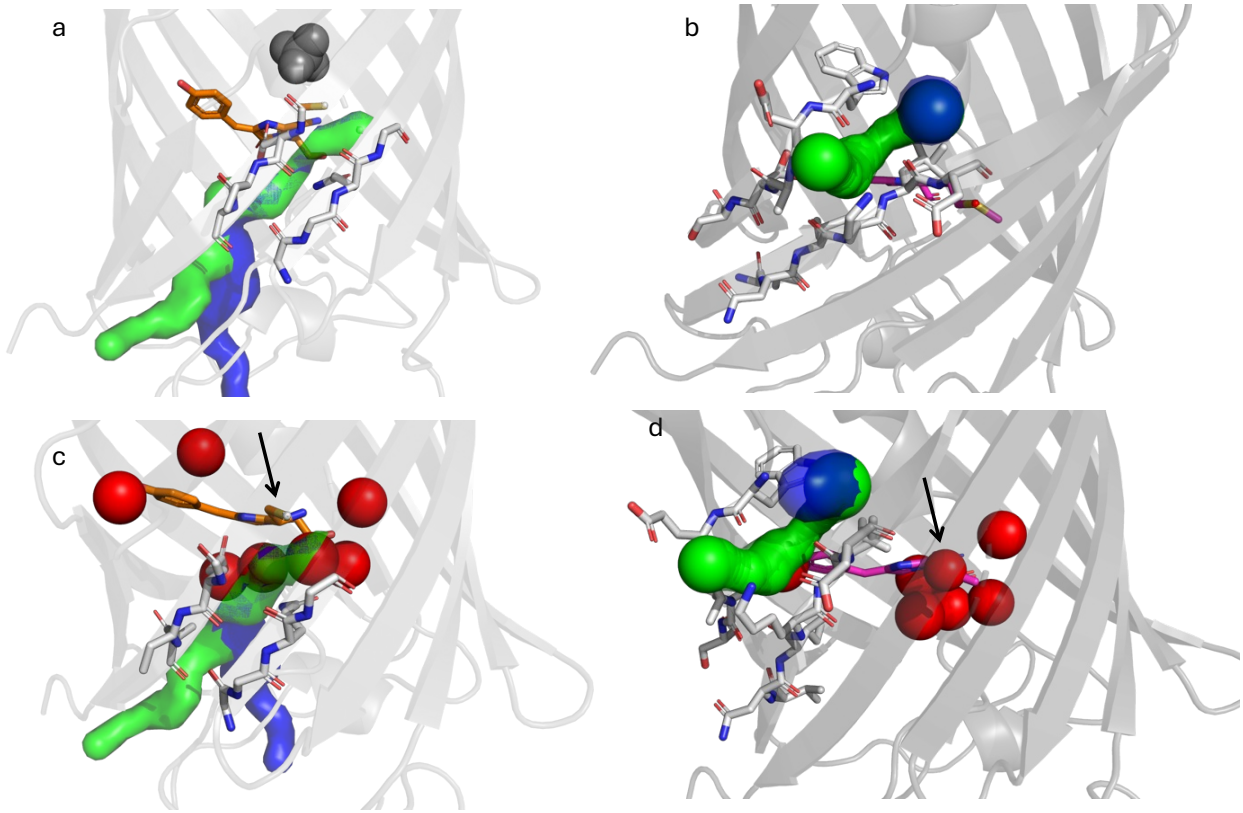

Figure 4.26: Tunnels in mCherry and mCherry^{M66C}. Tunnels are represented by green and blue spheres; waters are shown as red spheres and residues that make up the entry/exit points are presented as white sticks. a) and c) show the tunnels in mCherry^{M66C} while b) and d) show the tunnels in mCherry. mCherry^{M66C} chromophore (orange) and mCherry chromophore (pink). The black arrow indicates the side chain of cysteine and methionine.

Finally, although these analyses reveal a lot for the nature of the chromophore and allowed further understanding of the structure to function relationship of the mutant, the effects of hydrogen peroxide are still yet to be fully understood. It can be argued that oxidation of mCherry^{M66C} is more likely to be tolerated even with more water accessing the pocket and the thiol group can still be accessible to external solutes, while oxidation or other chemical modifications of mCherry might be less likely to be tolerated structurally due to steric clashes between the chromophore and the β-barrel (i.e. residue Q42).

4.3 Conclusions

In conclusion, we can see that this newly generated protein has unique properties compared to the original mCherry. Structural analysis together with functional studies and

molecular dynamics have allowed us to further understand the potential mechanisms that give rise to the observed spectral properties of the newly generated mutant.

Residue 66 plays a key role in determining the chromophore's immediate chemical environment and is thus a critical determinant of the protein's photophysical behaviour. Although primary changes in chromophore composition are a major driver of spectral properties, mounting evidence suggests that local environmental effects, particularly electrostatic interactions within the chromophore pocket, are equally influential (Shu et al. 2006; Drobizhev et al. 2009).

For example, mGrape1, mCherry and mHoneydew all have a M66, and all have different set of mutations compared to mRFP1.1 where they descended from. To generate mHoneydew the following mutations were introduced Y67W, E144A, S146T, S147T, T195I and I197G. For mGrape1 M150L, V177A, T195L and I197Y were introduced. From mRFP1.1 to mCherry, three intermediate proteins were engineered. In the process mRFP1.1 had undergone amino acid replacement both at its N and C-terminal, along with A225G (mRFP1.3), V7I and M182K (mRFP1.4), R17H and M163Q (mRFP1.5) and finally N6D, K194N, T195V, D196N (mCherry).

Some key mutations include V7I and M182K as they showed to be beneficial for protein folding while M163Q has showed an almost complete disappearance of the secondary absorbance peak at ~510 nm (Shaner et al. 2004). All the variety within these proteins is both reflected on their absorbance and emission peaks as well as the colour of the protein. While mGrape1 and mCherry still have RFP-like properties, mHoneydew is blue-shifted enough to exhibit GFP-like attributes with absorbance and emission peaks of 487/504 and 537/562 respectively as well as a yellowish colour.

On the other hand, mGrape1 and mCherry have absorbance and emission peaks at 595/625 nm and 587/610 nm respectively as well as purple-pink hue (Shaner et al. 2004; Lin et al. 2009). We can also see that key residues engineered for these variants such as 195-197 or 144-147 and 163, are residues that are involved in water transport within the protein (Section 4.2.4.1) and hydrogen bonding (Chica et al. 2010) respectively.

Therefore, having structural information to be able to relate and attribute these changes to specific interactions or residues can be a very strong informant for future protein engineering rationales for improving fluorescent protein properties.

5 Characterisation of superfolder GFP^{H148S}

Work in this chapter contributed to a published, peer-review paper where I am joint 1st author.

Ahmed, R. D., Jamieson, W. D., Vitsupakorn, D., **Zitti, A.,** Pawson, K. A., Castell, O. K., Watson, P. D., & Jones, D. D*. (2025). Molecular dynamics guided identification of a brighter variant of superfolder Green Fluorescent Protein with increased photobleaching resistance. Communications Chemistry, 8(1), 174. https://doi.org/10.1038/s42004-025-01573-4

* Corresponding author.
R.D.A., W.D.J, D.V. and A.Z contributed equally to this work. The order is based on the alphabetical order of their surnames.

Statement of work

Work in this chapter was performed in a collaborative effort by members in Prof Jones Lab, Dr Castell's Lab and Prof Watson. Initial design, modelling and engineering of the YuzuFP and steady-state spectral analysis was carried out by Dr Rochelle Ahmed. I produced the protein for downstream steady-state spectral analysis, single molecule fluorescence and generated the constructs for the mammalian cell imaging work. Single molecule fluorescence data collection and analysis were carried out by Dr David Jamieson. Danoo Vitsupakorn contributed to mutant production, steady-state spectral analysis and long scale molecular dynamics analysis. Kai Pawson contributed to the initial short scale molecular dynamics. Dr Oliver Castell contributed to single molecule measurements and analysis and directed the project. Mammalian cell imaging data were collected and

analysed by Prof Peter Watson. Long scale molecular dynamic analysis was conducted by Prof Dafydd Jones and Danoo Vitsupakorn.

5.1 Introduction

Since their discovery, FPs have played a pivotal role in our understanding of biochemical and molecular processes as well as characterisation of cellular structures. Over the years, protein engineering has been extensively used in an attempt to generate an array of fluorescent proteins to meet research requirements. Being the first FP to be discovered, green fluorescent protein (GFP) has always been a focus point when it comes to engineering new variants to match the ever-changing need for improved probes adapting to the needs of the discipline. Scientists always look to improve resistance to photobleaching, brightness, longer lasting fluorescence signals as well as the lifetime of fluorescence (Enterina et al. 2015; Duwé and Dedecker 2019).

GFP can exist in two different forms, the neutral protonated form of the chromophore (CroOH) or the anionic deprotonated form of the chromophore (CroO-). In its protonated form GFP absorbs at ~400 nm, while in its deprotonated form it absorbs at ~490 nm. However, in both cases a green fluorescence is emitted (510 nm). The reason for this is that the chromophore undergoes a photoexcitation shift referred to as excited-state proton transfer (ESPT). When CroOH is exposed to UV, it gets deprotonated and thus converts to the CroO- form hence emitting fluorescence (Remington 2011). Since exciting at 490 nm instead of 400 nm requires less energy, efforts were made to generate a protein where there was a shift towards the 490 nm excitation wavelength. This was achieved when GFP was engineered to form EGFP, where mutating F65T promoted the ground state $CroO^-$, in addition to making the protein mature faster and more stable. The mutation at residue 65 showed to be critical for the promotion of the $CroO^-$ state, through changing the nature of the charged state of E222 and in extend the network linked through to the phenol group of the chromophore (Arpino et al. 2012). Through studying the structure of EGFP, it was apparent that aside from E222, neighbouring residues such as H148 and water molecules

play a role in proton transfer regulation and in maintaining the anionic ground state of the chromophore (Zhang et al. 2005).

The histidine at residue 148, is one of the key determinants of the fluorescence properties of avGFP. Along with a structurally conserved water molecule, it interacts directly with the phenolate group of the chromophore and is considered to play a key role in promoting and stabilising the deprotonated phenolate form in the ground state (Seifert et al. 2003; Campanini et al. 2013; Hartley et al. 2016). Modifying the histidine to different amino acids has shown to give rise to reduced emission as well as expression yields and stability of the mature protein (Bulina et al. 2002). Mutating histidine to glycine leads to the loss of the H-bond to the chromophore, which subsequently gives rise to dual emission GFP, with a predominant protonated form (Hanson et al. 2002). H148 is also thought to be dynamic and has been observed to sample an "open" conformation which shifts the residue away from the chromophore, as well as the predominant "closed" form. It has been observed that the "open" form of the chromophore leaves enough space to form a channel allowing diffusion of small molecules and even larger metal ions (Barondeau et al. 2002; Seifert et al. 2003).

In addition to replacing the histidine with natural amino acids to showcase the residue's importance for the function of GFP, incorporation of nnAA at 148 of superfolder GFP has also been demonstrated before. In that case, histidine was replaced with azido phenylalanine (azF), which led to the loss of two hydrogen bonds from the chromophore, one from H148 and one from T203. This loss resulted in a blue shift in the major absorbance peak by about ~90 nm shifting the dominant deprotonated peak to the neutral chromophore ground state. However, it has been shown that this shift is reversible upon UV irradiation or through biorthogonal click chemistry (Hartley et al. 2016). This concept was further explored as means to generate a photo switchable biosensor, where a switch-on effect ($CroO^-$ predominates) takes place upon dimerisation of sfGFP. To achieve that, histidine at 148 was replaced with a cysteine, allowing the formation of disulphide bridges between them and thus encouraging dimerisation.

Since modifying residue 148 has proved to influence the properties of GFP, this chapter focuses on how replacing the imidazole side chain of H148 modulates the charged state of the chromophore between the protonated and deprotonated forms. The imidazole of histidine was replaced with the hydroxyl group present in serine, which led to a brighter protein with improved molar absorbance and quantum yield. This change in the chromophore environment has also improved resistance to photobleaching as revealed by single molecule fluorescence analysis.

In this chapter I will explore and analyse the improved properties that sfGFP^{H148S} (YuzuFP) has compared to sfGFP and discuss how MD modelling helped in deciding what residue to mutate H148 into as well as shed some light in the newly acquired dynamics of the chromophore environment. This new construct has showed to be 1.5 times brighter than the original protein variant (sfGFP), with a near 3-fold increase in resistance to photobleaching.

5.2 Results and Discussion

5.2.1 Engineering of H148S rationale

The contribution of residue H148 to the stabilization of the anionic chromophore form (CroO$^-$) in avGFP-derived fluorescent proteins is well documented in literature (Seifert et al. 2003; Campanini et al. 2013). However, H148 is not regarded as essential for proper folding or chromophore maturation in sGFP, as variants at this position can still yield functional, fluorescent proteins (Reddington et al. 2013; Hartley et al. 2016). Previous studies suggest that H148 exhibits conformational flexibility (Seifert et al. 2003) and can adopt multiple configurations, some of which disrupt its interaction with the chromophore(Arpino et al. 2014; Auhim et al. 2021). The conformation most frequently observed in crystal structures is illustrated in Figure 5.1a; however, this arrangement is suboptimal with respect to both geometry and interaction strength. Specifically, the angle of interaction is approximately 140^O, and the distance between the imidazole side chain and the chromophore ranges from 0.31-0.35 nm (Figure 5.1d).

Short timescale (10 ns) MD simulations further support this observation, revealing that the hydrogen bond between the phenolate oxygen of the chromophore and H148 is maintained for less than 5% of the simulation time (Table 5.1). In addition, the nearby structurally conserved water molecule, W1 (Figure 5.1b), which is hypothesized to participate in local stabilisation, shows limited residence time in this region (Table 5.2).

Efforts to predict the structural consequences of H148 mutations using tools like AlphaFold and Rosetta have previously been employed but are limited in the context of fluorescent proteins. This is due to their inability to incorporate the chromophore nor the local solvation environment, both crucial for fluorescent proteins. Moreover, these prediction methods provide static structures and do not account for conformational relaxation or the dynamic behaviour of residues. To address these limitations, short timescale (10 ns) MD simulations were performed to evaluate the structural effects of substituting H148 with each of the 19 other canonical amino acids (Figure 5.2) (Ahmed et al. 2025). Among all substitutions, only serine and asparagine adopted side chain orientations that permitted hydrogen bonding with the chromophore phenolate oxygen. In particular, the H148S variant appears to form a more favourable interaction where the hydroxyl group of the serine forms a hydrogen bond with the phenolate oxygen at a shorter O-O distance (~0.27 nm), with near-planar geometry (Figure 5.1b and c). Hydrogen bond analysis reveals that in the H148S model, the interaction between S148 and the chromophore phenolate oxygen is both more frequent and occasionally involves dual hydrogen bonds, one from the serine side chain and another from the backbone amine (Table 5.1).

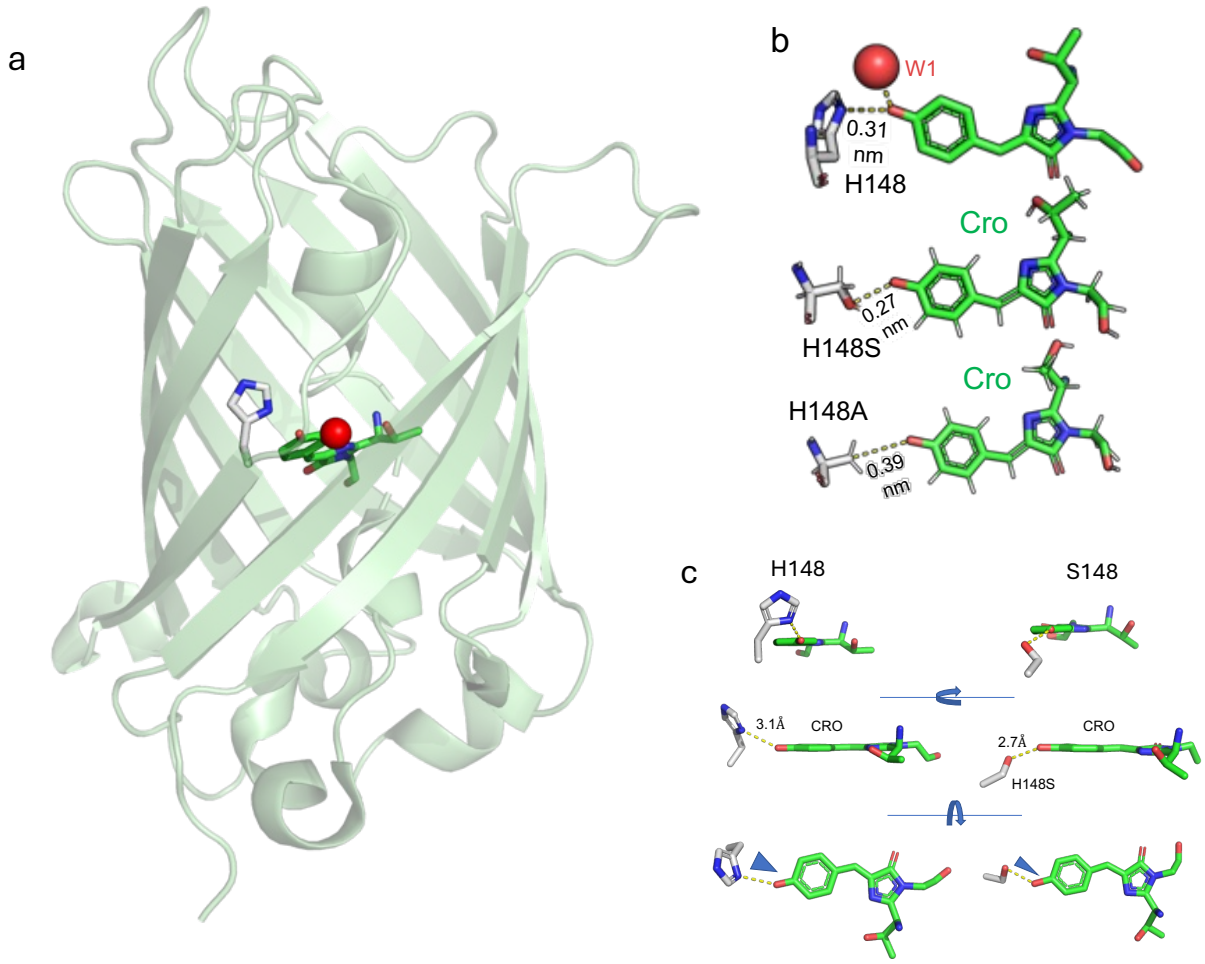

Figure 5.1: Structural representation of the interaction between the chromophore and residue 148. a) Crystal structure of sfGFP (PDB 2B3P) with the position of the chromophore (green sticks) relative to H148 (grey sticks). b) shows the interaction between the chromophore and H148, H148S and H148A respectively. c) Residue conformations of 148 (grey sticks) and chromophore (green sticks). On the left side is the position of H148 and on the right is the modelled structure of H148S from the clustered average of individual MD trajectory outputs from the 10 ns MD simulations. The dashed yellow lines represent the distances between the phenolate oxygen of the chromophore and the imidazole and hydroxyl group of the histidine and serine respectively. Red sphere represents a water molecule. Figure adapted from Ahmed et al. (2025).

In the case of the H148N variant, molecular dynamics simulations revealed increased distances between the side chain and the chromophore phenolate oxygen, along with altered side chain orientations compared to the native histidine (Figure 5.2). Initial models of H148N prior to MD (with no hydrogens present) suggested that either the amine or carbonyl can form the polar interaction depending on carboxamide rotation. However, following 10 ns of MD sampling with the chromophore in its phenolate state, only the amine

group was observed to form a hydrogen bond with the chromophore (Figure 5.2; Table 5.1). In contrast, the H148T substitution does not support stable hydrogen bonding between the side chain hydroxyl and the chromophore. Substitution with alanine preserves the β-carbon orientation towards the chromophore but eliminates any side chain mediated hydrogen boning capacity (Figure 5.1b; Table 5.1). For the H148C variant, the cysteine side chain tends to flip out, likely due to the large atomic radius of sulphur and the comparatively weaker Lewis base character of the thiol group relative to a hydroxyl (Figure 5.2; Table 5.1). Other amino acid substitutions at position 148 were similarly found to be incapable of forming hydrogen bonds with the chromophore phenolate (Figure 5.2).

Table 5.1: H-bond frequency between chromophore and residue 148 over 10ns simulations. Copied with permission from Ahmed et al. (2025)**.**

Variant	H-bonds = 0	H-bonds = 1	H-bonds = 2	% time H-bonding
sfGFP WT (H148)	952	49	0	4.9
sfGFP H148S	512	409	80	48.9
sfGFP H148T	993	8	0	0.8
sfGFP H148N [a]	222	779	0	77.9
sfGFP H148N [b]	992	9	0	0.9
sfGFP H148C	977	24	0	2.4
sfGFP H148A	1001	0	0	0

a, sfGFP H148N where starting rotamermic form has the NH group of the carboxamide group is closest to the chromophore phenol O.
b, sfGFP H148N where starting rotamermic form has the O group of the carboxamide group is closest to the chromophore phenol O.

In addition to direct residue interactions, we evaluated the behaviour of the conserved structural water molecule (W1) (Figure 5.1b), over the 10 ns simulation window. In general, W1 was observed to diffuse away from its crystallographic position within approximately 2.3 ns (Table 5.2), although replacement by other water molecules was occasionally noted. In the wild type sfGFP simulation, W1 formed hydrogen bonds with the chromophore phenolate for less than 6% of the simulation time (Table 5.2). Interestingly, increased hydrogen bond occupancy with the chromophore was observed for W1 in the H148N (19.6%), H148A (22.6%) and H148T (13.1%) mutants. In contrast, both forms of H148N, as well as H148C, exhibited W1-chromophore hydrogen bond frequencies comparable to the wild type sfGFP (Table 5.2).

Taken together, these short timescale MD simulations suggest that among the tested variants, H148S exhibits the greatest propensity to support stable local hydrogen bonding interactions with the chromophore phenolate group.

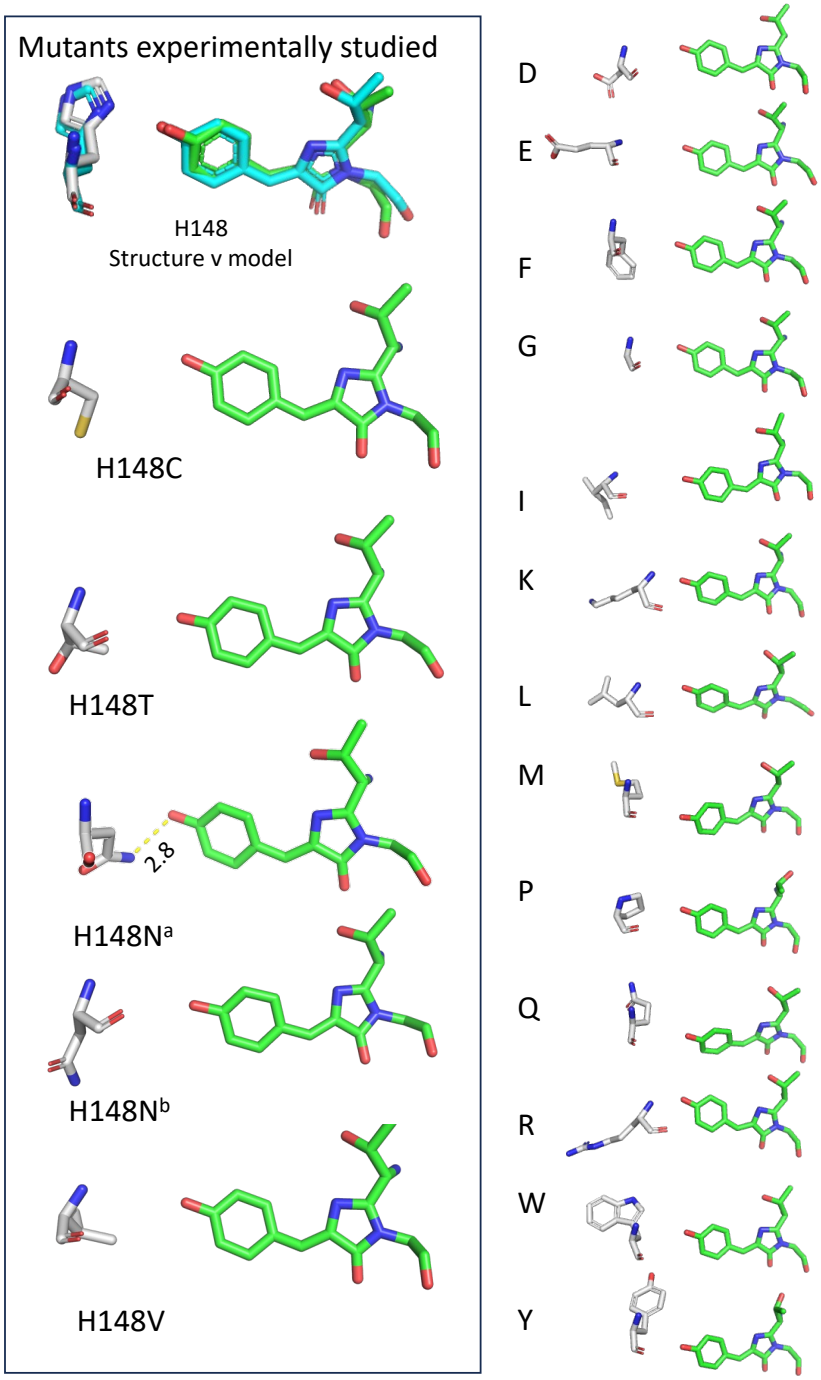

Figure 5.2: Modelling of the H148X mutations. Modelled structures of the H148X mutations are the clustered average of individual MD trajectory outputs after 10 ns of molecular dynamics. Mutations on the left-hand side have been experimentally analysed. For H148N, two model outcomes are shown: H148N[a] is from a starting point of the NH group of the carboxamide group closest to the phenol oxygen of the chromophore. H148N[b] is from a starting of the oxygen group of the carboxamide group closest to the chromophore phenol oxygen (Ahmed et al. 2025).

Table 5.2. H-bond frequency between chromophore and W1 water over 10ns simulations. Copied with permission from Ahmed et al. (2025)**.**

Variant	H-bonds = 0	H-bonds = 1	H-bonds = 2	% time H-bonding	Time before distance > 0.4 nm[c]
sfGFP WT (H148)	945	56	0	5.6	0.22 ns
sfGFP H148S	805	196	0	19.6	2.24 ns
sfGFP H148T	870	131	0	13.1	1.4 ns
sfGFP H148N [a]	946	55	0	5.5	0.83 ns
sfGFP H148N [b]	931	70	0	7	0.1 ns
sfGFP H148C	973	28	0	2.8	0.53 ns
sfGFP H148A	775	224	2	22.6	1.05 ns

a, sfGFP H148N where starting rotamermic form has the NH group of the carboxamide group is closest to the chromophore phenol O.
b, sfGFP H148N where starting rotamermic form has the O group of the carboxamide group is closest to the chromophore phenol O.
c, refers to distance between the chromophore phenol O atom and the O atom of W1. Time is based 0.4 nm distance threshold broken for at least 0.2 ns.

5.2.2 The effect of H148S mutation on the spectral properties of sfGFP

Serine was introduced into residue 148 via site-directed mutagenesis. Once pure protein was produced, spectral analysis was conducted to characterise the protein (Ahmed R. 2022). Absorbance and fluorescence measurements showed a 1.3 – fold and 1.5 – fold increase respectively at λ_{max} and λ_{EM} for both, for YuzuFP compared to sfGFP (Figure 5.3 and Table 5.3). In addition, YuzuFP has slightly red-shifted (4-5 nm) excitation and emission compared to sfGFP, which gives it its slight yellow hue compared to sfGFP, hence the name YuzuFP.

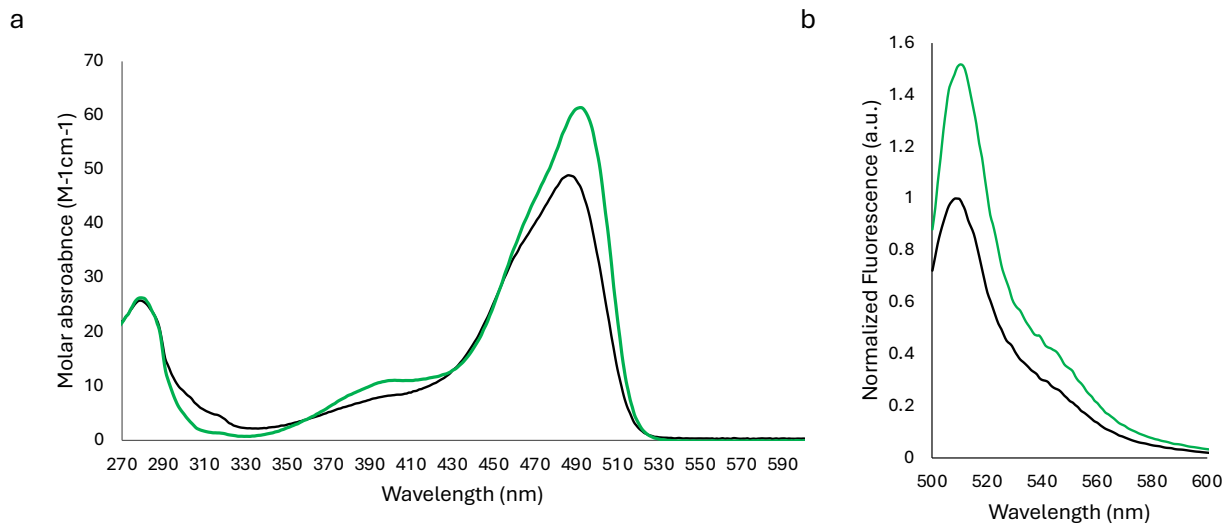

Figure 5.3: Spectral analysis of sfGFP vs YuzuFP. a) Absorbance spectra of sfGFP (black) against YuzuFP (green). b) Emission spectra of sfGFP (black) against YuzuFP (green) upon excitation at their respective λ_{max} of 485 nm for sfGFP and 492 nm for YuzuFP. Emission spectra were normalised to the sfGFP emission maximum.

Absorbance and fluorescence were recorded for both sfGFP and YuzuFP at a range of pH states (4.5-12). At pH states lower than pH 8, a decrease in both absorbance and emission is observed. As conditions become more acidic, a mixture of the protonated and deprotonated forms of the chromophore is present as a shift from the ~490 nm peak to the ~400 nm peak is observed. An overall decrease in both absorbance and emission was recorded for sfGFP when pH moved away from neutral, whereas a slight increase in absorbance was recorded for YuzuFP at higher pH states (10-12). At pH 7 absorbance of sfGFP decreases more than that of YuzuFP with a ~24 % and ~15 % decrease in absorbance respectively. This decrease in absorbance is also reflected in emission (on excitation at 490 nm) at pH 7 with a decrease of 24 % and 8 % for sfGFP and YuzuFP respectively. Moving on, at higher pH states of 11 and 12, absorbance of YuzuFP shows an increase, however, emission decreases in alkaline conditions. On the other hand, sfGFP shows a steady decrease in both absorbance and emission when pH becomes more acidic or alkaline. A critical information here is that the absorbance spectra show a clear isosbestic point at pH6 where the ratio of the ~400 nm to 490 nm is in equilibrium (Figure

5.4a and b), suggesting a defined switch between the two states (protonated and deprotonated chromophore).

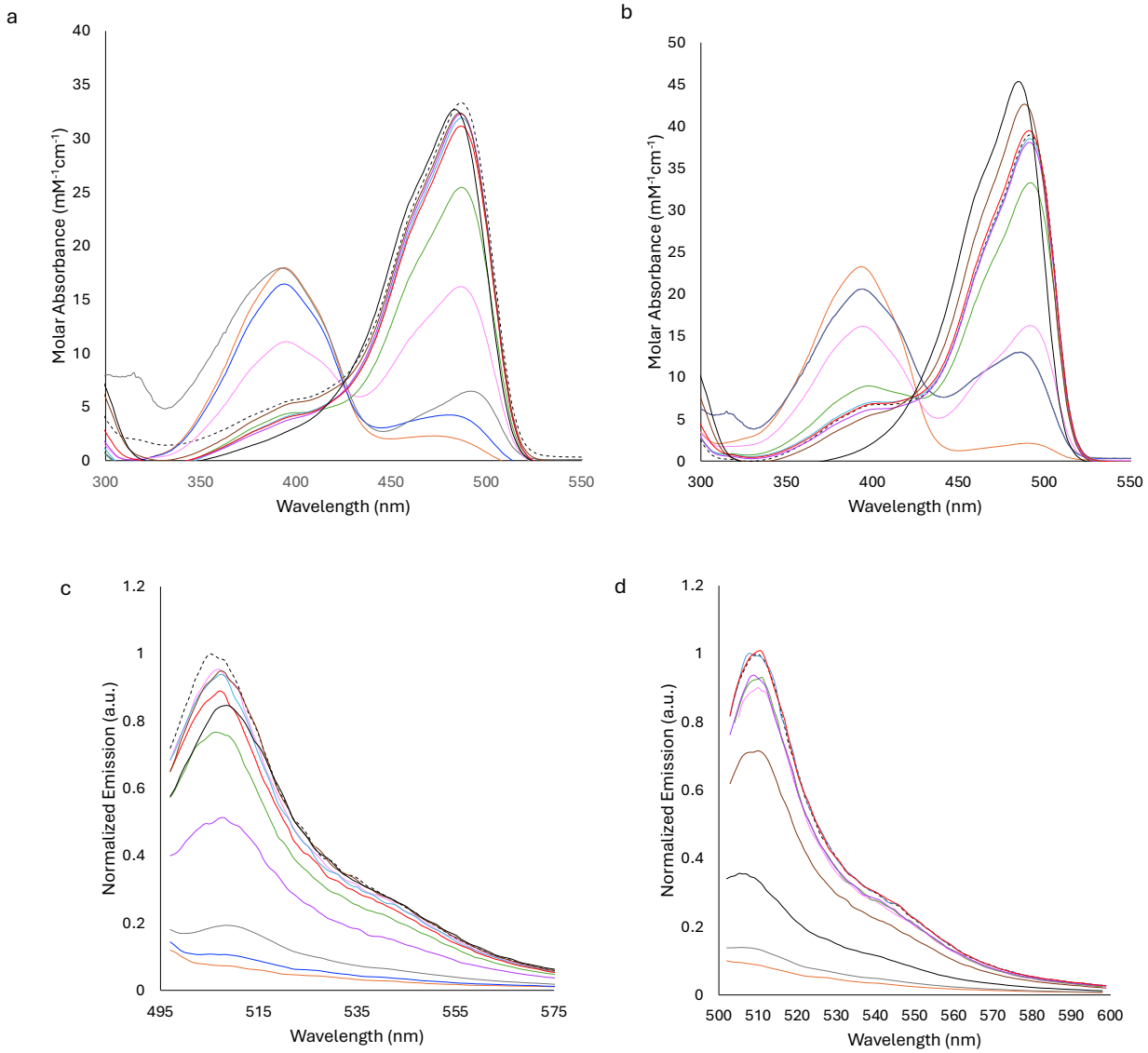

Figure 5.4: Spectral Analysis of sfGFP and YuzuFP at different pH states. Absorbance spectra of a) sfGFP and b) YuzuFP and emission spectra of c) sfGFP and d) YuzuFP at different pH states. pH8 (black dashed), pH4.5 (orange), pH5 (blue), pH5.5 (grey), pH6 (pink), pH7 (green), pH8 (teal), pH9 (purple), pH10 (red), pH11 (brown) and pH12 (black). Emission spectra were normalised to peak emission of protein at pH 8 for each variant.

Following analysis of the spectral properties of YuzuFP and its sensitivity to pH changes, its pKa was determined by plotting the change in absorbance at 490 nm against the pH state (Figure 5.5). Despite the slight variability between the spectral properties of the two proteins, their pKa values are relatively similar. YuzuFP has a pKa of 6.2 whereas sfGFP has a pKa of 6.1 (Table 5.3). This indicates that residue 148 does not play a major role in dictating pH-dependence in fluorescence or significantly affect the pKa of the chromophore.

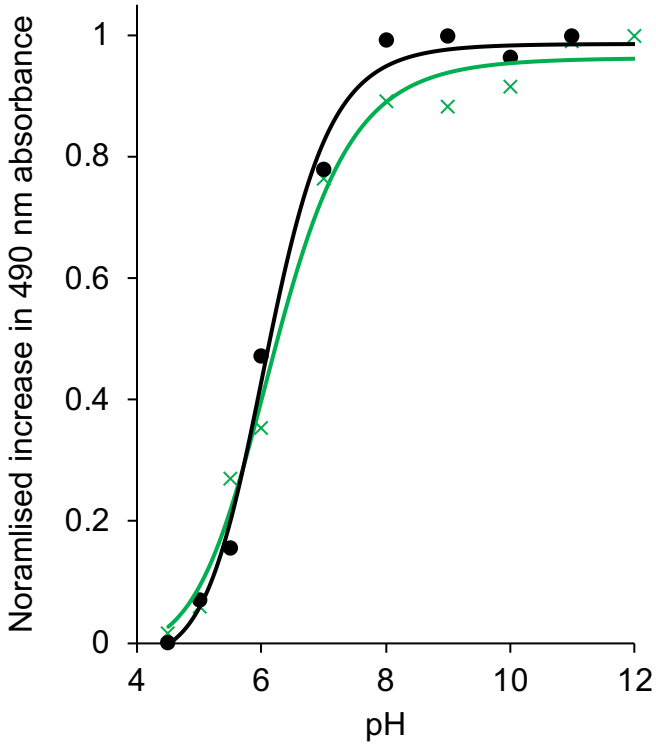

Figure 5.5: pH-dependent increase in absorbance at 490 nm of sfGFP and YuzuFP. Black dots represent the absorbance values at different pH for sfGFP and green Xs for YuzuFP. Black sigmoidal curve corresponds to sfGFP and green sigmoidal curve to YuzuFP.

Table 5.3: Spectral properties of sfGFP variants and other fluorescent proteins.

FP	λ_{max} (nm)	ε ($mM^{-1}cm^{-1}$)	λ_{EM} (nm)	QY	Brightness ($mM^{-1}cm^{-1}$)	pKa
YuzuFP (sfGFP^{H148S})	492	61.4	511	0.92	56.5	6.2
sfGFP	485	49.0	509	0.72	35.3	6.1
sfGFP^{H148A}	398 / 497	36.5 / 15.2	511	0.35 / 0.54	12.8 / 8.2	ND
EGFP	488	56	508	0.67	37.5	6.1

Since spectral analysis revealed that YuzuFP is brighter *in vitro* compared to both sfGFP and EGFP (1.7 times brighter), further analysis through different microscopy methods was undertaken to further explore and quantify its photostability compared to sfGFP.

5.2.3 Photostability of YuzuFP

High photostability and strong fluorescence intensity are essential properties for fluorescent proteins used in super resolution imaging techniques. These attributes enable effective imaging under intense illumination conditions, such as those employed in simulated emission depletion (STED) and total internal reflection fluorescence (TIRF) microscopy. Furthermore, enhanced photostability contributes to improved signal to background contrast, particularly in contexts where photobleaching limits the ability to distinguish fluorescent signal from background noise (Minoshima and Kikuchi 2017).

Photostability of fluorescent proteins Both sfGFP and YuzuFP were analysed at the single molecule level *in vitro* as well as in live cells. Single fluorescent protein analysis of sfGFP and YuzuFP was performed by Dr David Jamieson (Cardiff School of Pharmacy and Pharmaceutical Sciences) and it was done using TIRF. In addition to its enhanced brightness, YuzuFP displayed greater photostability at the single-molecule level compared to sfGFP. When comparing individual fluorescence traces, YuzuFP consistently remained fluorescent for longer durations than sfGFP (Figure 5.5a; Figure 5.6b). This indicates that YuzuFP is more resistant to photobleaching (Figure 5.5b), exhibiting a half-life of 15.0 ± 0.1

seconds under the single molecule imaging conditions, which is approximately twice as long as the 8.3 ± 0.1 seconds determined for sfGFP. Correspondingly, the lifetime increased from 12.0 ± 0.2 seconds for sfGFP to 21.5 ± 0.1 seconds for YuzuFP. Analysis of the single-molecule traces revealed that 24 % of YuzuFP molecules remained fluorescent for over 25 seconds (Figure 5.6a; Figure 5.7), even under laser intensities suitable for dynamic single molecule imaging at 60 ms time resolution. Given that sfGFP is already known as one of the most photostable green-yellow FPs, these findings represent a substantial improvement. The mechanism by which the H148 mutation enhances photostability is not yet fully understood. One hypothesis is that improved hydrogen bonding within YuzuFP's chromophore environment could reduce the likelihood of light-induced conformational changes or prevent photochemical alterations. Another possibility is that the mutation limits the access of reactive molecules, such as oxygen, to the chromophore pocket, thereby reducing possible photochemical modifications.

Furthermore, at the single molecule level, fluorescent proteins often exhibit a behaviour known as "blinking", where the fluorescence intermittently switches between bright ("on") and dark ("off") states. The blinking has previously been linked to ensemble-level fluorescence behaviour. YuzuFP demonstrates an extended duration in the "on" state compared to sfGFP, with the average on-time increasing from 3.6 seconds to 5.3 seconds (Figure 5.6c).

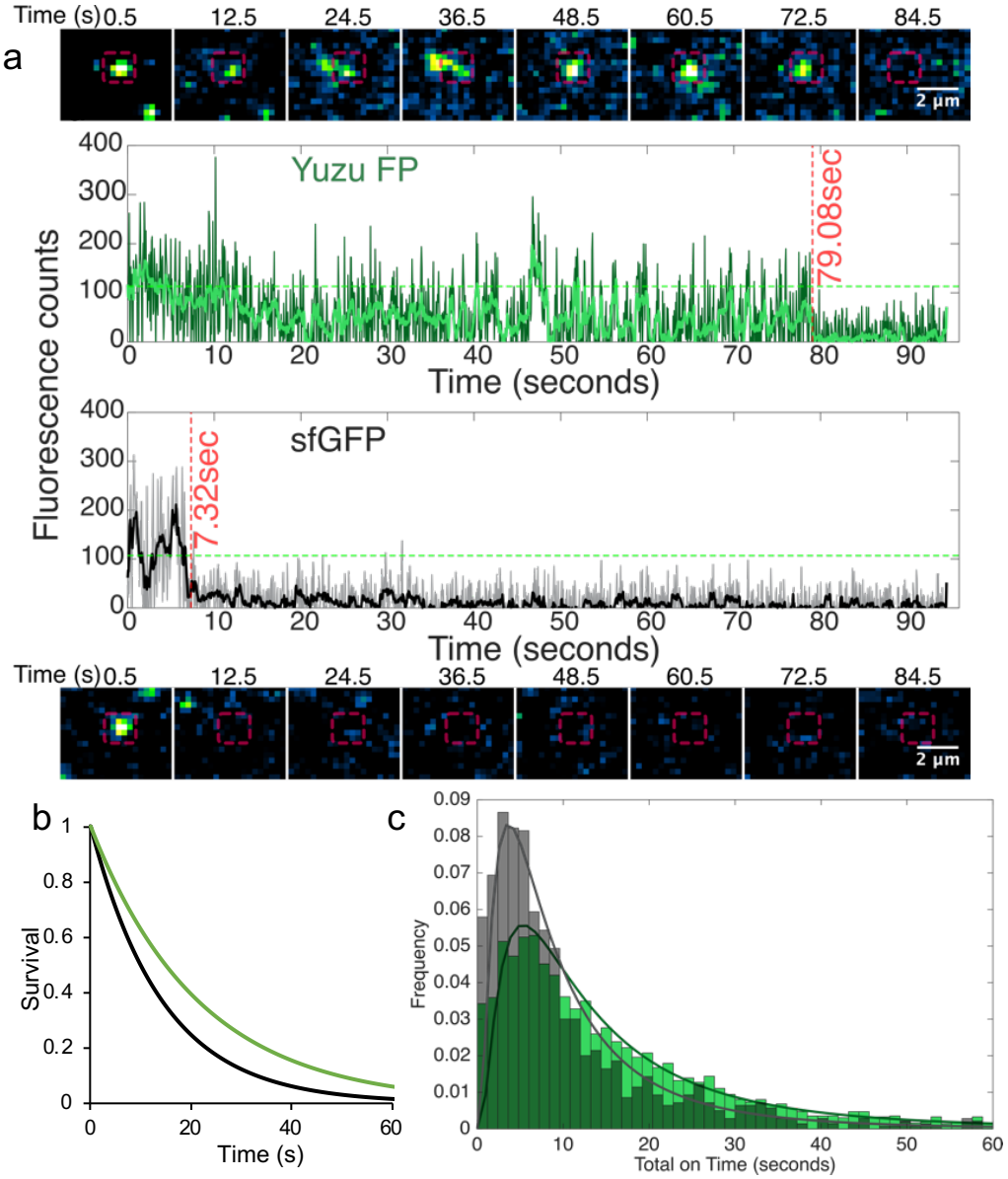

Figure 5.6: Single molecule fluorescence analysis of sfGFP and YuzuFP. a) Single FP fluorescence images at different timepoints, with their corresponding emission traces for YuzuFP (top,green) and sfGFP (bottom, black). The red box in the time course images, represents the region of interest representing the analysed area. The dark green and grey traces show the raw data for each protein while the lighter green and black lines show the denoised data after applying a Chung-Kennedy filter. The green dashed line represents the threshold separating values considered on and off. b) Single molecule photobleaching survival plot sfGFP (black) and YuzuFP (green) plotted proportionally to fluorescence retention (survival) of FPs over time. Decay curves were generated by fitting a single component exponential function to empirical cumulative distribution functions, comprised from the fluorescence lifetimes from 3766 and 1283 individual FP traces for YuzuFP and sfGFP, respectively. c) Frequency distribution of total "on" times for sfGFP (grey) and YuzuFP (green), representing the cumulative time molecules across each population spent in an "on" bright state prior to photobleaching. The solid lines represent a log-normal fit to the distribution. Figure used with permission from Ahmed et al. (2025).

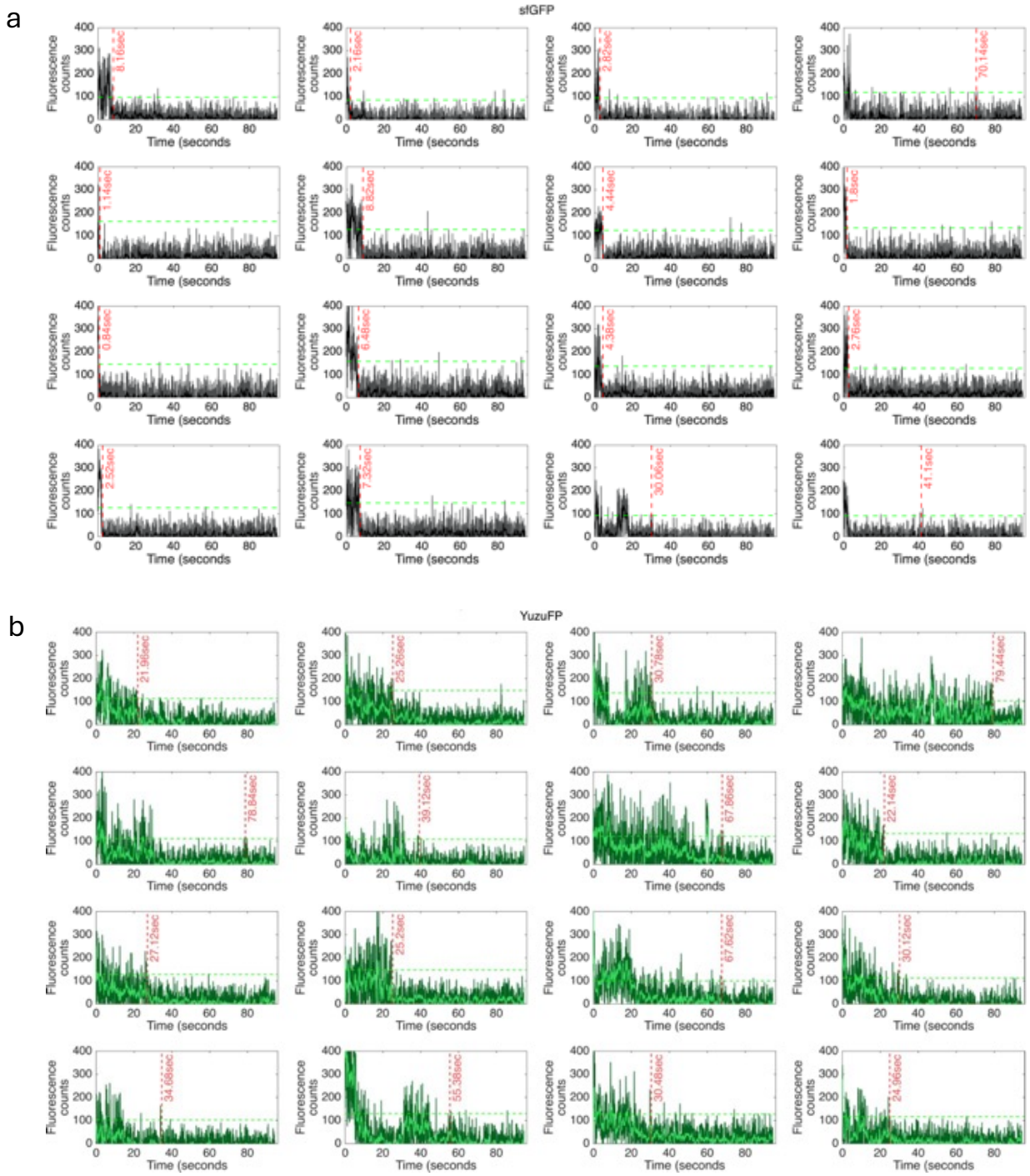

Figure 5.7: Single molecules fluorescence traces for (a) sfGFP and (b) YuzuFP . Single molecule data is extracted from a micrograph timeseries acquired using a total internal reflection fluorescence (TIRF) microscopy imaging system. Traces are generated from the mean intensities of 4 by 4-pixel regions of interest corresponding to individual fluorophores. Raw data (grey and dark green) are plotted alongside data passed through a forward-backward moving window average filter (black and light green). In order to clearly identify "on" and "off" fluorescence states thresholding of the raw data to 10 standard deviations beyond the mean background was used (green dashed line). Using this in

combination with a temporal threshold allowed for identification of photobleaching lifetimes of all single molecules (red dashed lines and times in seconds).

To assess whether the enhanced *in vitro* performance of YuzuFP translates to cellular imaging, we also imaged a LifeAct-YuzuFP fusion protein in live cells using wide-field fluorescence microscopy. For transient transfection, the amount of protein produced in individual cells can vary across the cell population, depending on the number of plasmids introduced into the cell, and the overall activity at the individual cell level.

Distinct actin filament structures were clearly visible in HeLa cells expressing the LifeAct fusion constructs (Figure 5.8 and Figure 5.9). In Figure 5.8 we show a comparison between the two fusions at different expression levels that we determined as high (Hi), medium (mid) and low (Lo) based on their brightness, across different time points up to 120 s of exposure (contrasted equally within columns 1 and 2 to highlight the differences in expression level). When contrasted to show the full dynamic range of the image, each cell can be seen to photobleach across the timeline of the movie. The sfGFP (Figure 5.8a) appears to photobleach more than YuzuFP (Figure 5.8b) over the course of the 120 s movie. While the filament structures became indistinct after 40 seconds in cells expressing the sfGFP-LifeAct fusion, cells expressing YuzuFP-LifeAct retained clear filament visibility for up to 120 seconds, consistent with its improved photostability (Figure 5.9).

Wide-field epi-fluorescence imaging of live cells has shown a higher decrease in fluorescence over time for the sfGFP fusions compared to the YuzuFP fusions, as expected based on our *in vitro* experiments and the single molecule data. However, one can argue that quantifying and comparing fluorescence and photobleaching between the two fusions in live cells should be done with caution. The reason for this is that fluorescence can be highly affected by concentration of protein fusions present as well as quality of expression by the cell. Data were collected from 23 individual regions of interest (ROI) which correspond to 23 different cells.

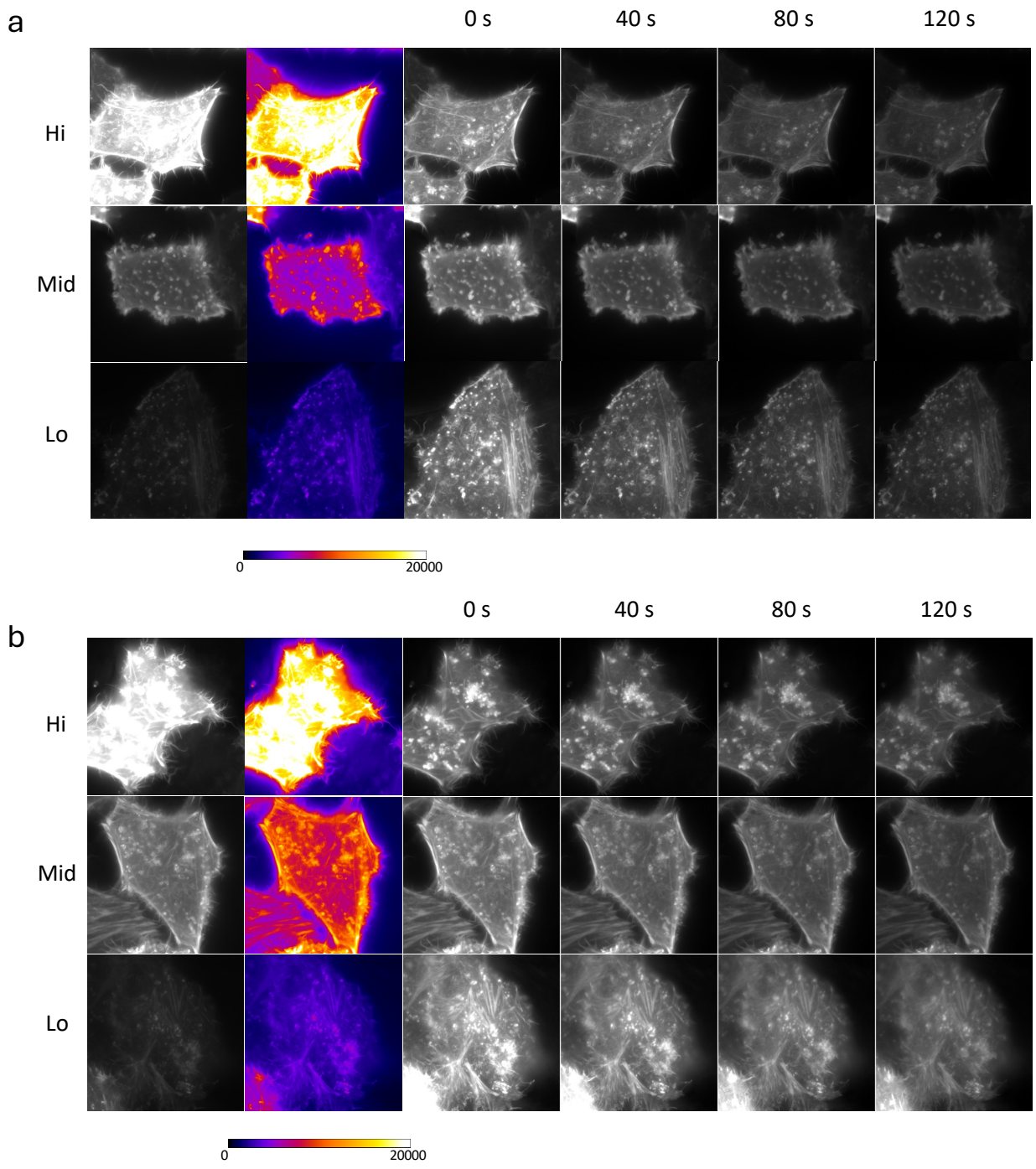

Figure 5.8: Expression of sfGFP and YuzuFP LifeAct Fusions. Wide field image of sfGFP-LifeAact a) and YuzuFP-LifeAct fusions b) at different exposure levels and over time. Second row panel is coloured using a 16-bit Fire look-up table. Black value is set to 0 and white value is set to 20000.

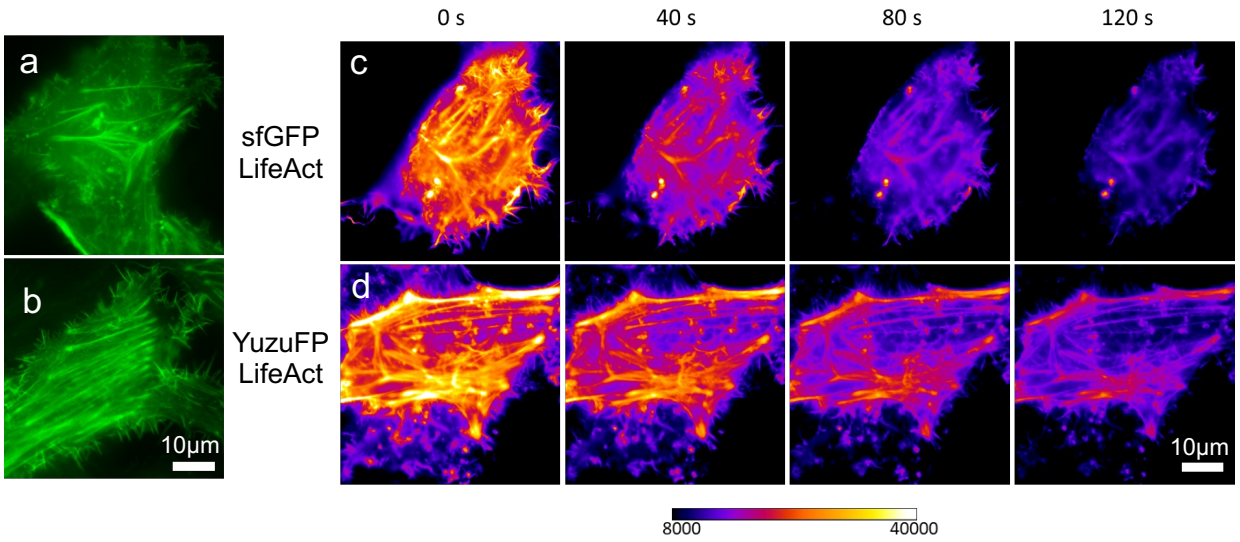

Figure 5.9: Live cell imaging of sfGFP and YuzuFP LifeAct Fusions. Wide-field image of a) sfGFP and b) YuzuFP LifeAct fusions. Time course of sfGFP (c) and YuzuFP (d) LifeAct fusions false coloured using a 16-bit Fire look-up table with the black value set to 8000, and the white value set to 40000, to visualise signal within the dynamic range of the microscope setup. Figure adapted from Ahmed et al. (2025).

The mean level of fluorescence intensity for each sample was measured and plotted to calculate the rate of decrease for each ROI (Figure 5.8). It is evident from Figure 5.8a and b, that there is variability in fluorescence intensities between different samples expressing the same construct. This is expected as the level of transfection and thus expression by each cell is not controlled when performing transient transfection, hence why we see such variability. Despite that variability, overall, it is still evident that YuzuFP photobleaches slower and less over time compared to sfGFP. This shows the importance of conducting a variety of qualitative and quantitative experiments to support a hypothesis. In order to visualise this more clearly, we standardised against expression level by normalising each ROI to timepoint to 0 to calculate the fraction of fluorescence that remains over time (Figure 5.10c and d), which can then be plotted together (Figure 5.10e) as a mean of all ROIs (Figure 5.10f).

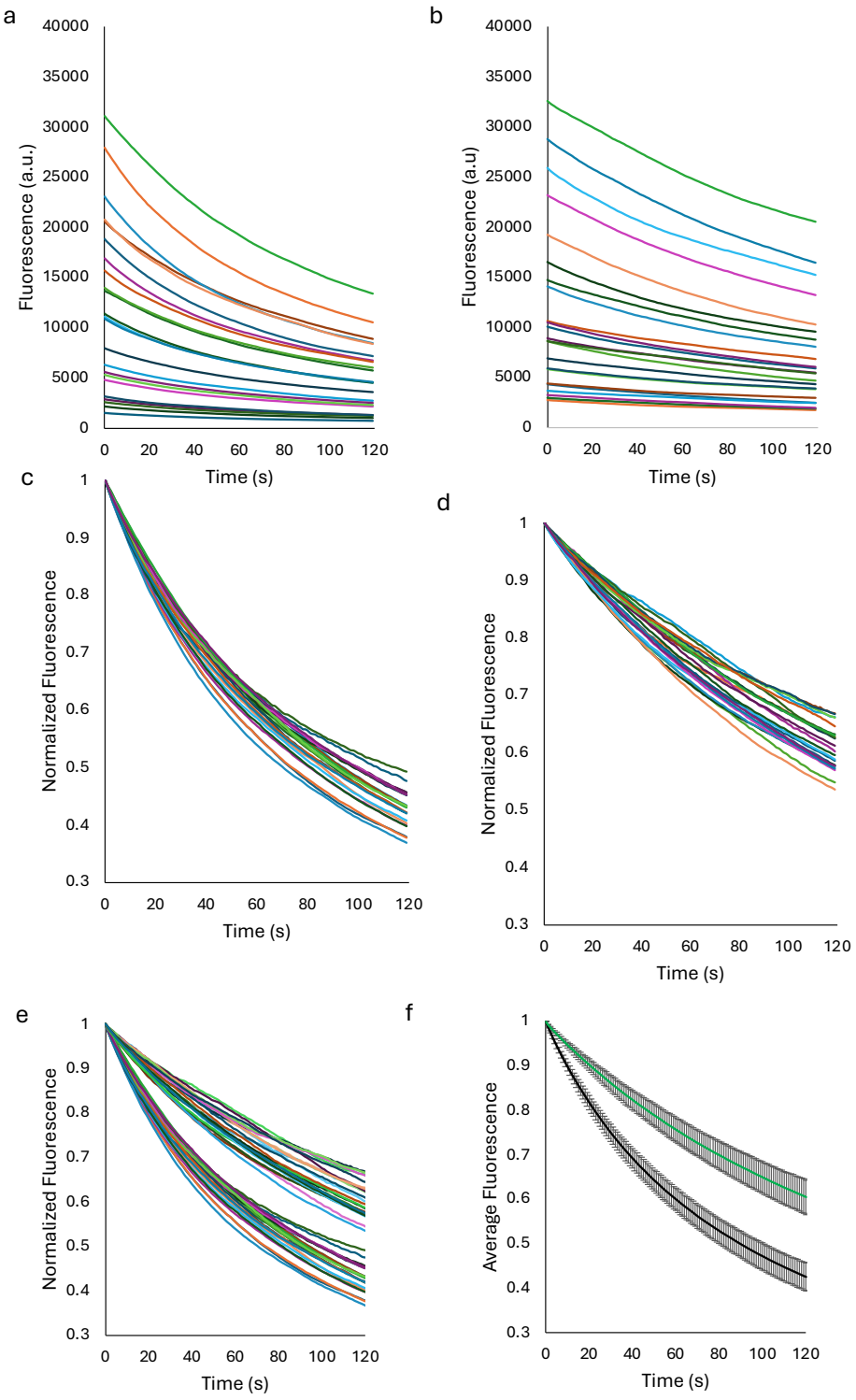

Figure 5.10: Photobleaching profiles of sfGFP versus YuzuFP. Photobleaching in cells expressing a) sfGFP and b) YuzuFP. Fluorescence intensity values were normalised to 1 (at timepoint 0 s) for cells expressing c) sfGFP and d) YuzuFP – LifeAct fusions. e) Normalised fluorescence of sfGFP against YuzuFP. f) Average fluorescence of sfGFP (black) against YuzuFP (green) with standard deviation (black lines).

Finally quantitative analysis revealed that the photobleaching half-life of the YuzuFP fusion was nearly three times longer than that of the sfGFP fusion at approximately 206 ± 3.7 s versus 74 ± 0.5 s respectively (Figure 5.9). Similarly, the fluorescence lifetime increased from 106 ± 0.7 seconds for sfGFP to 297 ± 5.2 seconds for YuzuFP, further confirming the enhanced photostability of YuzuFP relative to the already robust sfGFP.

The decay curve calculated provides additional information to the calculated half-life such as the rate of photobleaching as well as photostability of the proteins. The rate of photobleaching for YuzuFP (k = 0.004) was revealed to be half of that of sfGFP (k= 0.008), meaning that sfGFP photobleaches significantly faster than YuzuFP at any given exposure time. As YuzuFP has a lower photobleaching rate constant, it can therefore withstand longer periods of excitation light exposure before its fluorescence significantly decreases, whereas sfGFP will "fade" more quickly under similar exposure (Figure 5.11).

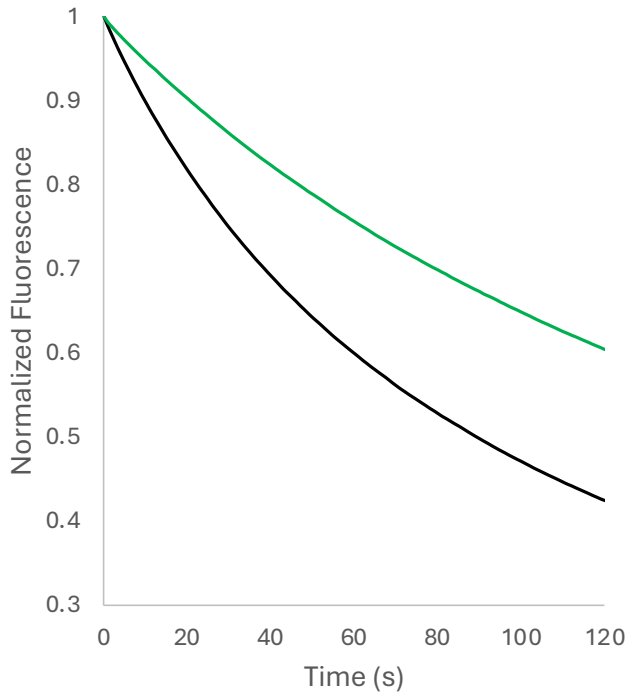

Figure 5.11: Exponential decay of sfGFP vs YuzuFP. Black line corresponds to sfGFP and green line to YuzuFP. sfGFP has a higher rate of photobleaching constant compared to YuzuFP of 0.008 versus 0.004 respectively. All intensity values were used an normalised to 1, averaged and fit into the decay curve.

5.2.4 Crystallization limitations

YuzuFP was set up for crystal trials both at the Medicines Discovery Institute (MDI) at Cardiff University, as well as at Diamond (DLS). Despite trialling a series of different conditions (see Section 2.5.1), crystalising YuzuFP was unsuccessful during this project. Solutions were incubated at room temperature, however, no crystals formed over time but instead, protein either dried out after being incubated for periods of over two months or ended up in an emulsion like state (Figure 5.12).

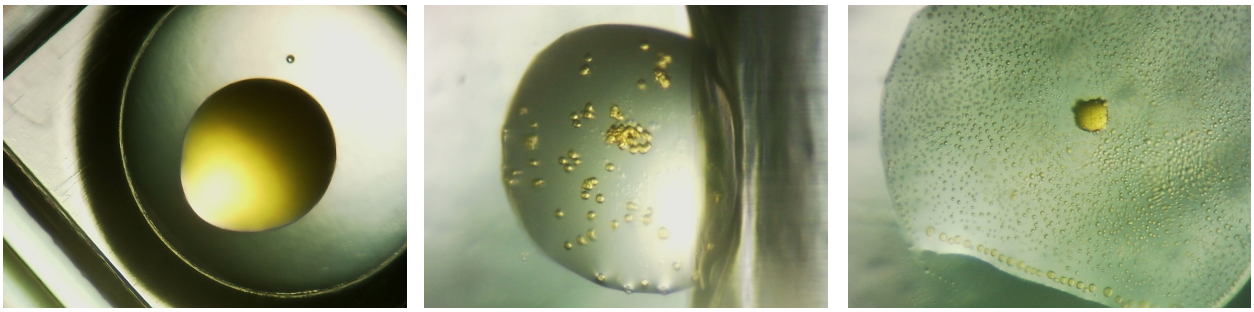

Figure 5.12: Crystal trials of YuzuFP. Examples of failed crystallisation of YuzuFP at different crystallisation conditions. All three pictures were taken at the same after the same period of incubation at one month after crystal trials were set up.

5.2.5 Long time scale molecular dynamics

Since no crystal structure was obtained to provide information as to why we might be seeing the spectral characteristics of YuzuFP, molecular dynamics analysis was utilised to help further understand how the H148S mutation can potentially affect the dynamics of the chromophore and the protein.

To better understand the potential mechanism by which the H148S mutation influences protein behaviour, we extended the initial molecular dynamics simulations to three independent 500 ns simulations on the YuzuFP CroO$^-$ model and compared to sfGFP. Analysis of the RMSF revealed that both proteins exhibit broadly similar backbone dynamics with most variations falling within sub Ångstrom level (Figure 5.12). The region encompassing the loop immediately preceding residue 148 and the subsequent β-strand

(residues 145-151) displayed some difference in flexibility between YuzuFP and sfGFP; however, the ΔRMSF values were minimal and generally less than 0.05 nm (Figure 5.13).

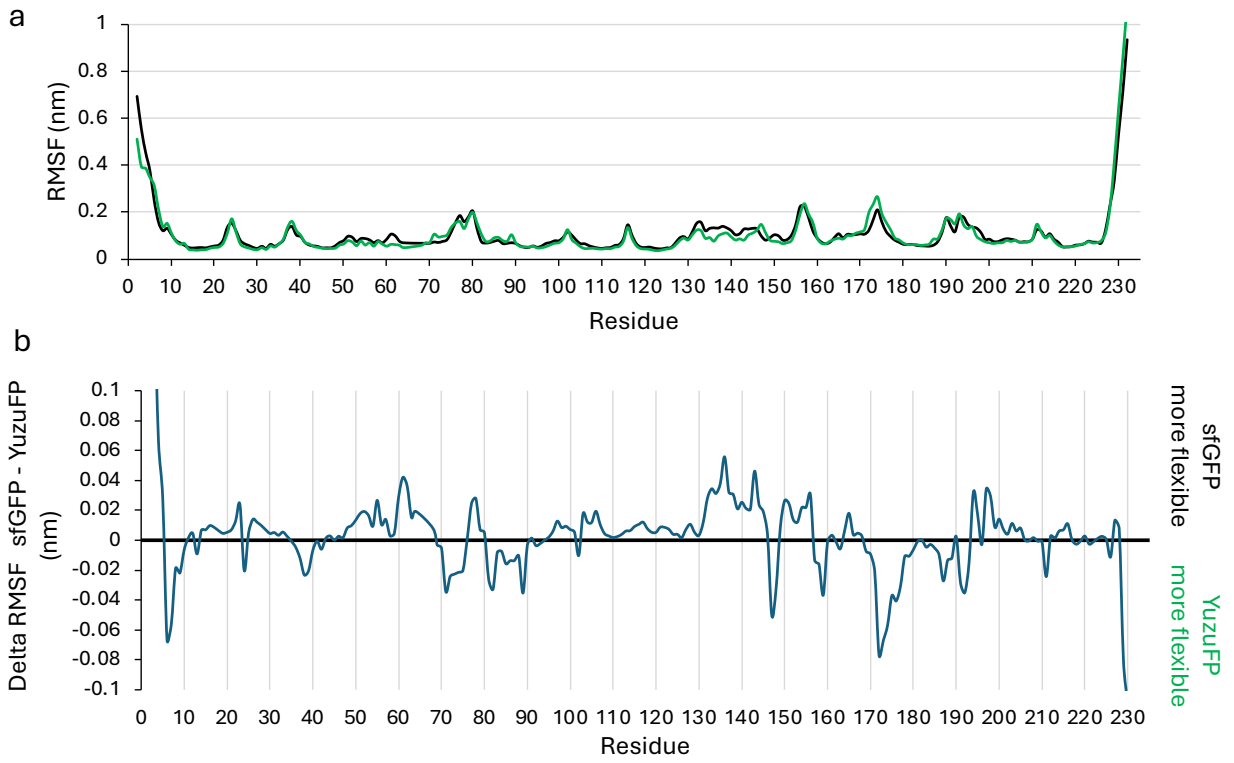

Figure 5.13: Backbone change in sfGFP and YuzuFP over the course of the MD simulation. a) Average per residue backbone (Cα) RMSF of sfGFP (black) and YuzuFP (green). b) Difference plot for per residue Cα RMSF. Difference between the Cα RMSF of the two proteins was calculated by subtracting the values of YuzuFP from that of sfGFP. Positive values indicate more flexibility in sfGFP whereas negative values indicate more flexibility in YuzuFP.

Throughout the simulations, the chromophore structure remained largely stable in both proteins, showing no significant conformational deviations (Figure 5.14a and b). While minor fluctuations were observed in the hydroxyethyl group of the original T65 residue and the backbone oxygen of G67, the majority of atoms within the chromophore remained structurally stable over the simulation period (Figure 5.14c).

A key distinction between YuzuFP and sfGFP lies in the interaction between the chromophore and residue 148. In sfGFP, the histidine at position 148 forms hydrogen bonds with the chromophore infrequently, doing so for only about 12 % of the simulation time. In contrast, the serine at position 148 in YuzuFP engages in hydrogen bonding with the chromophore for approximately 60 % of the time, with two simultaneous hydrogen bonds

forming for 5.5 % of the time (Figure 5.15a), which is consistent with the preliminary short length simulations.

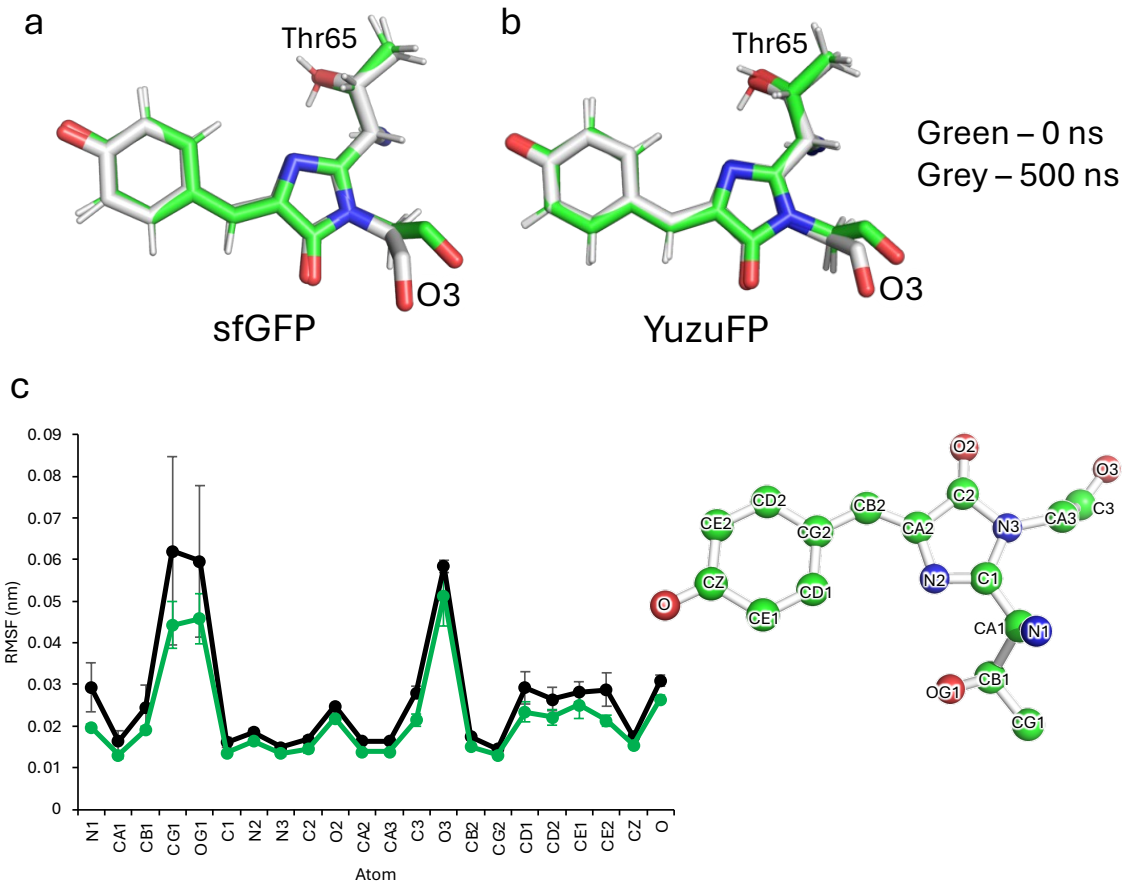

Figure 5.14: Changes in chromophore structure over the course of the MD simulation. Chromophore structure of a) sfGFP and b) YuzuFP at the beginning of a single simulation (green sticks) and after 500 ns (grey sticks). c) RMSF of the chromophore atoms corresponding to the annotated molecular structure (right) for sfGFP (black) and YuzuFP (green). Figure adapted from Ahmed et al. (2025).

These differences in hydrogen bonding behaviour are reflected in the pairwise distribution of distances between the phenolate oxygen of the chromophore and either the side chain atom (Nδ in H148 or Oγ in S148) or the backbone amide nitrogen (Figure 5.15a and b). On average, the side chain of S148 in YuzuFP is 0.1 nm closer to the chromophore phenolate oxygen than H148 in sfGFP (0.42 ± 0.13 nm versus 0.52 ± 0.14nm). Additionally, distance distributions show that the backbone amide nitrogen of S148 in YuzuFP falls within hydrogen bonding distance (<0.35 nm) of the chromophore more frequently, while the same is rarely true for H148 in sfGFP, which generally remains over 0.4 nm away (Figure

5.15b and c). The mean distance between the chromophore's phenolate oxygen and the backbone amide nitrogen is 0.40 ± 0.12 nm in YuzuFP, compared to 0.61 ± 0.10 nm in sfGFP.

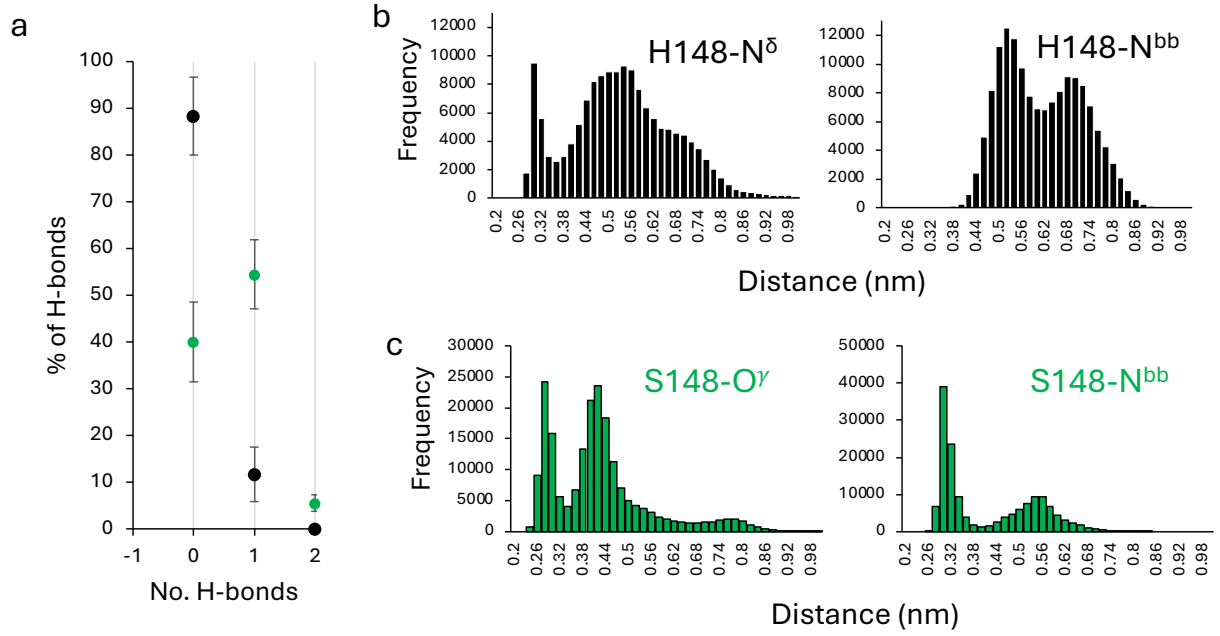

Figure 5.15: Hydrogen bonding between residue 148 and the chromophore. a) Percentage of time that H-bonds are formed between the chromophore and residue 148 through the course of the MD simulation for YuzuFP (green) and sfGFP (black). The error bars show the standard deviation between values measured for the three individual simulations. The pair-wise distribution of the distance in b) sfGFP and c) YuzuFP, between the chromophore phenolate oxygen and residue 148, and H-bond donor heavy side chain atoms across all simulation data. Figure adapted from Ahmed et al. (2025).

Importantly, YuzuFP's S148 can form hydrogen bonds with the chromophore phenolate oxygen through either its side chain or backbone amide; a behaviour not observed in sfGFP. In fact, over half of these hydrogen bonds (55% ± 21%) in YuzuFP involve the backbone amide group. This indicates that the H148S mutation enables dual hydrogen bonding configurations, with the side chain, backbone, or both simultaneously, contributing to chromophore interactions. These interactions are clearly visible in individual molecular dynamic trajectories, where the different bonding configurations can be directly observed (Figure 5.16).

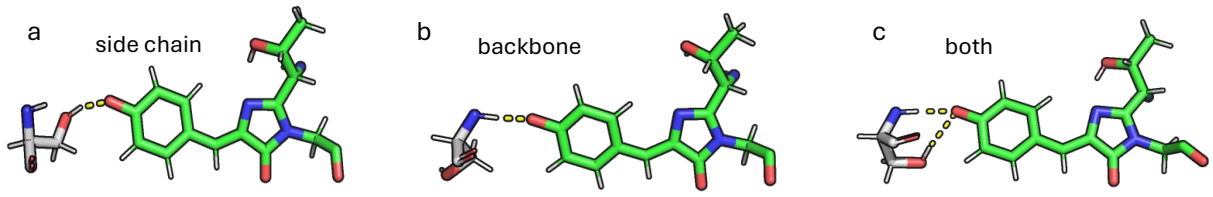

Figure 5.16: Hydrogen bonding between S148 and chromophore. Individual trajectories showing the H-bond between S148 (grey sticks) and the a) side chain hydroxyl group, b) the backbone amine group and c) both in YuzuFP. H-bonds are shown as yellow dotted lines.

The altered hydrogen bonding behaviour in YuzuFP appears to be primarily driven by changes in both the orientation and proximity of residues 148 relative to the chromophore. In sfGFP, the RMSD of H148 remains generally stable throughout most of the simulation but occasionally exhibits significant shifts exceeding 0.1 nm (Figure 5.17a). In contrast, S148 in YuzuFP rapidly adopts a conformation with and RMSD around 0.08, oscillating within a range that remains close to the initial structure (Figure 5.17b).

Distance measurements between the chromophore's phenolate oxygen and the Cα atom of residue 148 support this observation; in YuzuFP, S148 tends to move closer to the chromophore, while H148 in sfGFP tends to move further away (Figure 5.17c and d). The average Cα to phenolate oxygen distance is 0.45 ± 0.08 nm in YuzuFP which is approximately 0.15 nm shorter than the corresponding distance in sfGFP (0.59 ± 0.08 nm).

Despite slightly higher RMSD observed for YuzuFP, S148's hydrogen bonding capacity with the chromophore remains unaffected (Figure 5.17e). The larger fluctuations in RMSD observed for H148 in sfGFP are consistent with the "open" conformational state reported in previous crystal structures (Figure 5.17f) (Arpino et al. 2014). Altogether, the closer positioning of S148 to the chromophore and its strong tendency to form hydrogen bonds likely contribute significantly to the enhanced properties of YuzuFP.

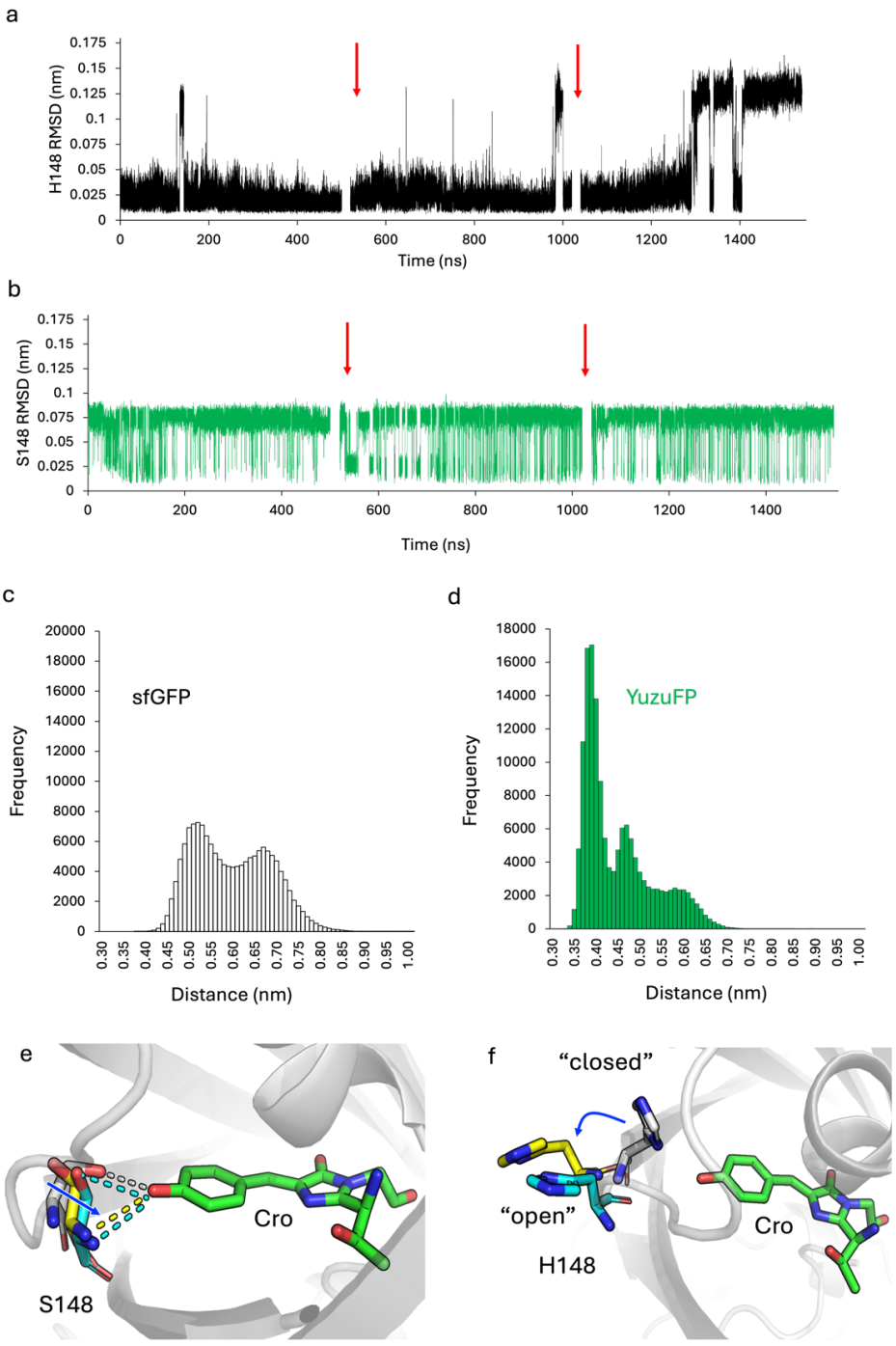

Figure 5.17: Conformational changes of residue 148. RMSD of a) H148 in sfGFP and b) in YuzuFP. Distance distribution between the phenolate oxygen of the chromophore and the Cα of residue 148 for all simulations in c) sfGFP and d) YuzuFP. The red arrows mark the start and end of each 500 ns simulation. e) and f) represent the trajectories of YuzuFP and sfGFP respectively at different time points of the simulation. The dashed lines represent the polar contacts determined in PyMOL. The blue arrow illustrates the direction of backbone movement with respect to the starting trajectory at 0 ns. Figure used with permission by Ahmed et al. (2025).

Additional structural changes may also play a role in these improvements. Notably, residue F145 which is adjacent to position 148, was mutated from tyrosine during the development of sfGFP. This substitution is known to increase protein stability and reduce solvent exposure of the chromophore. F145 is more closely associated with the chromophore in YuzuFP than in sfGFP with the average distance between the F145 aromatic Cζ to chromophore Cζ (Cz in Figure 5.15c) being 0.57 (± 0.07) nm compared to 0.90 (± 0.26) nm respectively. This suggests that the H148S mutation may influence local structure dynamics beyond direct chromophore contact, potentially contributing further to the favourable characteristics of YuzuFP.

5.2.6 The importance of water molecules

Despite the chromophore being buried within the β-barrel structure of the protein, the crystal structure of sfGFP (Pédelacq et al. 2006) has revealed that it is well-solvated, as there are six water molecules present that surround the chromophore (Figure 5.18b). A particular water molecule, labelled W1 in Figure 5.1b, is consistently found in crystal structures of avGFPs and other fluorescent proteins (Figure 5.18b). W1 forms polar interactions with both the chromophore's phenolate oxygen and the side chain of S205, and its positioning is believed to play a role in stabilizing the $CroO^-$ state by facilitating proton shuttling which is ultimately contributing to enhanced fluorescence.

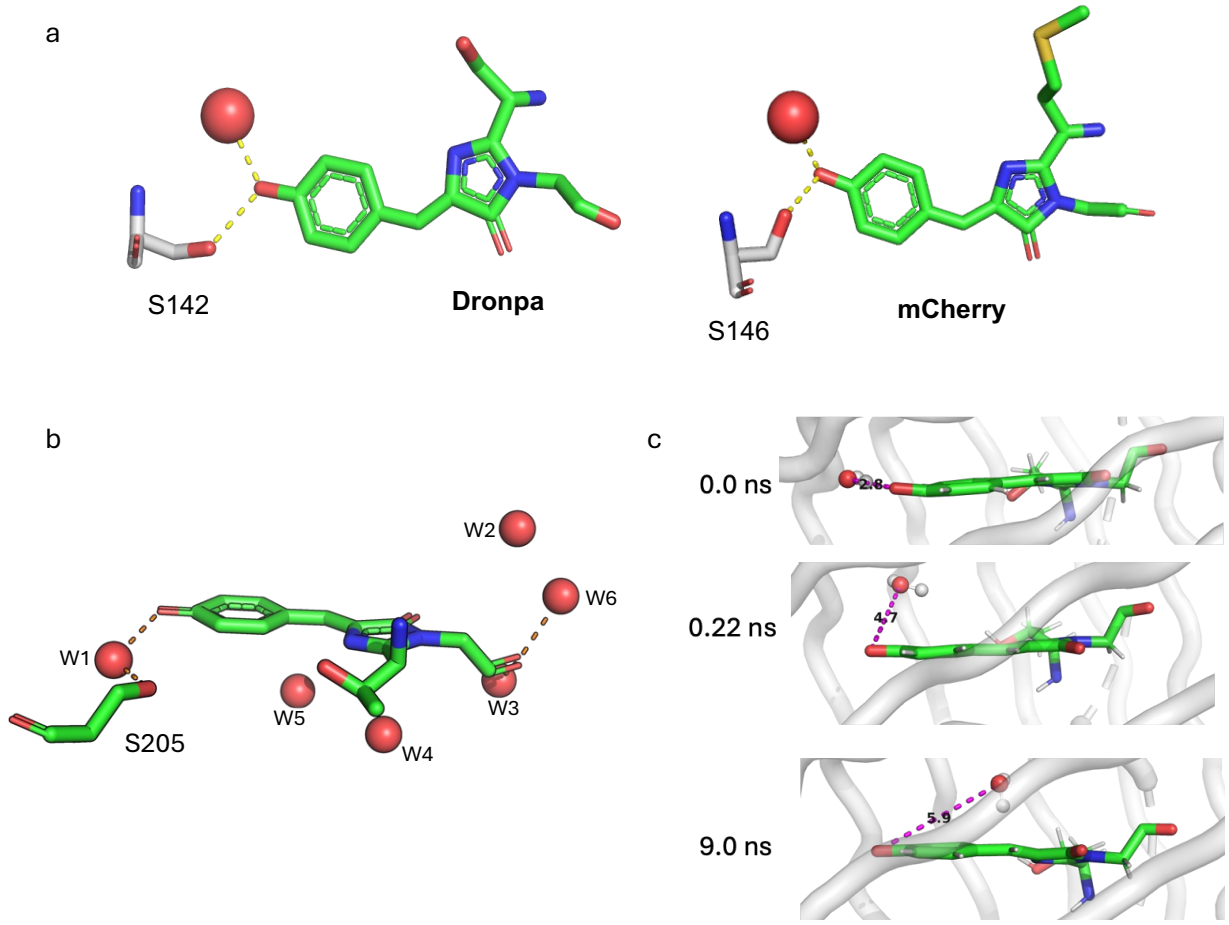

Figure 5.18: Chromophore and water interactions. a) Residues equivalent to H148 in Dronpa (PDB 2IE2) and mCherry (PDB 2H5Q) represented as grey sticks and their interactions (yellow dashed lines) with the phenol group of the chromophore (green sticks). b) The sfGFP (PDB 2B3P) chromophore and water molecules (red spheres) within proximity. Polar interactions between the chromophore and the water molecules are shown as red dashed lines. c) Changes in the position of the W1 water molecule, through the simulation. As the simulation progresses, the water molecule moves further away and thus can no longer H-bond with the phenol oxygen of the chromophore. Dashed lines and number represent the distance between the W1 water molecule and the chromophore oxygen in Å.

MD simulations conducted for this project revealed that W1 is only transiently associated with its original position near the chromophore in both sfGFP and YuzuFP before being replaced by other water molecules from the surrounding solvent (Figure 5.19). However, W1 exhibits a longer average residence time in YuzuFP of approximately 8.5 ns unlike sfGFP at 1.5 ns. Interpreting this behaviour is complex. For example, in one of the sfGFP simulations, W1 rapidly becomes internalized within 0.22 ns relocating near the chromophore (Figure 5.18c). Throughout the simulations, W1 is continually replaced by

other water molecules. Despite these dynamics, the chromophore's phenolate oxygen in YuzuFP is unbound to water for only 5.7 % of the time, while in sfGFP, this water-free period is slightly increased to 6.8 %.

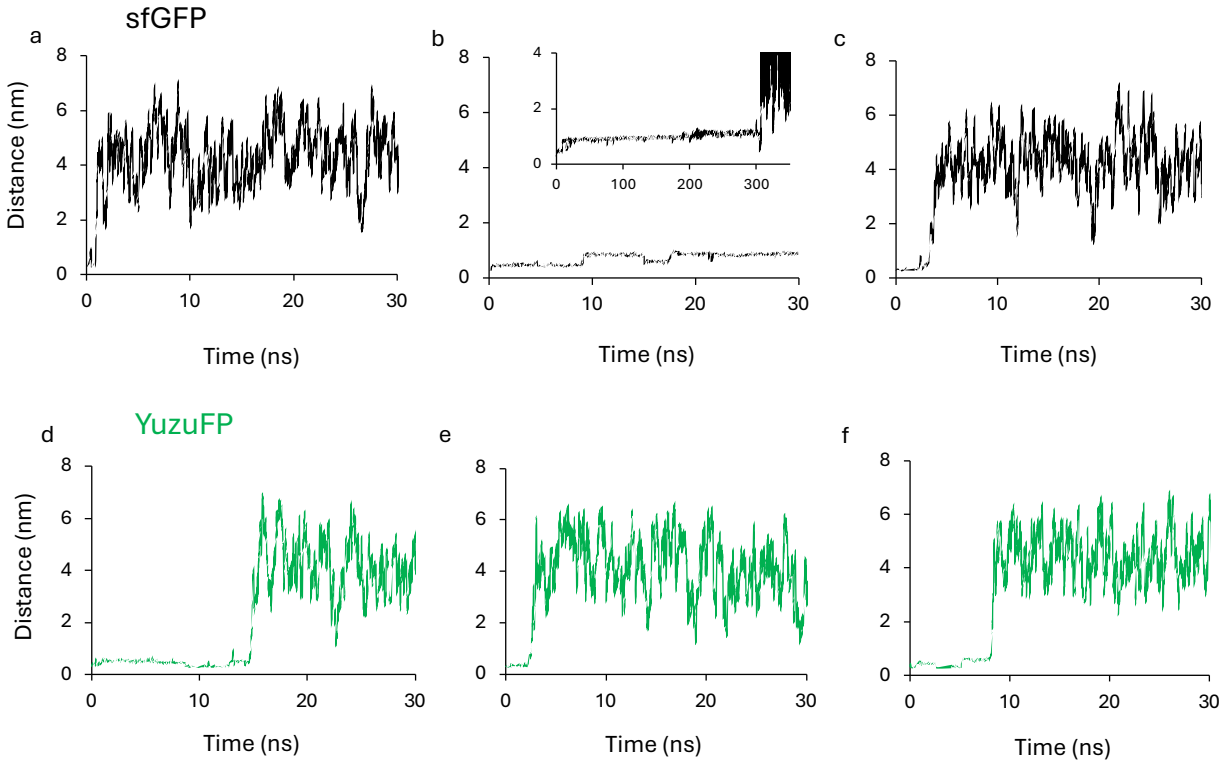

Figure 5.19: Residency of W1 water molecule around the chromophore phenolate oxygen. Graphs a-c) show the distance between the chromophore phenolate oxygen and the oxygen atom of the W1 water molecule in sfGFP and d-f) in YuzuFP in each of the 500 ns simulation. The insert in b shows the distance over an extended period of time (350 ns).

The remaining water molecules located within the β-barrel structure tend to exhibit longer residence times; however, most eventually exchange with the surrounding bulk solvent. Tunnel analysis performed using CAVER revealed a potential solvent-accessible pathway near residue H148 in sfGFP (Figure 5.20), which may serve as a route for water exchange with the external environment. Despite this, some water molecules remain in close proximity to the chromophore throughout the simulations. In YuzuFP, for example, the water molecule designated W2 remains buried for the entire duration of two simulations, while in sfGFP, W2 exchanges with the solvent within 172 ns (Table 5.4).

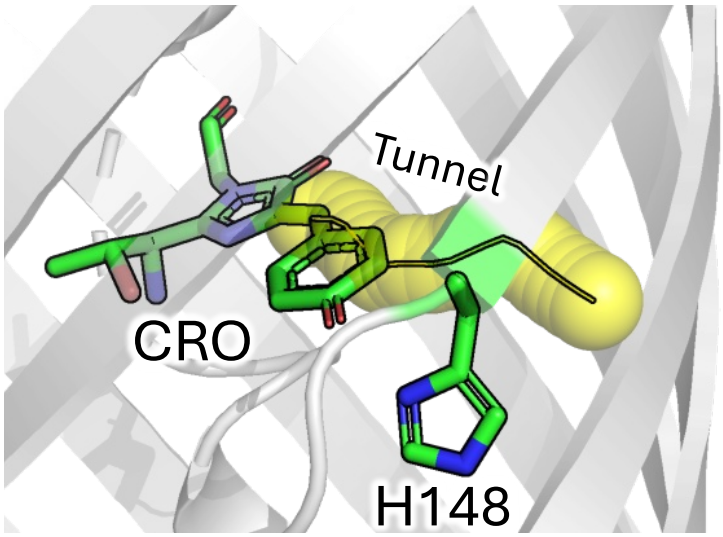

Figure 5.20: Tunnel calculated from CAVER analysis. Tunnel (yellow) between the chromophore and solvent.

Furthermore, during the third simulation of YuzuFP, two additional water molecules (W5 and W6) remained sequestered within the protein for the full 500 ns trajectory. On the other hand, only two buried water molecules persisted throughout a single simulation of sfGFP, with no such retention observed in the others (Table 5.4).

These differences in water exchange dynamics may reflect variations in chromophore pocket accessibility to external solutes such as reactive oxygen species, which are known to promote photobleaching through covalent modifications and disruption of hydrogen bonding networks. Therefore, mutations at position 148, through altering local structural dynamics may influence tunnel formation or aperture size, potentially modulating solvent and solute access to the chromophore environment.

Table 5.4: Residency time of water molecules around the chromophore.

	Residency time (ns)[a]							
	sfGFP				YuzuFP			
Water	Sim 1	Sim 2	Sim 3	**Average**	Sim 1	Sim 2	Sim 3	**Average**
W1	3.36	306.55 (0.22)[c]	0.89	**103.6 (1.49)**	14.79	8.15	2.64	**8.53**
W2	171.46	124.14	119.19	**138.26**	500[b]	63.83	500[b]	**354.61**
W3	33.54	200.19	500[b]	**244.58**	89.69	248.41	11.62	**116.57**
W4	50.77	188	119.19	**119.32**	37.14	162.39	2.88	**67.47**
W5	41.51	221.98	68.77	**110.75**	25.56	500[b]	42.15	**189.24**
W6	66.3	89.21	500[b]	**218.50**	73.43	500[b]	161.45	**244.96**

a, water residency time close to the chromophore a determined by a distance consistency (0.1 ns) above a distance of 1.5 nm from the b-methylene bridge atom.
b, retained residency over the course of the 500 ns simulation. The average is thus calculated with these values set to 500 ns.
c, W1 in sfGFP Sim 2 moves from its original position to more internal position close to the chromophore. W1 becomes internalised within 0.22 ns (see Figure S11).

5.3 Conclusions

This chapter presents the characterisation of YuzuFP, a fluorescent protein engineered for enhance photostability and performance. Compared to its parent protein sfGFP, YuzuFP exhibits significantly improved brightness, reduced blinking, and a longer fluorescence lifetime at the single-molecule level. These advantages were confirmed through TIRF microscopy and live-cell imaging, where YuzuFP-LifeAct fusions enabled extended visualisation of actin filaments in HeLa cells, outperforming sfGFP equivalents.

Molecular dynamics simulations revealed that these improvements are largely attributed to altered chromophore interactions resulting from the H148S mutation. In YuzuFP, S148 forms more frequent and stable hydrogen bonds with the chromophore than H148 in sfGFP, aided by closer positioning and reduced structural fluctuations. Additionally, nearby residue F145 showed tighter chromophore association in YuzuFP, possibly enhancing chromophore shielding.

Water dynamics within the β-barrel also differed significantly. YuzuFP retained internal water molecules longer and demonstrated reduced solvent exchange near the chromophore, potentially limiting access of photoreactive species. Tunnel analysis

suggested that changes near residue 148 may influence solvent accessibility and contribute to photoprotection.

Overall, this work demonstrates how targeted mutagenesis, and structural optimization can enhance the functional properties of fluorescent proteins offering practical benefits for long-term imaging in live-cell and single-molecule applications.

6 Discussion

6.1 General Overview

Protein engineering has been a powerful tool driving advancements in synthetic biology by enabling the design and modification of proteins with tailored properties for specific biological applications. It focuses on manipulating the amino acid sequence of proteins to alter their structure, function and interactions, thereby expanding the toolkit available for biological research and novel technological development.

Across the three chapters, residue-level targeting near or within the chromophore yielded functional improvements or introduced novel properties. The substitution of H148 with serine in YuzuFP (Ahmed R. 2022) stabilised fluorescence through dual hydrogen bonding. Similarly, M66C in mCherry altered chromophore protonation dynamics and enhanced sensitivity. These mutations underscore how local environment tuning; especially around the chromophore, can modulate fluorescence lifetime, quantum yield and environmental responsiveness.

In contrast, the incorporation of bulky or chemically orthogonal nnAAs like pCNPhe and pCCPhe proved disruptive, especially when placed within the chromophore. These modifications abolished fluorescence and failed to deliver Raman-active signals, suggesting that chemical compatibility and folding integrity must be carefully balanced.

6.2 Engineering of mCherry and mNeptune as Raman probes

Chapter 3 presents a rigorous exploration into expanding the toolkit of fluorescent proteins for multimodal imaging by introducing chemically orthogonal Raman active vibrational modes. Target residues in or near the chromophore pocket were selected based on structural modelling and previous studies (Reddington et al. 2015), and were mutated to incorporate the amber stop codon for nnAA insertion using an orthogonal tRNA/aminoacyl-tRNA synthetase system (Zitti and Jones 2023). Successful incorporation was confirmed by protein expression and SDS-PAGE analysis, although colour changes and spectral shifts indicated altered chromophore maturation. Spectral analysis revealed blue shifts in absorption and emission for most mCherry variants, accompanied by decreased

fluorescence. In mNeptune, incorporation of pCNPhe or pCCPhe at the chromophore tyrosine (Y67) abolished both absorbance and emission. Finally, epr-SRS spectroscopy showed minimal to no enhancement in the biologically silent region, with further limitations arising from low protein yields and suboptimal microscope excitation wavelengths. Moreover, despite both mCherry and mNeptune having the properties for pre-resonant coherent Raman scattering excitation, a huge hindrance to this study was caused due to the laser set-up we currently have at Cardiff University as it does not cover the required wavelength for our red-fluorescent proteins where their absorbance maximum is up to 600 nm. Because of that, proteins samples had to be sent to an external collaborator for measurements delaying data acquisition.

6.3 Development of an mCherry-based fluorescent biosensor

In chapter 4, the properties of the engineered mCherry^{M66C} mutant were explored and analysed. The strategic substitution of methionine with cysteine at residue 66 in mCherry represents a compelling case of chromophore-level protein engineering, capitalising on the unique vibrational and chemical properties of thiols (Bazylewski et al. 2017) as originally the aim was to generate an alternative Raman-probe to those generated for chapter 3 with the incorporation of nnAAs. The mutagenesis approach successfully generated a functionally fluorescent protein, as evidenced by the formation of pink colonies and robust spectral signatures. The 21 nm blue shift in absorbance and corresponding 25 nm emission shift are consistent with disruption of the conjugated π-system, reflecting the sensitive interplay between chromophore chemistry and photophysical properties.

The most notable advancement of this work lies in the dual pKa response of mCherry^{M66C}, signifying two discrete protonation events likely arising from the thiol and phenolic groups within the chromophore. This three-state model, validated by QM simulations, confers a level of environmental sensitivity absent in mCherry, which exhibits only a single acidic transition (pKa of 4.7). Such behaviour suggests potential utility in ratiometric pH imaging, particularly in systems with fluctuating intracellular pH values.

Contrary to initial expectations, Raman spectral enhancement was lower for the mutant, likely due to excitation mismatch (λ_{max} = 566 nm and pump beam = 640 nm) rather than diminished chromophore activity.

Nevertheless, the study pivoted effectively to explore environmental responsiveness, with detailed H_2O_2 experiments revealing enhanced oxidative resilience of mCherry^{M66C} over mCherry. The mass spectrometry data suggest comparable oxidative modifications (~58 Da loss), though the chemical nature remains uncertain, warranting further investigation.

Both of these two novel features of mCherry^{M66C}, make it an excellent candidate for as molecular biosensor. Current biosensors like pHluorin (Reifenrath and Boles 2018), and SypHer (Matlashov et al. 2015) are sensitive to changes in pH and H_2O_2 respectively. Unlike mCherry^{M66C} both SypHer and pHluorin have been developed from yellow fluorescent protein (YFP) and GFP respectively, and they emit at 509-516 nm. Both have dual excitation and can therefore enable ratiometric imaging. Having mCherry^{M66C} introduced into the toolkit will allow multiplexing possibilities across cellular environments were multiple components of similar pH or oxidative potential could be labelled to allow reporting of multiple components at a time. In addition to that, mCherry^{M66C} also has the potential to act as a multisensor as it is sensitive to both pH and H_2O_2 which can also eliminate the need for multiple reporters, which can be useful in cellular environments that cannot withstand multiple reporters or experience cell stress at high protein expression levels i.e. use of nuclear -localized β-galactosidase and EGFP in primary cortical neurons (Detrait et al. 2002).

Moreover, the X-ray crystal structure at 2.04 Å resolution provided high-confidence structural insights. It revealed a subtly enlarged chromophore pocket and increased hydration, confirmed by MD simulations. These simulations further illustrated enhanced rigidity and reduced chromophore flux in mCherry^{M66C}, especially around the phenol and thiol functional groups. Importantly, this rigidity aligns with the observed increase in quantum yield (from 0.22 in mCherry to 0.27).

The study also introduces the concept of dynamic water-mediated entry/exit points as determinants of chromophore accessibility and reactivity. In mCherry^{M66C}, a shorter thiol side chain permits a new tunnel and alternate water transport route, possibly influencing chromophore protonation kinetics. Tunnel mapping via CAVER reinforced this idea, linking structural voids with functional pKa transitions and supporting the biosensor potential of mCherryM66C.

Finally, this work situates its findings within the broader context of RFP evolution, comparing the role of residue 66 across variants such as mRFP1, mTangerine, and mGrape1. The insight that water accessibility, hydrogen bonding, and chromophore rigidity are modulated by residue 66 supports a new framework for rational fluorescent protein design.

6.4 Development of YuzuFP

The mechanism by which the H148S mutation enhances photostability is not yet fully understood. However, the successful development of YuzuFP underscores the power of structure-guided mutagenesis in tailoring fluorescent proteins for specific imaging needs. The substitution of histidine with serine at position 148 in sfGFP (Ahmed R. 2022) was both strategically rational and experimentally validated, addressing key limitations of the parent protein such as brightness and photostability, without sacrificing spectral properties or chromophore integrity.

Molecular dynamics simulations played a critical role in predicting the behaviour of residue 148 across 19 amino acid substitutions. Serine emerged as the most promising candidate due to its ability to form frequent and geometrically favourable hydrogen bonds with the chromophore phenolate. This contrasts sharply with the native histidine, which interacted with the chromophore for less than 5% of the simulation time. Importantly, YuzuFP's S148 engaged in hydrogen bonding nearly 49% of the time in short MD simulations and approximately 60% in long 500 ns simulations suggesting a significant stabilising influence. The structural dynamics also revealed dual bonding modes in YuzuFP, via both the side chain hydroxyl and backbone amine, contributing to chromophore

stability. This dual anchoring likely reduces flexibility in the chromophore pocket, impeding pathways for photobleaching-inducing molecular interactions such as oxidation or proton transfer disruptions (Duan et al. 2013).

Experimentally, spectroscopic data confirmed the computational predictions as YuzuFP showed increased brightness, small red shifts in excitation/emission (consistent with chromophore environment changes), and a similar pKa to sfGFP (6.1 and 6.2; Table 5.3), indicating that fluorescence pH-sensitivity was preserved. *In vitro* pH studies also revealed the presence of an isosbestic point, suggesting well-defined equilibrium between protonated and deprotonated chromophore states.

Moreover, its improved fluorescence properties (increased half-life and resistance to photobleaching) were further linked to altered water dynamics. MD simulations revealed that YuzuFP retained key internal water molecules (i.e. W1) longer than sfGFP, which may shield the chromophore from photodamaging reactive oxygen species. Tunnel mapping revealed that mutation at residue 148 modulate internal solvent accessibility, providing a plausible photoprotection mechanism.

While crystallographic confirmation was not achieved due to failed crystallisation attempts, the combination of extensive computational modelling, spectroscopy, and live-cell imaging built a compelling, multi-layered case for the superiority of YuzuFP over sfGFP and EGFP.

In broader terms, this work exemplifies how atomistic insights can be directly translated into functional gains, advancing the field of bioimaging. The dual approach of rational design and empirical validation employed here serves as a template for the next generation of genetically encoded fluorescent tools.

6.5 Synopsis and future work

In summary, this thesis highlighted the importance of utilising computational methods to help inform experimental data, to provide a rational prior to experimental design as well as drive research forward. Aside from that, both the research undertaken for mCherry as well as sfGFP, despite having different end-goals, the underlying research has

expanded our understanding of how small structural perturbations can influence local environment conditions and how these translate to the overall function. In addition to that, experimental and computational analysis has helped shape our understanding of how H-bonding and water molecule interactions affect the stabilisation of the protonation states of the chromophore allowing a more informed engineering rational when aiming to improve fluorescent properties.

One of the main limitations from this study was the inability to produce fluorescent proteins with incorporated non-natural amino acids within the chromophore that retained their fluorescence. This effect was attributed to steric clashes that disrupted the conjugated system of the chromophore; therefore, future efforts should try to limit this by incorporating alternative nnAAs with smaller side chains or incorporating amino acids with the functional group in a different carbon position. Current work undertaken by other members in the Jones lab, suggests the incorporation of nnAAs where the functional group is in the third position rather than the fourth with the aim to prevent steric clashes and prevent disruption of the chromophore conjugated system.

Work on our published YuzuFP continues, where changes in the local chromophore environment are tested experimentally, to allow comparisons and draw clearer conclusions when compared to the simulated data we have acquired. As well as incorporating serine in other avGFP-like proteins to further understand its effect on the chromophore environment as well as test if it gives similar advantages in spectral and structural properties. Finally, further efforts should be made in crystalising YuzuFP and solving its structure as that would offer more accurate structural information when running simulations.

7 Bibliography

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindahl, E. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, pp. 19–25. doi: 10.1016/j.softx.2015.06.001.

Ahmed, R.D. 2022. Understandingg and Engineering Protein Oligomers. PhD Thesis. Cardiff University

Ahmed, R.D. et al. 2025. Molecular dynamics guided identification of a brighter variant of superfolder Green Fluorescent Protein with increased photobleaching resistance. *Communications Chemistry* 8(1), p. 174. doi: 10.1038/s42004-025-01573-4.

Aksakal, O. 2024 Developing and Characterising Genetically Encoded Vibrational Probes for Next-Generation Bioimaging Applications using Electronic Pre-Resonant Raman Spectrocopy. PhD Thesis. Cardiff University.

Ando, R., Mizuno, H. and Miyawaki, A. 2004. Regulated Fast Nucleocytoplasmic Shuttling Observed by Reversible Protein Highlighting. *Science* 306(5700), pp. 1370–1373. doi: 10.1126/science.1102506.

Aronson, D.E., Costantini, L.M. and Snapp, E.L. 2011. Superfolder GFP is fluorescent in oxidizing environments when targeted via the Sec translocon. *Traffic (Copenhagen, Denmark)* 12(5), pp. 543–8. doi: 10.1111/j.1600-0854.2011.01168.x.

Arpino, J.A.J., Rizkallah, P.J. and Jones, D.D. 2012. Crystal Structure of Enhanced Green Fluorescent Protein to 1.35 Å Resolution Reveals Alternative Conformations for Glu222. *PLoS ONE* 7(10), p. e47132. doi: 10.1371/journal.pone.0047132.

Arpino, J.A.J., Rizkallah, P.J. and Jones, D.D. 2014. Structural and dynamic changes associated with beneficial engineered single-amino-acid deletion mutations in enhanced green fluorescent protein. *Acta Crystallographica Section D Biological Crystallography* 70(8), pp. 2152–2162. doi: 10.1107/S139900471401267X.

Auhim, H.S. et al. 2021. Stalling chromophore synthesis of the fluorescent protein Venus reveals the molecular basis of the final oxidation step. *Chemical Science* 12(22), pp. 7735–7745. doi: 10.1039/D0SC06693A.

Baird, G.S., Zacharias, D.A. and Tsien, R.Y. 1999. Circular permutation and receptor insertion within green fluorescent proteins. *Proceedings of the National Academy of Sciences* 96(20), pp. 11241–11246. doi: 10.1073/pnas.96.20.11241.

Baird, G.S., Zacharias, D.A. and Tsien, R.Y. 2000. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences* 97(22), pp. 11984–11989. doi: 10.1073/pnas.97.22.11984.

Barondeau, D.P., Kassmann, C.J., Tainer, J.A. and Getzoff, E.D. 2002. Structural Chemistry of a Green Fluorescent Protein Zn Biosensor. *Journal of the American Chemical Society* 124(14), pp. 3522–3524. doi: 10.1021/ja0176954.

Bazewicz, C.G., Lipkin, J.S., Smith, E.E., Liskov, M.T. and Brewer, S.H. 2012. Expanding the Utility of 4-Cyano- <scp>l</scp> -Phenylalanine As a Vibrational Reporter of Protein Environments. *The Journal of Physical Chemistry B* 116(35), pp. 10824–10831. doi: 10.1021/jp306886s.

Bazylewski, P., Divigalpitiya, R. and Fanchini, G. 2017. In situ Raman spectroscopy distinguishes between reversible and irreversible thiol modifications in l-cysteine. *RSC Advances* 7(5), pp. 2964–2970. doi: 10.1039/c6ra25879d.

Berman, H., Henrick, K. and Nakamura, H. 2003. Announcing the worldwide Protein Data Bank. *Nature Structural & Molecular Biology* 10(12), pp. 980–980. doi: 10.1038/nsb1203-980.

Berman, H.M. 2000. The Protein Data Bank. *Nucleic Acids Research* 28(1), pp. 235–242. doi: 10.1093/nar/28.1.235.

Bjelkmar, P., Larsson, P., Cuendet, M.A., Hess, B. and Lindahl, E. 2010. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *Journal of Chemical Theory and Computation* 6(2), pp. 459–466. doi: 10.1021/ct900549r.

Blight, S.K. et al. 2004. Direct charging of tRNACUA with pyrrolysine in vitro and in vivo. *Nature* 431(7006), pp. 333–335. doi: 10.1038/nature02895.

Blondel, A. and Karplus, M. 1996. New formulation for derivatives of torsion angles and improper torsion angles in molecular mechanics: Elimination of singularities. *Journal of Computational Chemistry* 17(9), pp. 1132–1141. doi: 10.1002/(SICI)1096-987X(19960715)17:9<1132::AID-JCC5>3.0.CO;2-T.

Bulina, M.E., Chudakov, D.M., Mudrik, N.N. and Lukyanov, K.A. 2002. Interconversion of Anthozoa GFP-like fluorescent and non-fluorescent proteins by mutagenesis. *BMC Biochemistry* 3(1), p. 7. doi: 10.1186/1471-2091-3-7.

Burgstaller, S. et al. 2019. pH-Lemon, a Fluorescent Protein-Based pH Reporter for Acidic Compartments. *ACS Sensors* 4(4), pp. 883–891. doi: 10.1021/acssensors.8b01599.

Campanella, B., Palleschi, V. and Legnaioli, S. 2021. Introduction to vibrational spectroscopies. *ChemTexts* 7(1), p. 5. doi: 10.1007/s40828-020-00129-4.

Campanini, B. et al. 2013. Role of histidine 148 in stability and dynamics of a highly fluorescent GFP variant. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1834(4), pp. 770–779. doi: 10.1016/j.bbapap.2013.01.014.

Campbell, R.E., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S., Zacharias, D.A. and Tsien, R.Y. 2002. A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences* 99(12), pp. 7877–7882. doi: 10.1073/pnas.082243699.

Chica, R.A., Moore, M.M., Allen, B.D. and Mayo, S.L. 2010. Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proceedings of the National Academy of Sciences of the United States of America* 107(47), pp. 20257–62. doi: 10.1073/pnas.1013910107.

Choorakuttil, A.J.X. et al. 2023. Electronically preresonant stimulated Raman scattering microscopy in the visible. *Applied Physics Letters* 123(16). doi: 10.1063/5.0171725.

Chung, S.H. and Kennedy, R.A. 1991. Forward-backward non-linear filtering technique for extracting small biological signals from noise. *Journal of Neuroscience Methods* 40(1), pp. 71–86. doi: 10.1016/0165-0270(91)90118-J.

Craggs, T.D. 2009. Green fluorescent protein: structure, folding and chromophore maturation. *Chemical Society Reviews* 38(10), p. 2865. doi: 10.1039/b903641p.

Crameri, A., Whitehorn, E.A., Tate, E. and Stemmer, W.P. 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature biotechnology* 14(3), pp. 315–9. doi: 10.1038/nbt0396-315.

Cranfill, P.J. et al. 2016. Quantitative assessment of fluorescent proteins. *Nature Methods* 13(7), pp. 557–562. doi: 10.1038/nmeth.3891.

Dang, X. et al. 2019. Deep-tissue optical imaging of near cellular-sized features. *Scientific Reports* 9(1), p. 3873. doi: 10.1038/s41598-019-39502-w.

DeLeon, E.R. et al. 2016. A case of mistaken identity: are reactive oxygen species actually reactive sulfide species? *American journal of physiology. Regulatory, integrative and comparative physiology* 310(7), pp. R549-60. doi: 10.1152/ajpregu.00455.2015.

Demchenko, A.P. 2020. Photobleaching of organic fluorophores: quantitative characterization, mechanisms, protection. *Methods and Applications in Fluorescence* 8(2), p. 022001. doi: 10.1088/2050-6120/ab7365.

Detrait, E.R., Bowers, W.J., Halterman, M.W., Giuliano, R.E., Bennice, L., Federoff, H.J. and Richfield, E.K. 2002. Reporter Gene Transfer Induces Apoptosis in Primary Cortical Neurons. *Molecular Therapy* 5(6), pp. 723–730. doi: 10.1006/mthe.2002.0609.

Dietzek, B., Cialla, D., Schmitt, M. and Popp, J. 2018. Introduction to the Fundamentals of Raman Spectroscopy. pp. 47–68. doi: 10.1007/978-3-319-75380-5_3.

Dodo, K., Fujita, K. and Sodeoka, M. 2022. Raman Spectroscopy for Chemical Biology Research. *Journal of the American Chemical Society* 144(43), pp. 19651–19667. doi: 10.1021/jacs.2c05359.

Drobizhev, M., Tillo, S., Makarov, N.S., Hughes, T.E. and Rebane, A. 2009. Color Hues in Red Fluorescent Proteins Are Due to Internal Quadratic Stark Effect. *The Journal of Physical Chemistry B* 113(39), pp. 12860–12864. doi: 10.1021/jp907085p.

Duan, C. et al. 2013. Structural Evidence for a Two-Regime Photobleaching Mechanism in a Reversibly Switchable Fluorescent Protein. *Journal of the American Chemical Society* 135(42), pp. 15841–15850. doi: 10.1021/ja406860e.

Duwé, S. and Dedecker, P. 2019. Optimizing the fluorescent protein toolbox and its use. *Current Opinion in Biotechnology* 58, pp. 183–191. doi: 10.1016/j.copbio.2019.04.006.

Enterina, J.R., Wu, L. and Campbell, R.E. 2015. Emerging fluorescent protein technologies. *Current Opinion in Chemical Biology* 27, pp. 10–17. doi: 10.1016/j.cbpa.2015.05.001.

Ermakova, Y.G. et al. 2014. Red fluorescent genetically encoded indicator for intracellular hydrogen peroxide. *Nature Communications* 5(1), p. 5222. doi: 10.1038/ncomms6222.

Evans, C.L. and Xie, X.S. 2008. Coherent Anti-Stokes Raman Scattering Microscopy: Chemical Imaging for Biology and Medicine. *Annual Review of Analytical Chemistry* 1(1), pp. 883–909. doi: 10.1146/annurev.anchem.1.031207.112754.

Falb, M. et al. 2006. Archaeal N-terminal Protein Maturation Commonly Involves N-terminal Acetylation: A Large-scale Proteomics Survey. *Journal of Molecular Biology* 362(5), pp. 915–924. doi: 10.1016/j.jmb.2006.07.086.

Favreau, P., Hernandez, C., Lindsey, A.S., Alvarez, D.F., Rich, T., Prabhat, P. and Leavesley, S.J. 2013. Thin-film tunable filters for hyperspectral fluorescence microscopy. *Journal of Biomedical Optics* 19(1), p. 011017. doi: 10.1117/1.JBO.19.1.011017.

Follenius-Wund, A. et al. 2003. Fluorescent Derivatives of the GFP Chromophore Give a New Insight into the GFP Fluorescence Process. *Biophysical Journal* 85(3), pp. 1839–1850. doi: 10.1016/S0006-3495(03)74612-8.

Gao, L., Kester, R.T., Hagen, N. and Tkaczyk, T.S. 2010. Snapshot Image Mapping Spectrometer (IMS) with high sampling density for hyperspectral microscopy. *Optics Express* 18(14), p. 14330. doi: 10.1364/OE.18.014330.

García-González, F., Otero, J.C., Ávila Ferrer, F.J., Santoro, F. and Aranda, D. 2024. Linear Vibronic Coupling Approach for Surface-Enhanced Raman Scattering: Quantifying the Charge-Transfer Enhancement Mechanism. *Journal of Chemical Theory and Computation* 20(9), pp. 3850–3863. doi: 10.1021/acs.jctc.4c00061.

Gnanasambandam, R., Nielsen, M.S., Nicolai, C., Sachs, F., Hofgaard, J.P. and Dreyer, J.K. 2017. Unsupervised Idealization of Ion Channel Recordings by Minimum Description Length: Application to Human PIEZO1-Channels. *Frontiers in Neuroinformatics* 11. doi: 10.3389/fninf.2017.00031.

Goulding, A., Shrestha, S., Dria, K., Hunt, E. and Deo, S.K. 2008. Red fluorescent protein variants with incorporated non-natural amino acid analogues. *Protein Engineering Design and Selection* 21(2), pp. 101–106. doi: 10.1093/protein/gzm075.

Greenwald, E.C., Mehta, S. and Zhang, J. 2018. Genetically Encoded Fluorescent Biosensors Illuminate the Spatiotemporal Regulation of Signaling Networks. *Chemical Reviews* 118(24), pp. 11707–11794. doi: 10.1021/acs.chemrev.8b00333.

Griffiths, P.R. 1992. The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules. *Vibrational Spectroscopy* 4(1), p. 121. doi: 10.1016/0924-2031(92)87021-7.

Grigorenko, B.L., Nemukhin, A. V., Polyakov, I. V., Khrenova, M.G. and Krylov, A.I. 2015. A Light-Induced Reaction with Oxygen Leads to Chromophore Decomposition and Irreversible Photobleaching in GFP-Type Proteins. *The Journal of Physical Chemistry B* 119(17), pp. 5444–5452. doi: 10.1021/acs.jpcb.5b02271.

Gross, L.A., Baird, G.S., Hoffman, R.C., Baldridge, K.K. and Tsien, R.Y. 2000. The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences* 97(22), pp. 11990–11995. doi: 10.1073/pnas.97.22.11990.

Handzo, B., Peters, J. and Kalyanaraman, R. 2022. A Fingerprint in a Fingerprint: A Raman Spectral Analysis of Pharmaceutical Ingredients. *Spectroscopy*, pp. 24-30,45. doi: 10.56530/spectroscopy.cm2680s2.

Hanson, G.T. et al. 2002. Green Fluorescent Protein Variants as Ratiometric Dual Emission pH Sensors. 1. Structural Characterization and Preliminary Application. *Biochemistry* 41(52), pp. 15477–15488. doi: 10.1021/bi026609p.

Hartley, A.M., Worthy, H.L., Reddington, S.C., Rizkallah, P.J. and Jones, D.D. 2016. Molecular basis for functional switching of GFP by two disparate non-native post-translational modifications of a phenyl azide reaction handle. *Chemical Science* 7(10), pp. 6484–6491. doi: 10.1039/C6SC00944A.

Heim, R., Cubitt, A.B. and Tsien, R.Y. 1995. Improved green fluorescence. *Nature* 373(6516), pp. 663–664. doi: 10.1038/373663b0.

Heim, R., Prasher, D.C. and Tsien, R.Y. 1994. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences* 91(26), pp. 12501–12504. doi: 10.1073/pnas.91.26.12501.

Heo, S., Kim, S. and Kang, D. 2020. The Role of Hydrogen Peroxide and Peroxiredoxins throughout the Cell Cycle. *Antioxidants* 9(4), p. 280. doi: 10.3390/antiox9040280.

Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M. 1997. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* 18(12), pp. 1463–1472. doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

Hochreiter, B., Pardo-Garcia, A. and Schmid, J. 2015. Fluorescent Proteins as Genetically Encoded FRET Biosensors in Life Sciences. *Sensors* 15(10), pp. 26281–26314. doi: 10.3390/s151026281.

Holzapfel, H.Y., Stern, A.D., Bouhaddou, M., Anglin, C.M., Putur, D., Comer, S. and Birtwistle, M.R. 2018. Fluorescence Multiplexing with Spectral Imaging and Combinatorics. *ACS Combinatorial Science* 20(11), pp. 653–659. doi: 10.1021/acscombsci.8b00101.

Houston, J.P., Valentino, S. and Bitton, A. 2024. Fluorescence Lifetime Measurements and Analyses: Protocols Using Flow Cytometry and High-Throughput Microscopy. pp. 323–351. doi: 10.1007/978-1-0716-3738-8_15.

Icha, J., Weber, M., Waters, J.C. and Norden, C. 2017. Phototoxicity in live fluorescence microscopy, and how to avoid it. *BioEssays* 39(8). doi: 10.1002/bies.201700003.

Jach, G., Pesch, M., Richter, K., Frings, S. and Uhrig, J.F. 2006. An improved mRFP1 adds red to bimolecular fluorescence complementation. *Nature Methods* 3(8), pp. 597–600. doi: 10.1038/nmeth901.

Jenkins, A.L., Larsen, R.A. and Williams, T.B. 2005. Characterization of amino acids using Raman spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 61(7), pp. 1585–1594. doi: 10.1016/j.saa.2004.11.055.

Johnson, F.H., Shimomura, O., Saiga, Y., Gershman, L.C., Reynolds, G.T. and Waters, J.R. 1962. Quantum efficiency of *Cypridina* luminescence, with a note on that of *Aequorea*.

Journal of Cellular and Comparative Physiology 60(1), pp. 85–103. doi: 10.1002/jcp.1030600111.

Jun, Y.W., Kim, H.R., Reo, Y.J., Dai, M. and Ahn, K.H. 2017. Addressing the autofluorescence issue in deep tissue imaging by two-photon microscopy: the significance of far-red emitting dyes. *Chem. Sci.* 8(11), pp. 7696–7704. doi: 10.1039/C7SC03362A.

Kapuscinski, J. 1995. DAPI: a DNA-Specific Fluorescent Probe. *Biotechnic & Histochemistry* 70(5), pp. 220–233. doi: 10.3109/10520299509108199.

Karsten, L., Goett-Zink, L., Schmitz, J., Hoffrogge, R., Grünberger, A., Kottke, T. and Müller, K.M. 2022. Genetically Encoded Ratiometric pH Sensors for the Measurement of Intra- and Extracellular pH and Internalization Rates. *Biosensors* 12(5), p. 271. doi: 10.3390/bios12050271.

Keresztury, G. 2006. Raman Spectroscopy: Theory. doi: 10.1002/9780470027325.s0109.

Kerppola, T.K. 2008. Bimolecular Fluorescence Complementation (BiFC) Analysis as a Probe of Protein Interactions in Living Cells. *Annual Review of Biophysics* 37(1), pp. 465–487. doi: 10.1146/annurev.biophys.37.032807.125842.

El Khatib, M., Martins, A., Bourgeois, D., Colletier, J.-P. and Adam, V. 2016. Rational design of ultrastable and reversibly photoswitchable fluorescent proteins for super-resolution imaging of the bacterial periplasm. *Scientific Reports* 6(1), p. 18459. doi: 10.1038/srep18459.

Kim, H., Ju, J., Lee, H.N., Chun, H. and Seong, J. 2021. Genetically encoded biosensors based on fluorescent proteins. *Sensors (Switzerland)* 21(3), pp. 1–18. doi: 10.3390/s21030795.

Lakowicz, J.R. 2006. Instrumentation for Fluorescence Spectroscopy. In: *Principles of Fluorescence Spectroscopy*. Boston, MA: Springer US, pp. 27–61. doi: 10.1007/978-0-387-46312-4_2.

Lavis, L.D. 2021. Live and Let Dye. *Biochemistry* 60(46), pp. 3539–3546. doi: 10.1021/acs.biochem.1c00299.

Liguori, I. et al. 2018. Oxidative stress, aging, and diseases. *Clinical Interventions in Aging* Volume 13, pp. 757–772. doi: 10.2147/CIA.S158513.

Lin, M.Z. et al. 2009. Autofluorescent Proteins with Excitation in the Optical Window for Intravital Imaging in Mammals. *Chemistry & Biology* 16(11), pp. 1169–1179. doi: 10.1016/j.chembiol.2009.10.009.

Lindley, M., de Pablo, J.G., Peterson, J.W., Isozaki, A., Hiramatsu, K. and Goda, K. 2021. High-throughput Raman-activated cell sorting in the fingerprint region. doi: 10.1101/2021.05.16.444384.

Long, D.A. 2002. *The Raman Effect*. Wiley. doi: 10.1002/0470845767.

Luker, K.E. et al. 2015. Comparative study reveals better far-red fluorescent protein for whole body imaging. *Scientific Reports* 5(1), p. 10332. doi: 10.1038/srep10332.

Mackerell, A.D., Feig, M. and Brooks, C.L. 2004. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry* 25(11), pp. 1400–1415. doi: 10.1002/jcc.20065.

van Manen, H.-J., Kraan, Y.M., Roos, D. and Otto, C. 2005. Single-cell Raman and fluorescence microscopy reveal the association of lipid bodies with phagosomes in leukocytes. *Proceedings of the National Academy of Sciences* 102(29), pp. 10159–10164. doi: 10.1073/pnas.0502746102.

Di Marzo, N., Chisci, E. and Giovannoni, R. 2018. The Role of Hydrogen Peroxide in Redox-Dependent Signaling: Homeostatic and Pathological Responses in Mammalian Cells. *Cells* 7(10), p. 156. doi: 10.3390/cells7100156.

Matlashov, M.E. et al. 2015. Fluorescent ratiometric pH indicator SypHer2: Applications in neuroscience and regenerative biology. *Biochimica et biophysica acta* 1850(11), pp. 2318–28. doi: 10.1016/j.bbagen.2015.08.002.

Matz, M. V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L. and Lukyanov, S.A. 1999. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nature Biotechnology* 17(10), pp. 969–973. doi: 10.1038/13657.

Merzlyak, E.M. et al. 2007. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nature Methods* 4(7), pp. 555–557. doi: 10.1038/nmeth1062.

Minoshima, M. and Kikuchi, K. 2017. Photostable and photoswitching fluorescent dyes for super-resolution imaging. *JBIC Journal of Biological Inorganic Chemistry* 22(5), pp. 639–652. doi: 10.1007/s00775-016-1435-y.

Miyake-Stoner, S.J., Refakis, C.A., Hammill, J.T., Lusic, H., Hazen, J.L., Deiters, A. and Mehl, R.A. 2010. Generating Permissive Site-Specific Unnatural Aminoacyl-tRNA Synthetases. *Biochemistry* 49(8), pp. 1667–1677. doi: 10.1021/bi901947r.

Miyawaki, A. and Niino, Y. 2015. Molecular Spies for Bioimaging—Fluorescent Protein-Based Probes. *Molecular Cell* 58(4), pp. 632–643. doi: 10.1016/j.molcel.2015.03.002.

Miyawaki, A., Shcherbakova, D.M. and Verkhusha, V. V. 2012. Red fluorescent proteins: chromophore formation and cellular applications. *Current Opinion in Structural Biology* 22(5), pp. 679–688. doi: 10.1016/j.sbi.2012.09.002.

Morin, J.G. and Hastings, J.W. 1971. Energy transfer in a bioluminescent system. *Journal of Cellular Physiology* 77(3), pp. 313–318. doi: 10.1002/jcp.1040770305.

Mukherjee, S., Manna, P., Douglas, N., Chapagain, P.P. and Jimenez, R. 2023. Conformational Dynamics of mCherry Variants: A Link between Side-Chain Motions and Fluorescence Brightness. *The Journal of Physical Chemistry B* 127(1), pp. 52–61. doi: 10.1021/acs.jpcb.2c05584.

Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K. and Miyawaki, A. 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology* 20(1), pp. 87–90. doi: 10.1038/nbt0102-87.

Ong, W.J.-H. et al. 2011. Function and structure of GFP-like proteins in the protein data bank. *Molecular BioSystems* 7(4), p. 984. doi: 10.1039/c1mb05012e.

Ormö, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y. and Remington, S.J. 1996. Crystal Structure of the *Aequorea victoria* Green Fluorescent Protein. *Science* 273(5280), pp. 1392–1395. doi: 10.1126/science.273.5280.1392.

Orth, A. et al. 2018. Super-multiplexed fluorescence microscopy via photostability contrast. *Biomedical Optics Express* 9(7), p. 2943. doi: 10.1364/BOE.9.002943.

Pearson, D.A., Lister-James, J., McBride, W.J., Wilson, D.M., Martel, L.J., Civitello, E.R. and Dean, R.T. 1996. Thrombus Imaging Using Technetium-99m-Labeled High-Potency GPIIb/IIIa Receptor Antagonists. Chemistry and Initial Biological Studies. *Journal of Medicinal Chemistry* 39(7), pp. 1372–1382. doi: 10.1021/jm950112e.

Pédelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. and Waldo, G.S. 2006. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* 24(1), pp. 79–88. doi: 10.1038/nbt1172.

Piatkevich, K.D., Efremenko, E.N., Verkhusha, V. V and Varfolomeev, S.D. 2010. Red fluorescent proteins and their properties. *Russian Chemical Reviews* 79(3), pp. 243–258. doi: 10.1070/RC2010v079n03ABEH004095.

Pletnev, S., Subach, F. V, Dauter, Z., Wlodawer, A. and Verkhusha, V. V. 2010. Understanding blue-to-red conversion in monomeric fluorescent timers and hydrolytic degradation of their chromophores. *Journal of the American Chemical Society* 132(7), pp. 2243–53. doi: 10.1021/ja908418r.

Porcelli, A.M., Ghelli, A., Zanna, C., Pinton, P., Rizzuto, R. and Rugolo, M. 2005. pH difference across the outer mitochondrial membrane measured with a green fluorescent protein mutant. *Biochemical and Biophysical Research Communications* 326(4), pp. 799–804. doi: 10.1016/j.bbrc.2004.11.105.

Pronk, S. et al. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7), pp. 845–854. doi: 10.1093/bioinformatics/btt055.

Pruccoli, A., Kocademir, M., Winterhalder, M.J. and Zumbusch, A. 2023. Electronically Preresonant Stimulated Raman Scattering Microscopy of Weakly Fluorescing Chromophores. *The Journal of Physical Chemistry B* 127(27), pp. 6029–6037. doi: 10.1021/acs.jpcb.3c01407.

Raček, T., Schindler, O., Toušek, D., Horský, V., Berka, K., Koča, J. and Svobodová, R. 2020. Atomic Charge Calculator II: web-based tool for the calculation of partial atomic charges. *Nucleic Acids Research* 48(W1), pp. W591–W596. doi: 10.1093/nar/gkaa367.

RAMAN, C. V. and KRISHNAN, K.S. 1928. A New Type of Secondary Radiation. *Nature* 121(3048), pp. 501–502. doi: 10.1038/121501c0.

Reddington, S.C., Driezis, S., Hartley, A.M., Watson, P.D., Rizkallah, P.J. and Jones, D.D. 2015. Genetically encoded phenyl azide photochemistry drives positive and negative functional modulation of a red fluorescent protein. *RSC Advances* 5(95), pp. 77734–77738. doi: 10.1039/C5RA13552D.

Reddington, S.C., Rizkallah, P.J., Watson, P.D., Pearson, R., Tippmann, E.M. and Jones, D.D. 2013. Different Photochemical Events of a Genetically Encoded Phenyl Azide Define and Modulate GFP Fluorescence. *Angewandte Chemie International Edition* 52(23), pp. 5974–5977. doi: 10.1002/anie.201301490.

Reifenrath, M. and Boles, E. 2018. A superfolder variant of pH-sensitive pHluorin for in vivo pH measurements in the endoplasmic reticulum. *Scientific Reports* 8(1), p. 11985. doi: 10.1038/s41598-018-30367-z.

Remington, S.J. 2006. Fluorescent proteins: maturation, photochemistry and photophysics. *Current Opinion in Structural Biology* 16(6), pp. 714–721. doi: 10.1016/j.sbi.2006.10.001.

Remington, S.J. 2011. Green fluorescent protein: A perspective. *Protein Science* 20(9), pp. 1509–1519. doi: 10.1002/pro.684.

Rodriguez, E.A. et al. 2017. The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins. *Trends in Biochemical Sciences* 42(2), pp. 111–129. doi: 10.1016/j.tibs.2016.09.010.

Ruff, F., Fábián, A., Farkas, Ö. and Kucsman, Á. (2009), Mechanism for the Oxidation of Sulfides and Sulfoxides with Periodates: Reactivity of the Oxidizing Species. Eur. J. Org. Chem., 2009: pp. 2102-2111. doi: 10.1002/ejoc.200801180

Sankaranarayanan, S., De Angelis, D., Rothman, J.E. and Ryan, T.A. 2000. The Use of pHluorins for Optical Measurements of Presynaptic Activity. *Biophysical Journal* 79(4), pp. 2199–2208. doi: 10.1016/S0006-3495(00)76468-X.

Sarangi, B.R. 2022. Principles and Applications of Fluorescence Microscopy. In: *Optical Spectroscopic and Microscopic Techniques*. Singapore: Springer Nature Singapore, pp. 197–213. doi: 10.1007/978-981-16-4550-1_9.

Schindelin, J. et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* 9(7), pp. 676–682. doi: 10.1038/nmeth.2019.

Seifert, M.H.J., Georgescu, J., Ksiazek, D., Smialowski, P., Rehm, T., Steipe, B. and Holak, T.A. 2003. Backbone Dynamics of Green Fluorescent Protein and the Effect of Histidine 148 Substitution. *Biochemistry* 42(9), pp. 2500–2512. doi: 10.1021/bi026481b.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E. and Tsien, R.Y. 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology* 22(12), pp. 1567–1572. doi: 10.1038/nbt1037.

Shaner, N.C., Patterson, G.H. and Davidson, M.W. 2007. Advances in fluorescent protein technology. *Journal of Cell Science* 120(24), pp. 4247–4260. doi: 10.1242/jcs.005801.

Shaner, N.C., Steinbach, P.A. and Tsien, R.Y. 2005. A guide to choosing fluorescent proteins. *Nature Methods* 2(12), pp. 905–909. doi: 10.1038/nmeth819.

Shcherbakova, D.M. and Verkhusha, V. V. 2013. Near-infrared fluorescent proteins for multicolor in vivo imaging. *Nature Methods* 10(8), pp. 751–754. doi: 10.1038/nmeth.2521.

Shimomura, O., Johnson, F.H. and Saiga, Y. 1962. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, *Aequorea*. *Journal of Cellular and Comparative Physiology* 59(3), pp. 223–239. doi: 10.1002/jcp.1030590302.

Shinoda, H., Shannon, M. and Nagai, T. 2018. Fluorescent Proteins for Investigating Biological Events in Acidic Environments. *International Journal of Molecular Sciences* 19(6), p. 1548. doi: 10.3390/ijms19061548.

Shu, X. et al. 2007. Ultrafast Excited-State Dynamics in the Green Fluorescent Protein Variant S65T/H148D. 1. Mutagenesis and Structural Studies ,. *Biochemistry* 46(43), pp. 12005–12013. doi: 10.1021/bi7009037.
Shu, X., Shaner, N.C., Yarbrough, C.A., Tsien, R.Y. and Remington, S.J. 2006. Novel Chromophores and Buried Charges Control Color in mFruits ,. *Biochemistry* 45(32), pp. 9639–9647. doi: 10.1021/bi060773l.

Singh, R., Yadav, V., Dhillon, A.K., Sharma, A., Ahuja, T. and Siddhanta, S. 2023. Emergence of Raman Spectroscopy as a Probing Tool for Theranostics. *Nanotheranostics* 7(3), pp. 216–235. doi: 10.7150/ntno.81936.

Smith, E. and Dent, G. 2005. *Modern Raman Spectroscopy – A Practical Approach*. Wiley. doi: 10.1002/0470011831.

Smith, J.C. and Karplus, M. 1992. Empirical force field study of geometries and conformational transitions of some organic molecules. *Journal of the American Chemical Society* 114(3), pp. 801–812. doi: 10.1021/ja00029a001.

Sousa da Silva, A.W. and Vranken, W.F. 2012. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes* 5(1), p. 367. doi: 10.1186/1756-0500-5-367.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J.C. 2005. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* 26(16), pp. 1701–1718. doi: 10.1002/jcc.20291.

Stokes, A.H. et al. 2000. *p* -Ethynylphenylalanine. *Journal of Neurochemistry* 74(5), pp. 2067–2073. doi: 10.1046/j.1471-4159.2000.0742067.x.

Strack, R.L., Strongin, D.E., Mets, L., Glick, B.S. and Keenan, R.J. 2010. Chromophore Formation in DsRed Occurs by a Branched Pathway. *Journal of the American Chemical Society* 132(24), pp. 8496–8505. doi: 10.1021/ja1030084.

Strianese, M., Staiano, M., Ruggiero, G., Labella, T., Pellecchia, C. and D'Auria, S. 2012. Fluorescence-Based Biosensors. pp. 193–216. doi: 10.1007/978-1-61779-806-1_9.

Strongin, D.E. et al. 2007. Structural rearrangements near the chromophore influence the maturation speed and brightness of DsRed variants. *Protein Engineering, Design and Selection* 20(11), pp. 525–534. doi: 10.1093/protein/gzm046.

Subach, F. V, Piatkevich, K.D. and Verkhusha, V. V. 2011. Directed molecular evolution to design advanced red fluorescent proteins. *Nature Methods* 8(12), pp. 1019–1026. doi: 10.1038/nmeth.1776.

Subach, F. V. and Verkhusha, V. V. 2012. Chromophore Transformations in Red Fluorescent Proteins. *Chemical Reviews* 112(7), pp. 4308–4327. doi: 10.1021/cr2001965.

Suhalim, J.L., Boik, J.C., Tromberg, B.J. and Potma, E.O. 2012. The need for speed. *Journal of Biophotonics* 5(5–6), pp. 387–395. doi: 10.1002/jbio.201200002.

Szymanski, H.A. ed. 1970. *Raman Spectroscopy*. Boston, MA: Springer US. doi: 10.1007/978-1-4684-3027-1.

Thorn, K. 2017. Genetically encoded fluorescent tags. *Molecular biology of the cell* 28(7), pp. 848–857. doi: 10.1091/mbc.E16-07-0504.

Tian, L. et al. 2009. Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* 6(12), pp. 875–881. doi: 10.1038/nmeth.1398.

Tinevez, J.-Y. et al. 2017. TrackMate: An open and extensible platform for single-particle tracking. *Methods (San Diego, Calif.)* 115, pp. 80–90. doi: 10.1016/j.ymeth.2016.09.016.

Toseland, C.P. 2013. Fluorescent labeling and modification of proteins. *Journal of Chemical Biology* 6(3), pp. 85–95. doi: 10.1007/s12154-013-0094-5.

Tsien, R.Y. 1998. THE GREEN FLUORESCENT PROTEIN. *Annual Review of Biochemistry* 67(1), pp. 509–544. doi: 10.1146/annurev.biochem.67.1.509.

Valeur, B. and Berberan-Santos, M.N. 2012. *Molecular Fluorescence*. Wiley. doi: 10.1002/9783527650002.

Vardaki, M.Z., Gregoriou, V.G. and Chochos, C.L. 2024. Biomedical applications, perspectives and tag design concepts in the cell – silent Raman window. *RSC Chemical Biology* 5(4), pp. 273–292. doi: 10.1039/D3CB00217A.

Vegh, R.B. et al. 2014. Chromophore Photoreduction in Red Fluorescent Proteins Is Responsible for Bleaching and Phototoxicity. *The Journal of Physical Chemistry B* 118(17), pp. 4527–4534. doi: 10.1021/jp500919a.

Wachter, R.M., Watkins, J.L. and Kim, H. 2010. Mechanistic Diversity of Red Fluorescence Acquisition by GFP-like Proteins. *Biochemistry* 49(35), pp. 7417–7427. doi: 10.1021/bi100901h.

Waldo, G.S., Standish, B.M., Berendzen, J. and Terwilliger, T.C. 1999. Rapid protein-folding assay using green fluorescent protein. *Nature biotechnology* 17(7), pp. 691–5. doi: 10.1038/10904.

Wei, L. et al. 2017. Super-multiplex vibrational imaging. *Nature* 544(7651), pp. 465–470. doi: 10.1038/nature22051.
Wei, L. and Min, W. 2018. Electronic Preresonance Stimulated Raman Scattering Microscopy. *Journal of Physical Chemistry Letters* 9(15), pp. 4294–4301. doi: 10.1021/acs.jpclett.8b00204.

Wu, S.-Y. et al. 2022. A sensitive and specific genetically-encoded potassium ion biosensor for in vivo applications across the tree of life. *PLOS Biology* 20(9), p. e3001772. doi: 10.1371/journal.pbio.3001772.

Würth, C., Grabolle, M., Pauli, J., Spieles, M. and Resch-Genger, U. 2013. Relative and absolute determination of fluorescence quantum yields of transparent samples. *Nature Protocols* 8(8), pp. 1535–1550. doi: 10.1038/nprot.2013.087.

Xiong, H., Shi, L., Wei, L., Shen, Y., Long, R., Zhao, Z. and Min, W. 2019. Stimulated Raman Excited Fluorescence Spectroscopy and Imaging. *Nature photonics* 13(6), pp. 412–417. doi: 10.1038/s41566-019-0396-4.

XIONG, Z., PIGNATARO, G., LI, M., CHANG, S. and SIMON, R. 2008. Acid-sensing ion channels (ASICs) as pharmacological targets for neurodegenerative diseases. *Current Opinion in Pharmacology* 8(1), pp. 25–32. doi: 10.1016/j.coph.2007.09.001.

Yang, F., Moss, L.G. and Phillips, G.N. 1996. The molecular structure of green fluorescent protein. *Nature Biotechnology* 14(10), pp. 1246–1251. doi: 10.1038/nbt1096-1246.

Yarbrough, D., Wachter, R.M., Kallio, K., Matz, M. V. and Remington, S.J. 2001. Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-Å resolution. *Proceedings of the National Academy of Sciences* 98(2), pp. 462–467. doi: 10.1073/pnas.98.2.462.

Zanetti-Domingues, L.C., Tynan, C.J., Rolfe, D.J., Clarke, D.T. and Martin-Fernandez, M. 2013. Hydrophobic Fluorescent Probes Introduce Artifacts into Single Molecule Tracking Experiments Due to Non-Specific Binding. *PLoS ONE* 8(9), p. e74200. doi: 10.1371/journal.pone.0074200.

Zhang, R., Nguyen, M.T. and Ceulemans, A. 2005. A concerted mechanism of proton transfer in green fluorescent protein. A theoretical study. *Chemical Physics Letters* 404(4–6), pp. 250–256. doi: 10.1016/j.cplett.2005.01.094.

Zhang, S., Qi, Y., Tan, S.P.H., Bi, R. and Olivo, M. 2023. Molecular Fingerprint Detection Using Raman and Infrared Spectroscopy Technologies for Cancer Detection: A Progress Review. *Biosensors* 13(5), p. 557. doi: 10.3390/bios13050557.

Zhang, Y., Hong, H. and Cai, W. 2010. *Imaging with Raman Spectroscopy*.

Zheng, Q. and Blanchard, S.C. 2013. Single Fluorophore Photobleaching. In: *Encyclopedia of Biophysics*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 2324–2326. doi: 10.1007/978-3-642-16712-6_482.

Zhu, H., Tamura, T., Fujisawa, A., Nishikawa, Y., Cheng, R., Takato, M. and Hamachi, I. 2020. Imaging and Profiling of Proteins under Oxidative Conditions in Cells and Tissues by Hydrogen-Peroxide-Responsive Labeling. *Journal of the American Chemical Society* 142(37), pp. 15711–15721. doi: 10.1021/jacs.0c02547.

Zitti, A. and Jones, D. 2023. Expanding the genetic code: a non-natural amino acid story. *The Biochemist* 45(1), pp. 2–6. doi: 10.1042/bio_2023_102.